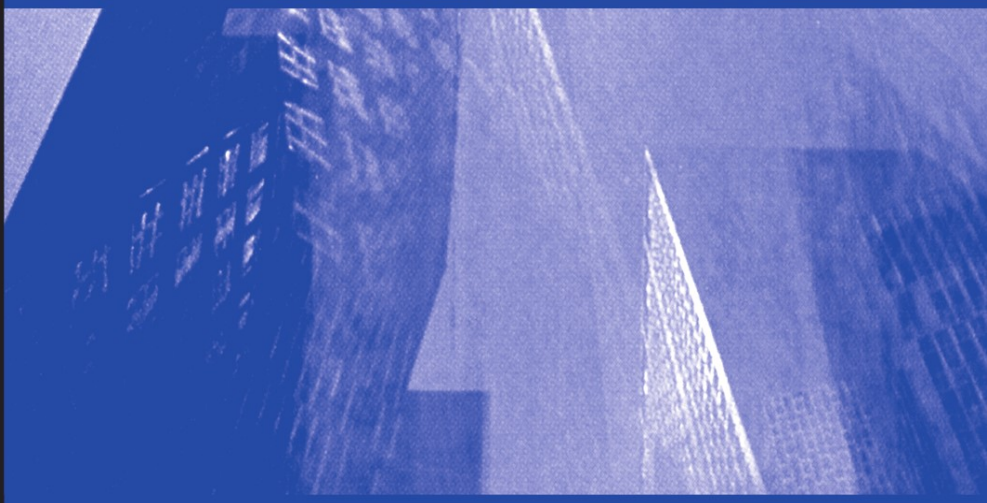
ADVANCED TOPICS IN SCIENCE AND TECHNOLOGY IN CHINA

Qihu Li

Digital Sonar Design in Underwater Acoustics

Principles and Applications



ZHEJIANG UNIVERSITY PRESS
浙江大学出版社
Springer

ADVANCED TOPICS
IN SCIENCE AND TECHNOLOGY IN CHINA

ADVANCED TOPICS IN SCIENCE AND TECHNOLOGY IN CHINA

Zhejiang University is one of the leading universities in China. In Advanced Topics in Science and Technology in China, Zhejiang University Press and Springer jointly publish monographs by Chinese scholars and professors, as well as invited authors and editors from abroad who are outstanding experts and scholars in their fields. This series will be of interest to researchers, lecturers, and graduate students alike.

Advanced Topics in Science and Technology in China aims to present the latest and most cutting-edge theories, techniques, and methodologies in various research areas in China. It covers all disciplines in the fields of natural science and technology, including but not limited to, computer science, materials science, life sciences, engineering, environmental sciences, mathematics, and physics.

Qihu Li

Digital Sonar Design in Underwater Acoustics

Principles and Applications

With 418 figures

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

Springer

Author
Prof. Qihu Li
Institute of Acoustics,
Chinese Academy of Sciences
100190, Beijing, China
E-mail: lqh@ocean.ioa.ac.cn

ISSN 1995-6819 e-ISSN 1995-6827
Advanced Topics in Science and Technology in China

ISBN 978-7-308-07988-4
Zhejiang University Press, Hangzhou

ISBN 978-3-642-18289-1 e-ISBN 978-3-642-18290-7
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011920978

© Zhejiang University Press, Hangzhou and Springer-Verlag Berlin Heidelberg 2012
This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer Violations are liable to prosecution under the German Copyright Law.
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Printed on acid-free paper

Springer is a part of Springer Science+Business Media (www.springer.com)

To my wife Xiaoping,
my children Hai and Ye,
my grandson Mike

Preface

Sound wave is the only physical medium which can propagate over a long distance in the ocean. Other physical media, such as electromagnetic, optical or laser beams will quickly dissipate in the ocean because sea water is a good electrical conductor.

Sound navigation and ranging (Sonar) is a technique which is based on sound propagation to navigate, to communicate with or to detect other vessels. It is used extensively in ocean development and military applications.

Digital sonar is a sonar system that uses digital signal processing theory and techniques and that realizes system integration in a digital manner. Compared with the analogue processing of signal, digital processing has many advantages, as digital data are easy to store, transmit, and process.

With the rapid development of the semiconductor industry, digital chips have become more and more powerful and can now complete very complicated algorithms which were impossible to perform in the past. As a result, the performance of digital sonar has improved greatly.

The design of a modern digital sonar depends not only on the theory of underwater acoustic signal processing but also on our knowledge of the ocean environment, including ambient noise, the acoustic channel, the characteristics of the sea surface and sea floor, etc.

This book aims to describe the basic design principles of digital sonar and its applications. We have tried to focus our description on the basic theory and design techniques of modern digital sonar by stripping away all unimportant details.

Most theoretical results are illustrated with a practical example so that the reader should understand the application background of the basic theory.

I would like to emphasize that the best way of learning the theory of digital sonar design is to design a modern digital sonar for a particular application area. Experiments in lake or sea water are essential. I believe that experiments at sea provide some of the most important lessons for aspiring sonar designers.

The theoretical and experimental results described in this book include a wide

variety of contributions from my colleagues and graduate students at the Institute of Acoustics, Chinese Academy of Sciences. I would like to express my sincerely thanks to Dr. C.H. Zhang, J.B. Liu, B.X. Xie, B.L. Tu, X.P. Chen, T. Xu, C.Y. Sun, S.Q. Li, L. Yin, H.N. Huang, J.S. Tang, J.Y. Liu, J.D. Luan, G.Q. Sun, X.T. Yang, M. Li, H.B. Yu, G. Liu, L. Wang, X.H. Chen, C.H. Wei and L.P. Dong.

Any comments or criticisms about this book are most welcome.

Qihu Li
June, 2011

Contents

1 Brief History of Digital Sonar Development 1

1.1 Evolution of Digital Sonar Systems 1
1.2 Main Features of Digital Sonar 5
1.3 Today and Tomorrow of Digital Sonar 10
References 15

2 Basic Theory of Digital Signal Processing 19

2.1 Digital Conversion of Analogue Signal: Quantization and Sampling 19
2.1.1 Signal Sampling 21
2.1.2 Signal Quantization 26
2.1.3 Signal Companding 28
2.1.4 $\Sigma - \Delta$ Modulation 30
2.2 Digital Filtering of Signal 32
2.2.1 Linear Digital Filtering 32
2.2.2 Transfer Function of a Linear System 34
2.2.3 Classification of Digital Filters 36
2.2.4 Cascade of Digital Filters 39
2.2.5 Examples of Digital Filters 40
2.3 Characteristics of Digital Signals in Time Domain and Frequency Domain 41
2.3.1 Fourier Transform of Signal 41
2.3.2 Wiener-Khinchine Theorem 43
2.3.3 Discrete Fourier Transform 43
2.3.4 Digital Feature of Signal Represented by Discrete Samples 45
2.3.5 Algorithm of Fast Fourier Transform 47
2.3.6 Calculation of DFT for Real Value Data 50
2.4 Basic Processing Technique for One-Dimensional Digital Signal 51
2.4.1 Local Average Filtering 51
2.4.2 Median Value Filtering 54
2.4.3 Threshold Filtering and Truncate Filtering 56

2.5 Two-Dimensional Digital Image Signal Processing 58
2.5.1 Definition of Digital Images 58
2.5.2 2D Fourier Transform 59
2.5.3 2D Cosine Discrere Transforms 60
2.5.4 Typical Image Processing Techniques 62
2.5.5 Time / Bearing Display in Digital Sonar 65
2.6 New Topics of Digital Signal Processing: Wavelet Transform and Fractal Transform 67
2.6.1 Fractal Transform 67
2.6.2 Wavelet Transform 72
References 77

3 Detection and Estimation Theory of Digital Signals 81

3.1 Some Basic Results from Probability Theory and Mathematical Statistics 81
3.1.1 Basic Definition of Probability 82
3.1.2 Conditional Probability 85
3.1.3 Random Variable and Distribution Function 86
3.1.4 Digital Characteristics of Random Variables 89
3.1.5 Large Number Law and Central Limit Theorem 92
3.1.6 Random Process (Stochastic Process) 94
3.2 Introduction to the Basic Concepts of Information Theory 96
3.2.1 Information and Entropy 97
3.2.2 The Coding Theorem of a Discrete Information Source 103
3.3 The Optimum Receiving Theory of Weak Signal in Background Noise 106
3.3.1 Basic Concepts of Statistical Hypothesis Tests 106
3.3.2 Optimum Detection Criterion 108
3.3.3 N-P Test 109
3.3.4 Multiple Observations 113
3.3.5 Wald Sequential Test 115
3.4 Wiener Filtering, Matched Filtering and Adaptive Filtering for Stationary Random Signal 116
3.4.1 Basic Relation of Input / Output of a Linear System for Stationary Random Signal 116
3.4.2 Wiener Filtering 119
3.4.3 Matched Filter 122
3.4.4 Adaptive Filtering 126
3.5 Kalman Filtering for Non-stationary Digital Signal 132
3.5.1 Kalman Filtering of a One-Dimensional Observation Model 133
3.5.2 Kalman Filtering of Multiple Channels 136

3.6 Parameter Estimation of Random Signal 138
3.6.1 Test of Stationariness and Ergodicness of a Random Signal 138
3.6.2 Basic Requirements for a Statistic 139
3.6.3 Some Estimates Used Frequently in Sonar Design 140
3.6.4 Cramer-Rao Low Bound 142
3.6.5 Example (Mean Value Estimate) 144
3.6.6 Model-Free Estimates 145
References 147

4 General Principles of Sonar Design 151

4.1 Determination of Sonar System Specifications 151
4.1.1 Category of Sonar 152
4.1.2 Relationship between Tactical and Technical Specifications 154
4.1.3 Technical Specification Related Concepts in Sonar Design 159
4.1.4 Basic Concepts of Sonar Specifications 166
4.2 Design Procedure of Digital Sonar: the Sonar Equation 168
4.2.1 Parameters of the Sonar Equation 168
4.2.2 Active Sonar Equation 169
4.2.3 Passive Sonar Qquation 171
4.2.4 Calculation of the Sonar Ranging Distance 172
4.3 Ambient Noise 175
4.3.1 Main Source of Ambient Noise in the Ocean 175
4.3.2 Frequency spectrum of ambient noise 176
4.3.3 Minimum Ambient Noise 177
4.3.4 Homogeneous and Isotropic Noise Fields 178
4.3.5 Cylindrical and Spherical Model of Ambient Noise 181
4.3.6 Vertical Directivity of Ambient Noise 182
4.4 Radiated Noise from Underwater Target and Platform Noise 183
4.4.1 Sources of Radiated Noise 183
4.4.2 Radiated Noise from Submarines 185
4.4.3 Radiated Noise of Surface Ships 187
4.4.4 Radiated Noise of Torpedo 188
4.4.5 Self-noise of Vessels 188
4.4.6 Auto-correlation Function of Target Noise 191
4.5 Reverberation 193
4.5.1 Sources of Reverberation 193
4.5.2 Short Distance Reverberation Theory 193
4.5.3 Volume and Boundary Reverberation Levels 195
4.5.4 Relationship between Reverberation Strength and Impulse Duration 197

4.5.5 Statistical Characteristics of Reverberations 198
4.6 Sound Propagation in the Ocean and Underwater Acoustic Channel 202
4.6.1 Sound Wave and Vibration 202
4.6.2 Velocity of Sound in the Sea: Sound Speed Profile 203
4.6.3 Wave and Ray Theories of Underwater Sound Fields 207
4.6.4 Transmission Loss 212
4.6.5 Sound Absorption in Sea Water 213
4.6.6 Upper Boundary of Acoustic Channel: the Sea Surface and Its Acoustic Characteristics 214
4.6.7 Lower Boundary of Acoustic Channel: the Sea Floor and Its Characteristics 216
4.6.8 Use of Propagation Characteristics in Sonar Design 216
4.6.9 Average Structure of a Sound Field in Shallow Water 219
4.6.10 Use of Transmission Loss in Sonar Ranging Distance Prediction 221
4.7 Hydrophone Array and Beamforming 226
4.7.1 Directivity Function (Beam Pattern) 226
4.7.2 Conventional Beamforming 230
4.7.3 Equal-Spaced Line Array 232
4.7.4 Uniformly Distributed Discrete Circle Array 238
4.7.5 Circle Array Baffling and Arc Array 242
4.7.6 Product Theorem of Directivity Function of a Line Array 244
4.7.7 Weighting of an Array 247
4.7.8 General Expression of Directivity Function 253
4.7.9 Continuous Distributed Array 258
4.8 Calculation of Sonar System Gain 259
4.8.1 Spatial Gain of Sonar System 260
4.8.2 Calculation of Time Processing Gain of Passive Sonar 261
4.8.3 Calculation of Time Processing Gain of Active Sonar 267
4.9 Gain Loss of a Sonar System in the Interface of Various Sub-systems 270
4.9.1 Relationship between Sonar System Gain and Input signal-to-noise Ratio 270
4.9.2 Gain Loss at the Interface of a Hydrophone and an A / D Converter 271
4.9.3 Interface Loss Due to Time Integration 272
4.9.4 Loss at the Interface of the Signal Processor and the Display System 273
4.10 Explosive Source of Underwater Sound 273
4.10.1 Main Characteristics of Explosive Sources of Underwater Sound 274

4.10.2 Measurement of Transmission Loss by Using Explosive Source 277
References 278

5 Design of Digital Sonar 285

5.1 System Architecture of Digital Sonar 285
5.1.1 Parallel Operation and Sequential Operation 286
5.1.2 Differences in Beamforming in Time Domain and Frequency Domain 287
5.1.3 Data Interchange among Sub-systems 291
5.1.4 Real Time Data Display of the Sonar System 291
5.1.5 COST Technology 292
5.2 Parameter Selection of Programmable Digital Sonar 292
5.3 Pre-processing and FIR filtering 298
5.3.1 Anti-aliazing Filtering 299
5.3.2 A / D Conversion 300
5.3.3 AGC and A / D Converter 301
5.3.4 FIR filtering 303
5.4 High Precision RAM Dynamic Beamforming and Optimum Spatial Filtering 307
5.4.1 Digital Beamforming Using SRAM 307
5.4.2 Optimum Spatial Filter 311
5.5 Target Characteristic Analysis: LOFAR, DEMON and Adaptive Line Enhancer 317
5.5.1 LOFAR and DEMON Analysis 318
5.5.2 Adaptive Line Enhancing 324
5.5.3 Combined Architecture of ANC and ALE 328
5.6 Post Processing of Multibeam Data and Wild Value Filtering 332
5.6.1 Square Detection and Absolute Value Detection 332
5.6.2 A Method of Stationarity for Non-stationary Data 334
5.6.3 Basic Technique of Wild Value Filtering 336
5.6.4 Wild Value Filtering for Two-Dimensional Time / Bearing Data 339
5.7 Display and Control: High Resolution Gray Scale Display and Background Equalization 340
5.7.1 Main Function of the Display / Control System 341
5.7.2 Applications of Human / Machine Interface Theory 342
5.7.3 Use of Color in Sonar Display 343
5.7.4 Gray Scale Conversion 344
5.7.5 Background Equalization 347
5.8 Reliability Design of Digital Sonar 349
5.8.1 The Series Standard for Reliability 350
5.8.2 Brief Introduction to Reliability Engineering 350

5.8.3 Hardware Reliability 351
5.8.4 Software Reliability 358
5.8.5 Technical Means of Hardware Reliability Design 358
5.8.6 Failure Self Test 360
5.9 Integrated Design of Digital Sonar: Multi-sensor, Multi-layer Data Fusion 361
5.9.1 Basic Concept of Data Fusion 361
5.9.2 Optimum Linear Data Fusion 363
5.9.3 Relationship of Weight Coefficients with SNR 369
5.9.4 Joint Detection and Estimation Problems 369
References 374

6 Implementation Methods of Various Functions of Digital Sonar 379

6.1 Precise Bearing for Target 379
6.1.1 Concept of the Bearing Accuracy of Sonar 380
6.1.2 Optimum Precise Bearing of Line Array and Circle Array 383
6.1.3 Implementation of Precise Bearing 386
6.1.4 Example of Realization of Precise Bearing Function in Digital Sonar 391
6.1.5 Method for Improving Bearing Accuracy 395
6.2 Automatic Tracking and Multi-target Resolution 398
6.2.1 Tracking Algorithm of Passive Sonar 398
6.2.2 Multi-target Resolution 404
6.2.3 Results of System Simulation 405
6.2.4 Method for Increasing the Resolution Ability of Multiple Targets 409
6.3 Moving Target Analysis: Pre-set Tracking 413
6.3.1 Concept of TMA 414
6.3.2 Pre-setting Tracking 415
6.3.3 TMA Research 418
6.4 Fast Moving Target Tracking: Torpedo Alarming 420
6.4.1 Possibility of Torpedo Alarming 421
6.4.2 Double Check Decision Strategy in Torpedo Alarming 422
6.4.3 Target Traces Extraction Technique 424
6.4.4 Example of System Simulation 426
6.5 Passive Ranging of Target Based on the Time Delay Estimation 427
6.5.1 Principle of Three-Point Ranging 427
6.5.2 Four-Point Passive Ranging 430
6.5.3 Analysis of Passive Ranging Accuracy 431
6.5.4 Realization of Passive Ranging 435
6.6 Target Classification and Recognition 436
6.6.1 Basic Requirement of Target Noise Classification 437
6.6.2 Concept of Artificial Neural Network 438

6.6.3 Simplified Expert System 440
6.6.4 An Example of an Expert System Based on Nearest Neighborhood Criteria 441
6.7 Active Sonar Signal Capture and Analysis 446
6.7.1 Theoretical Basis of Active Sonar Signal Capture 447
6.7.2 A Combined Time / Frequency Domain Signal Capture Technique 448
6.7.3 Signal Feature Extraction of Active Sonar Signal 453
6.8 Underwater Acoustic Communication and Adaptive Equalization of Underwater Acoustic Channel 455
6.8.1 Introduction of Underwater Acoustic Communication 456
6.8.2 Coding and Decoding 458
6.8.3 Frequency Spread Technique 461
6.8.4 Equalization of Underwater Acoustic Channel 463
6.9 Introduction of Multistatic Sonar 466
6.9.1 General Description of Multistatic Sonar Problem 467
6.9.2 Basic Formula for Calculating Detection Range in Multistatic Sonar 468
6.9.3 Realization of Bistatic Sonar Function by Using Monostatic Sonar 470
References 471

7 System Simulation Techniques in Digital Sonar Design 479

7.1 General Concept of System Simulation 479
7.1.1 Brief Introduction of System Simulation Technique 480
7.1.2 Development of Modern System Simulation Technique 481
7.2 Theoretical Basis of System Simulation: Monte Carlo Method 482
7.2.1 Method for Generating a Random Number 482
7.2.2 Basic Time Series in Digital System Simulation 485
7.3 Frequent Use Simulation Method in Underwater Acoustic Signal Processing 493
7.3.1 Simulation of the Sonar Environment Field 493
7.3.2 Simulation of Signal Processing Module in a Sonar System 497
7.3.3 Simulation of Post Processing 502
7.3.4 Specific Software for System Simulation 503
7.4 Design of Sonar Signal Simulator 503
7.4.1 General Concept of Sonar Simulator 504
7.4.2 Design Principle of General Purpose Sonar Simulator 505
7.4.3 Algorithm for Generating Infinite Non-cyclic Random Number 507
7.4.4 Generating of Array Signals 509
7.4.5 An Application Example of Sonar Simulator Design 512
References 514

8 Introduction of Typical Modern Digital Sonar 517

8.1 Concept of Model-Based Sonar 517
8.1.1 Basic Concept of Model-Based Sonar 518
8.1.2 Robust Signal Processing 520
8.1.3 Signal Processing System of Model-Based Sonar 521
8.1.4 Matched Field Filtering 522
8.2 Linear Towed Array Sonar 523
8.2.1 Brief History of Linear Towed Array Sonar 524
8.2.2 General Concept of Towed Array Sonar 525
8.2.3 Status of a Tow Cable under the Sea 527
8.2.4 Shape Distortion of Towed Array and Its Effect on Beamforming 530
8.2.5 Some Special Problems with Towed Array Sonar Design 533
8.2.6 Examples of Towed Array Sonar 542
8.3 Flank Array Sonar for Submarine 543
8.3.1 General Concept of Flank Array Sonar 543
8.3.2 Adaptive Canceling of Self Noise 544
8.3.3 Vibration Isolation of Hydrophone and Acoustic Damping Materials 548
8.3.4 Beamforming of Multiple Elements with Different SNR 549
8.3.5 System Configuration of Flank Array Sonar 550
8.3.6 Example of Typical Flank Array Sonar 551
8.4 Fixed Array Sonar 552
8.4.1 Brief History of Shore-Based Sonar System 553
8.4.2 Outline of Shore-Based Sonar Station 553
8.4.3 Some Specific Topics Concerning a Shore-Based Sonar Station 555
8.4.4 Example of Typical Fixed Array Sonar 559
8.5 Synthetic Aperture Sonar 559
8.5.1 Brief History of SAS 560
8.5.2 Basic Principle of SAS 561
8.5.3 Calculation of Directivity Function of Passive SAS 566
8.5.4 Basic Algorithm of Active SAS 568
8.5.5 An Experimental SAS Prototype 569
8.6 Concept of Network Centric Warfare and Its Application in Sonar 573
8.6.1 Basic Concept of NCW 573
8.6.2 Theory of Complexity System 575
8.6.3 US Navy Research Project in NCW 576
8.6.4 Applications of the NCW Concept in Underwater Acoustics 578
8.7 A New Topic in Sonar Design: Diver Detection Sonar 580
8.7.1 Basic Design Principle of DDS 580

8.7.2 Design Outline of DDS 581
References 583

9 Software and Hardware Support and Performance Evaluation in Digital Sonar Design 593

9.1 Software Package Matlab in Digital Signal Processing 594
9.1.1 Several Frequently Used Commands in Signal Processing 594
9.1.2 Examples 595
9.1.3 Other Features of Matlab 597
9.2 Programming Skills for C Language 599
9.3 General Purpose DSP chips 602
9.3.1 Concept of the General Purpose DSP Chip 602
9.3.2 Main Specifications of DSP Chips 603
9.3.3 Development Tools for DSP Chips 604
9.3.4 System Integration 604
9.4 Hardware Adjustment of Digital sonar 605
9.4.1 Evaluation Procedure of Digital Sonar 605
9.4.2 Step by Step Testing in a Laboratory of a Digital Sonar System 607
9.4.3 Wet End Testing of the Sonar System in a Laboratory 611
9.4.4 Limited Function Associate Testing of the Dry End and Wet End 612
9.5 Tactical and Technical Performance Evaluation of Digital Sonar: Comparison of Results of Laboratory Testing, Lake Testing and Experiment at Sea 612
9.5.1 From Sonar Design to Experiments at Sea 613
9.5.2 Measurement of Source Level and Sound Pressure Index 614
9.5.3 Measurement of Target Strength and Underwater Signal Channel 618
9.5.4 Calibration of Sonar Specification in Terms of the Result of Experiments at Sea 621
9.5.5 Statistical Average Problem in Experiment at Sea 624
9.5.6 Central Moment and Original Moment 627
9.5.7 Sea State Description and Some Results of Sonar Ranging Prediction 630
References 632

Index 635

Abbreviations

A
A-RCI: Acoustic rapid COTS insertion
ABF: Adaptive beamforming
ADCAP: Advanced capability
ADS: Advanced Deployable System
ADSL: Asynchronous digital subscriber line
AEGIS: Air early guard integration system
AERONET: Aerosol Robotic Network
AGC: Automatic gain control
AIP: Air Independent Propulsion
AMFP: Adaptive matched field processing
ALE: Adaptive line enhancer
ANC: Adaptive noise canceling
ASDIC: Anti-submarine division-ics
ASK: Amplitude Shift Keying
ATM: Asynchronous transfer mode
ATOC: Acoustic thermometry of ocean climate
AUTEC: Atlantic undersea test and evaluation center

B
BATS: Bermuda Atlantic Time Series

C
C2: Command and control
C3: Command, control and communication
C4ISR: Command, control, communication and computer, information, surveillance and recognition
CBF: Conventional beamforming
CDMA: Code division multiple access
CGS: Centimeter Gram Second
CORDIC: Coordinate rotation digital computer
COTS: Commercial off the shelf

CSDM: Cross spectral density matrix

D
DARPA: Defense advanced research project agency
DCT: Discrete Cosine Transform
DDS: Diver detection sonar
DEOS: Dynamics of Earth and Ocean System
DELTIC: Delay Line Time Compressor
DEMON: Demodulation on noise
DFT: Discrete Fourier transforms
DICANNE: Digital Interference Canceling Adaptive Network Nulling Equipment
DICASS: Directional command-activated sonobuoy system
DIFAR: Directional frequency analysis and ranging
DIMUS: Digital multi-beam system
DOA: Direction of arrival
DoD: Department of defense
DOFIX: Doppler fixing
DRAM: Dynamic random access memory
DWDM: Dense Wave Division Multiplexing
DS: Direct sequence
DSP: Digital signal processing, Digital Signal processor
DWDM: Dense wave division multiplexing

E
E3: Effective, Engagement, Envelope
ENIAC: Electronic Numerical Integrator and Calculator
ESONET: European sea floor observatory network
ESPRIT: Estimation of signal parameters via rotational invariant techniques
EVM: Evaluation module

F
FDDI: Fiber distributed data interface
FDS: Fixed Distributed System
FFT: Fast Fourier Transform
FH: Frequency hopper
FIR: Finite Impulse Response
FM: Frequency modulation
FOM: Figure of Merit
FRONT: Front resolving observatory network with telemetry
FSK: Frequency shift keying
FSS: Fixed Surveillance System

G
GOOS: Global ocean observatory system
GPS: Global positioning system

GSC: Grey scale conversion
GSS: Generic sonar simulator

H
H2O: Hawaii-2 observatory
HCI: Human computer interface
HMS: Hull mounted sonar

I
IBF: Inverse beamforming
IDFT: Inverse discrete Fourier transform
IDS: Intrude detection sonar
IFS: Iterated Function System
IIR: Infinite Impulse Response
IP: Internet protocol
ISDN: Integrated service digital network
IT-21: Information technology for the 21th century Initiative
IUSS: Integrated Undersea Surveillance System

J
JAMSTEC: Japan marine science & technology center
JPEG: Joint Photographic Experts Group
JV 2010: Joint vision 2010

L
LAN: Local area network
LOFAR: Low frequency analysis record
LOFIX: LOFAR Fixing

M
MAP: Maximum à posterior probability
MARS: Monterey accelerated research system
MBP: Model based processing
MDW: Mass destruction weapon
MFLOPS: Million Floating-Point Operations Per Second
MFP: Matched field processing
MIPS: Mega Instruction Per Second
MIUW: Mobile in-shore undersea warfare
MOPS: Mega Operation Per Second
MPEG: Moving Pictures Experts Group
MQPSK: Modified quadric-phase shift keying
MSE: Minimum mean square error
MTBCF: Mean Time Between Critical Failure
MTBF: Mean Time Between Failures
MTTF: Mean Time to Failure

MTTR: Mean Time to Repair
MUSIC: Multiple signal classification

N
NEPTUNE: North east Pacific time series undersea network experiments
NATO: North Atlantic Treaty Organization
NCOIC: Network Center Operation Industry Consortium
NCSL: Naval Coastal System Laboratory
NCW: Network Centric Warfare
NOPP: National oceanographic partnership program
NUWC: Navy underwater warfare center
NURC: NATO Undersea Research Center

O
ODN: Own Doppler null
OFDM: Orthogonal frequency division multiplexing
ONR: Office of naval research
ORBIS: Object-oriental rule based interactive system
OTA: Order truncate average

P
PCI: Personal Computer Interface
PDM: Pulse duration modulation
PFA: Passive Fixed Array
PLL: Phase lock loop
PRS: Passive ranging sonar
PSK: Phase Shift Keying
PVDF: Polyvinylidene fluoride

Q
QoS: Quality of service
QPSK: Quadraphase shift keying

R
ROC: Receiver operating curve

S
SABSOON: South Atlantic Bight Synoptic offshore observatory
SAS: Synthetic Aperture Sonar
SDS: Swimmer detection sonar
FRONT: Front-Resolving Observational Network with Telemetry
SHARC: Super Harvard architecture
SNR: Signal-to-noise ratio
SOFAR: Sound frequency and ranging channel
Sonar: Sound navigation and ranging

SONET: Synchronous optical network
SOSS: Soviet Ocean Surveillance System
SOSUS: Sound Surveillance Sonar
SPAWAR: Space and naval warfare systems center
SRAM: Static random access memory
SSBN: Nuclear ballistic missile submarine
SSN: Nuclear attack submarine
SSP: Sound speed profile
STDV: System technology demonstration vehicle
STFT: Short Time Fourier Transform
SURTASS: Surveillance Towed Array Sensor System
SVP: Sound velocity profile
SWATH: Small water area twin hull

T
TAGOS: Tactical Auxiliary General Ocean Surveillance
TALON: Tactical acoustic littoral ocean network
TCP/IP: Transmission control protocol/Internet protocol
TDMA: Time division multiple access
TMA: Target moving analysis
TOGA: Tropical ocean and the global atmosphere
TRM: Time reversal mirror
TVG: Time Varying Gain

U
UDT: Underwater defense technology
UUV: Unmanned undersea vehicle
UUVI: Unmanned undersea vehicle initiative

V
VDS: Variable depth sonar
VENUS: Victoria experimental network under the sea
VIM: Vibration isolate module
Virtual Collaboration
VME: Versa Module Europe
VTC: Video teleconference

W
WAA: Wide Aperture Array
WHOI: Woods hole oceanographic institute
WOCE: World ocean circulation experiment
WTD: Wehrternisshe Dienstelle

1

Brief History of Digital Sonar Development

Sonar is a technique / equipment based on sound propagation (usually in sea water) that is used to navigate, communicate with or detect other vessels[1-16], and the research work focuses on transducers, acoustic signal processing, acoustical oceanography, bioacoustics, micro-computers, and human / machine engineering[17-24]. The oretical basis of sonar design is closely related to other fields, such as acoustics, informatics, cybernetics and ocean engineering. In the 1960s, sonar designers began to use digital techniques to implement their designs. With the rapid development of digital signal processing theory and the microprocessor, digitization procedures for sonar systems developed quickly. The DSP (digital signal processor) chip is extensively used in sonar systems and considerably increases the processing capability and performance. Also, the design philosophy of digital sonar and its applications advances research in underwater acoustics and digital signal processing. This research has produced many new ideas and created many new opportunities, such as digital underwater acoustic communication, NCW (network centric warfare), underwater GPS (global position system) and digital DDS (diver detection sonar)[25-39]. This chapter will shortly introduce the history of sonar, will give an outline of digital sonar architecture, and will discuss the prospects for the theory and its application to digital sonar development[40-46].

1.1 Evolution of Digital Sonar Systems

A sound wave is the only physical medium which can be propagated over long distances in water. Other kinds of energy emission, such as optical and electromagnetic waves will quickly dissipate in the ocean. Although electromagnetic waves with very long wavelengths (100 m – 100 km) and some optical waves in specific wave windows (blue and green laser with wavelengths of 440 – 580 nm) can propagate a little further, the distances they can reach are several orders of magnitude less than those reached by underwater acoustic waves. Therefore, applications of non-acoustic waves underwater are very limited.

As the propagation media, sea water with the sea surface and sea bottom comprises the transmission channel of an acoustic wave, sometimes called the underwater acoustic channel. This transmission channel has very complicated behavior. It is very important in both theory and application to study the rules of generation, transmission and reception of underwater sound. Sonar is one of most important application areas of such studies. Sonar requires special equipment, usually consisting of two parts, the "wet end", which is installed underwater, and the "dry end", which is installed on a platform, out of the water (Fig. 1.1).

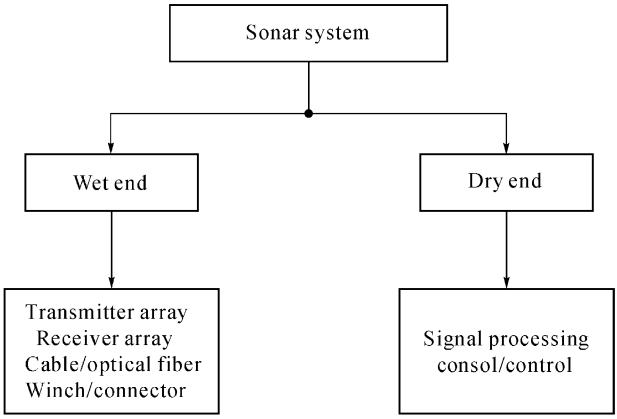


Fig. 1.1. Configuration of sonar systems

The wet end includes the components that are installed underwater and the dry end includes the components that are installed in the vessels, on the shore, in aircraft, etc. In some cases the distinction between these two parts may be a little confusing. For example, the transmission cable of a shore-based sonar system is still recognized as a part of the wet end even when it is housed onshore. The towed array winch is also recognized as a part of the wet end, although it is on the deck. Therefore, the definition of wet end and dry end is not absolutely distinguished by whether the equipment is in water.

Underwater sound has probably been used by marine animals for millions of years. The science of underwater acoustics can be traced back to 1490, when Leonardo Da Vinci wrote[11], "If you cause your ship to stop and place the head of a long tube in the water and place the outer extremity to your ear, you will hear ships at a great distance from you."

In 1687, Isaac Newton wrote the *Mathematical Principles of Natural Philosophy*, which included the first mathematical treatment of sound. The next major step in the development of underwater acoustics was made by Daniel Colladon, a Swiss physicist, and Charles Sturm, a French mathematician. In 1826, on Lake Geneva, they measured the time elapsed between a flash of light and the sound of a submerged ship's bell heard using an underwater listening horn. They measured a sound speed of 1,435 m/s over a 17 km distance, providing the first quantitative

measurement of the speed of sound in water. The result they obtained was within about 2% of currently accepted values. In 1877, Lord Rayleigh wrote the *Theory of Sound* and established the basis of modern acoustic theory.

The sinking of the Titanic in 1912 and the start of World War I provided the impetus for the next wave of progress in underwater acoustics. On 14 April, 1912, the huge vessel, the Titanic, sank in the Atlantic ocean at 42°N, 49°E, with the loss of 1,513 of the 2,208 people on board. This disaster helped to advance the study of echo sound sonar.

Anti-submarine listening systems were developed. Between 1912 and 1914, a number of echolocation patents were granted in Europe and the U.S.A., culminating in Reginald A. Fessenden's echo-ranger in 1914. Pioneering work was carried out during this time in France by Paul Langevin and in Britain by A. B. Wood and his associates. The development of both active ASDIC (anti-submarine division-ics) and passive sonar proceeded during the war, driven by the first large-scale deployments of submarines. Other advances in underwater acoustics included the development of acoustic mines. It is estimated that during World War I, 3,000 warships had somes types of sonar installed.

In 1919, the first scientific paper on underwater acoustics was published[6], theoretically describing the refraction of sound rays produced by temperature and salinity gradients in the ocean. The range predictions of the paper were experimentally validated by transmission loss measurements. The next two decades saw the development of several applications of underwater acoustics. The fathometer, or depth sounder, was developed commercially during the 1920s. Originally, natural materials were used for the transducers, but by the 1930s sonar systems incorporating piezoelectric transducers made from synthetic materials were being used for passive listening systems and for active echo-ranging systems. These systems were used to good effect during World War II by both submarines and anti-submarine vessels. Many advances in underwater acoustics were made which were summarised later by Bergmann[1].

From the end of World War I to the eve of World War II, the practical applications of underwater acoustics slowly but steadily developed. From the scientific point of view, the most important achievement in this period was the understanding and knowledge gained about underwater sound propagation in the ocean. The impetus provided by World War II considerably advanced the development of underwater acoustics and sonar design. Because the participant countries wanted to establish their naval power, their scientists and engineers paid most attention to the study of underwater acoustics. Today, almost all important concepts in underwater acoustics can be traced back to scientific discoveries made during World War II, including transmission loss, absorption, reverberation and target strength, etc.

After World War II, the development of sonar systems was driven largely by the cold war, resulting in advances in the theoretical and practical understanding of underwater acoustics, aided by computer-based techniques. New techniques and equipments were introduced in the military sonar area, such as towed array sonar, flank array sonar, variable depth sonar, conformal array sonar and imaging

sonar. In the civil application area, sonar systems were also of interest to many scientists and engineers because of applications in underwater gas and oil exploration, marine disaster rescue and recovery, underwater archaeology, etc.

As a synthetic electronic system, the development of a sonar system is affected and limited by many related sciences, including electrical engineering, ocean engineering, radio communication, material sciences, microprocessor technology, and signal processing theory. Up to the 1960s, the technique adopted in sonar systems was analogue signal processing. The components used in sonar hardware were electronic tubes or transistors. Since the use of analogue signal processing is limited by various conditions, including the frequency bandwidth of the signal, data transmission speed and memory capacity, the performance of such systems remains at a relatively low level. Nevertheless, the basic framework of sonar signal processing was already established during this period, and had become an independent field in underwater acoustics. The research field of statistical detection theory of underwater acoustics and the basic model of ambient noise in isolated and homogeneous media had been established.

In an overview paper at the 75th meeting of the American Society of Acoustics, Goodman[17] listed many milestone events in the development of underwater acoustics, including the discovery of the SOFAR (sound fixing and ranging) channel, sea surface scattering, DIMUS (digital multibeam steering) technology, parametric arrays, mode theory, matched field processing, wave number integration, time reversal mirror, internal waves, soliton in shallow water, SAS (synthetic aperture sonar) and the effects of sound on the marine environment. These developments and achievements in research work on underwater acoustics and ocean engineering play a very important role in modern underwater acoustic theory and practice, but there is still much to be done.

In the 1960s, sonar designers began to test the application of digital signal processing. The pioneering studies were carried out by V. C. Anderson at the Scripps Institute at the University of San Diego. Although the very preliminary work involved only a 1-bit system, it was the beginning of the digitized procedures of modern sonar. The 1-bit system included DELTIC (delay line time compressor), DIMUS and thereafter DICANNE (digital interference canceling adaptive network nulling equipment). In some ways, the application of digital techniques in sonar is easier than in radar. Because the signal frequency used in a sonar system is much lower than that used in radar, the hardware realization is much easier. The rapid growth of digital sonar techniques is also benefitting from the development of LSIC large scale integration and VLSI very largescale integration. Many specific digital signal processing chips have been launched which can meet a range of application requirements.

In the development of digital sonar, the invention of the FFT (fast Fourier transform) algorithm and progress in the theory of adaptive filtering, time delay estimation, Wiener filtering theory and Kalman filtering algorithm, optimum array processing theory and other digital signal processing methods played a very important role. The digitization of sonar signal processing resulted in substantial changes in sonar systems. A digital signal is more convenient for transmission,

storage, and processing. Therefore, it provides the possibility to design new concept sonar systems with friendlier human/machine interfaces and better performance. Also, the extensive application of digital signal processing technique in sonar systems considerably advanced the development of digital signal processing theory itself, microcomputer design, and even the semiconductor industry[37-38, 46-47].

Clearly, the digitized procedure is just one step in the ongoing evolution of sonar from analogue to analogue/digital hybrid and then to pure digital. Digital sonar is by no means the final phase of sonar design. With the development of signal detection theory and research work in underwater acoustics, and a better understanding of the ocean environment and underwater acoustic channels, the nature of sonar signal processing will continue to change.

1.2 Main Features of Digital Sonar

A sonar system is one of the most important items of naval equipment. In fact, a modern sonar system is really a multi-function computer. Well designed, high performance sonar systems have a very important role in anti-submarine warfare. Sonar technology is a combined product of underwater acoustic physics, underwater acoustic engineering, electrical engineering and transducer manufacture.

A definition of sonar is: the method or equipment which detects the existence of, locates the position of, and identifies an underwater object using underwater acoustic waves. In a broad sense, any system that uses underwater acoustic energy to observe or to communicate is called a sonar system.

An obvious question is: why do we use an acoustic wave instead of an electromagnetic or optical wave? Sea water is a good conductor of electromagnetic and optical waves, the electrical and optical energy is soon dissipated in the form of heat. At the same frequency, the attenuation of an electromagnetic or optical wave is much greater than that of an acoustic wave, and therefore the propagation distance is much smaller. For example, at 1,000 Hz, an acoustic wave, regardless of the expansion loss, has an absorption attenuation loss of about 0.67 dB/km. The absorption for an electromagnetic wave is very strong: even for a green (wavelength 400 nm) or blue (wavelength 500 nm) laser in so called "transmitting windows" the absorption attenuation is more than 100 dB/km. Therefore, it is easy to understand why we use underwater acoustic waves in sonar systems for both military and civil applications.

As an example, Fig. 1.2 illustrates the basic configuration of a sonar system in a surface ship and a submarine. For the surface ship and submarine, the wet end of the bow array sonar system is installed under the bow of vessel. The dry end of the sonar system is installed in the cabin of the vessel. The wet ends of the flank array sonar and passive raging sonar are mounted on both sides of the submarine. The wet end of the towed array sonar is usually installed on the rear deck of a surface ship, and in a special winch room in a submarine.

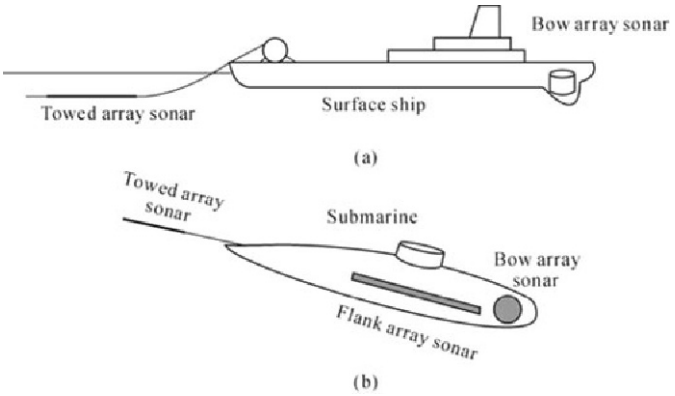


Fig. 1.2. Basic configuration of a sonar system.(a) Surface ship; (b) Submarine

Because sonar is used for different purposes, the function and performance of each sonar system is different. Even in the same submarine, a different installation position will result in a different configuration of sonar components. Therefore, if we attempt to classify the sonar based on its performance, it will have many different classes. Based on the vessel (the platform of installation of sonar), there is surface ship sonar, submarine sonar, shore-based sonar, helicopter decoy sonar and sonobuoy, etc. Based on the sonar function, there is integrated sonar, active sonar, passive sonar, communication sonar, and interception sonar, etc. Based on the installation position of the wet end, there are conformal sonar, flank array sonar, cylindrical array sonar, variable depth sonar, towed array sonar and shore-based sonar, etc. But all sonar systems have some common characteristics, for example, each sonar system has the basic architecture–the wet end and the dry end. The differences are in the configuration and the complexity, enabling it to operate with different missions and to complete different tasks.

Fig. 1.3 illustrates the sonar systems and sensors which are installed in a modern submarine. Almost all the hull is installed with different kinds of sensors and sonar arrays, including bow cylindrical array, flank array, distributed passive ranging array, mine avoidance sonar array, and interception sonar array. There are also many sensors used to perform auxiliary functions, such as depth sensors and self noise monitoring sensors.

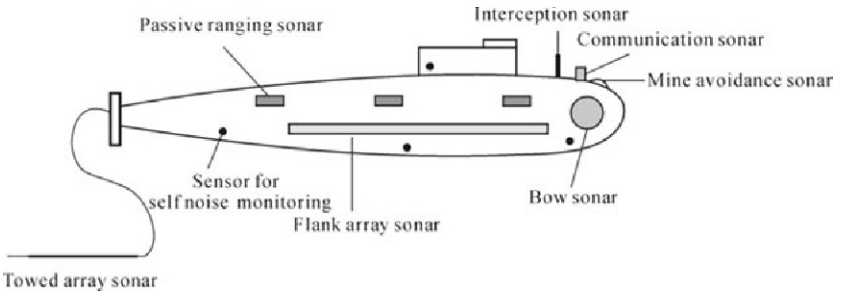


Fig. 1.3. Configuration of various sonar sensors / arrays in a submarine

Fig. 1.4 illustrates the active/passive towed array sonar of a surface ship. The wet end of the towed array sonar consists of tow cable (usually including light and heavy cable of different densities), winch, hydraulic drive/control and underwater acoustic module, VIM (vibration isolated module), and equipment module. The dry end consists of transmitter, signal processor and consol/control units.

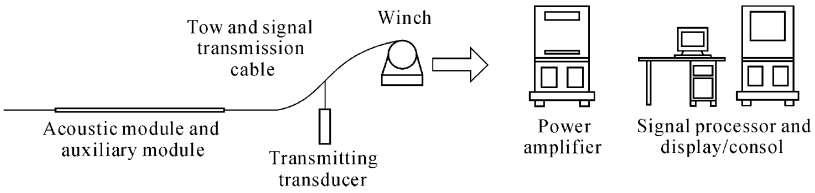


Fig. 1.4. Configuration of passive/active towed array sonar

The signal processing system of a modern sonar system is usually very complicated. To perform a wide variety of missions, it is necessary to have strong hardware support. Fortunately, developments in modern microelectronics are able to provide this kind of support. Here is an example of a signal processing system for modern active/passive sonar. Fig. 1.5 shows a typical digital signal processor for an active/passive sonar system. The transmitting signal of active sonar is generated by the specific software and transmitted by a phased control transducer array. The receiving signal of radiated noise or echo is passed through the bandpass filter which suppresses interference noise. The A/D component converts the signal from analogue to digital. In the case of active sonar detection, due to the relative movement of the sonar platform and the target, it is necessary to have an ODN (own Doppler null) circuit to cancel the Doppler frequency shift, following the coherence processing of the echo signal. In the case of passive detection, it is usual to perform a broadband beamforming algorithm and produce a waterfall time/bearing display.

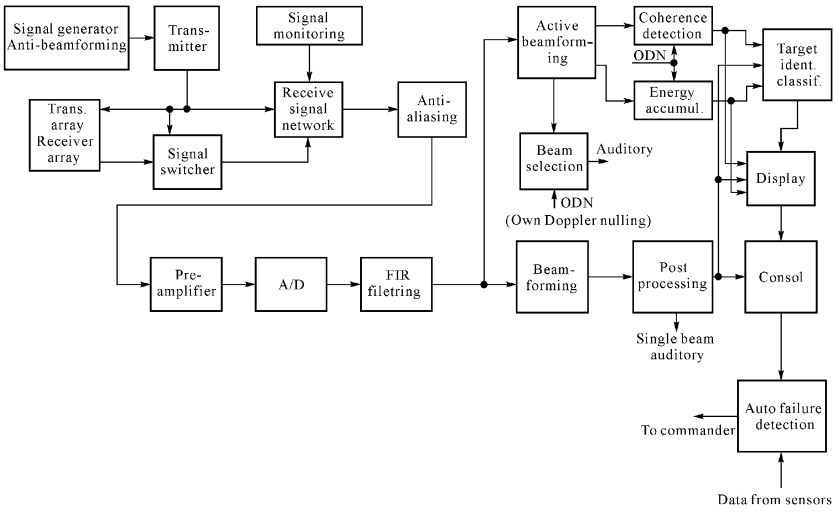


Fig. 1.5. Block diagram of signal processor for typical active/passive sonar system

The auditory signal for a sonar operator plays an important role in signal detection, especially for target identification and classification. It usually comes from a single beam system, which steers in the direction of the target. The single beamforming module is a part of the sonar system. The beam number is often much greater than that of a multibeam system, to obtain a finer beam direction that gives good listening quality. In addition to the main system of the signal processor, the system also has the ability to provide the necessary auxiliary interface and hardware ports to command and vessel control facilities. The parameters from/to other facilities include log and vessel maneuver.

For a complete sonar system it is also necessary to have a system test function, including built-in real time test and offline test.

Digital sonar processing has the following advantages compared with analogue processing:

(1) Digital processing of signal avoids some types of errors that are common in analogue processing. It restricts the phase shift error and channel uncertainty to the preliminary stage in the sonar system. These kinds of errors result in the degradation of the precision of beamforming. So we can focus our attention in the preliminary stages on designing a specific circuit to reduce the effect of the error. In the digital sonar system, before the A/D converter the signal is still in analogue form. After the A/D converter, all the information processing is in digital form. Of course, digital processing also has some kinds of bias or errors, but it is usually easy to predict and estimate, and therefore is easy to prevent. The main errors in an analogue channel are the phase shift within the bandpass, non-uniform multi-channel filtering, and differences in the multi-channel preamplifier. All these kinds of errors will eventually result in a beamforming error, so we have to make the errors as small as possible. After the A/D converter, since the implementation of the FIR (finite impulse response) filter, there is no additional phase shift between channels. In the beamforming, as the time delay between channels depends only on the sampling frequency, the error obeys a certain probability distribution when the number of channels is large enough, so that the effect of time delay on the beamforming (such as the direction of maximum response, or bearing accuracy) is predictable.

(2) A digital system can store necessary data and characteristic parameters in the sonar signal processing procedure for later output, if required. This will be helpful for designing a sonar with a more friendly and comprehensive human-machine interface. The sonar operator will have more opportunities to contribute to parameter selection, performance optimization and auxiliary decision-making.

(3) With the rapid development of the microelectronics industry, DSP chips are changing the system architecture of modern sonar. In particular, COTS (commercial off the shelf) technology has already become an important feature of modern sonar. With the development of COTS products in civilian areas, new products are available with better performance, higher reliability and low cost. Almost all of the products used in military sonar can be found on the commercial market except a few components which need to be specially designed. Therefore, sonar designers should focus their attention on software

development and system integration.

(4) With digital sonar it is easy to realize modules, series, and standard designs. Therefore, it is possible to design a single general purpose sonar signal processing platform which can realize different technical requirements. This kind of platform, a so-called hardware / software configurable system, can considerably reduce the design period and provide a wide variety of channels to improve sonar performance.

(5) The application of digital technology has an advantage in that it can improve system reliability and maintainability. Clearly, as an important piece of military equipment, the reliability and maintainability of a sonar system is very important. As well as having highly reliable components, the system architecture of digital sonar provides an opportunity to increase the reliability of the whole system. This opportunity arises from the block building form of extendable hardware / software architecture. The use of large quantities of the same DSP chips reduces the number of printed circuit boards with different designs, making maintenance easier.

(6) It is possible to realize an online fault test with the use of programmable DSP chips. Therefore, fault diagnosis can be carried out in real time and the position of failures can be located precisely. Today, in a digital sonar system, an online fault test can locate the failure level to a printed board or even a lower level. An early warning of a fault in the main DSP chips is possible.

(7) It is more convenient to set up various interfaces for the peripheral equipment of a digital sonar system. In the sonar operation, it is necessary to communicate data and parameters, record ambient noise and signal in real time. For a digital system, it is easier to storage, record and communicate data when some standards were carried out.

(8) Digital sonar systems are benefiting from progress in digital theory. It is easier to realize a new algorithm in digital form. New developments in sonar technology need the compact combination of signal processing and underwater acoustics. New techniques and algorithms are needed in concept innovation to meet new requirements in ocean engineering such as matched filtering processing, synthetic aperture sonar, adaptive filtering, target identification / classification, and trace extraction of underwater fast-moving targets. It is necessary to have complex algorithms that require multi-channel time / space processing. Only the powerful high speed DSP chip can support such techniques.

The features described above, enable digital sonar to complete many new missions that traditional analogue sonar cannot complete. The performance of digital sonar is developing to be more intelligent, but like everything else in the world, the digital processing of information by no means perfect, and there are also some problems in digital processing itself. For example, digitization of an analogue signal will introduce the “quantity noise”, sampling (quantity in time domain) will yield frequency distortion, and so on. We will discuss these problems later. In summary, digitized processing in modern sonar is a part of a tide of digitization of information data processing, one link in the whole chain of information theory development, and is not a beginning or end point.

1.3 Today and Tomorrow of Digital Sonar

The development of digital sonar started at the end of the 1950s and the beginning of the 1960s, but the digitized processing of signal can be traced back to the 1940s. Shannon introduced the concept of entropy to information theory[34]. This important concept originated in thermo-mechanics. Shannon defined the bit as a basic measurement of uncertainty and his paper is referred to as the basic work in modern information theory. N. Negroponte from MIT, called the bit the "DNA in information theory"[36].

In 1946, when computer experts like Von Neumann developed the first computer in the world, the ENIAC (electronic numerical integrator and calculator), the total weight of a computer was about 30 tons. In 1976, Intel introduced an 8-bit 8048 microprocessor with a similar performance but which weighed only 30 g (Table 1.1).

Table 1.1 Comparisons of ENIAC and Intel 8048

Parameter	ENIAC	Intel 8048
Volume	8.5×10^{7} m^3	6.9 cm^3
Weight	2.7×10^{3} kg	30 g
Power consumption	140 kW	1.5 W
ROM	16 kb	8 kb
RAM	1 kb	0.5 kb
Addition	200 μs	200 μs
Electronic tube / transistor	18,000 electronic tubes	5,000 transistors

Microprocessors can be categorized according to the number of transistors per circuit (Table 1.2).

Table 1.2 Category of integrated circuit

Abbreviation	Name	Number of transistors
SSI	Small scale	1 – 100
MSI	Middle scale	100 – 1,000
LSI	Large scale	1,000 – 10,000
VLSI	Very large scale	≥ 10,000

Today, all microprocessors used in digital sonar systems are VLSI chips. The development of the micro-industry obeys "Moore's law". This law describes an important rule in the history of computer hardware development. Since the invention of the integrated circuit in 1958, the number of transistors that can be inexpensively placed on an integrated circuit has increased exponentially, doubling about every two years. The trend was first observed in 1965 by the co-founder of Intel Corp., Gordon E. Moore[50]. It has continued for half a century and is not expected to stop for another decade at least and perhaps much longer. Moore said: "The complexity for minimum component costs has increased at a rate of roughly a factor of two per year... Certainly over the short term this rate can be expected to

continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer." Fig. 1.6 shows Moore's formal figure in his paper.

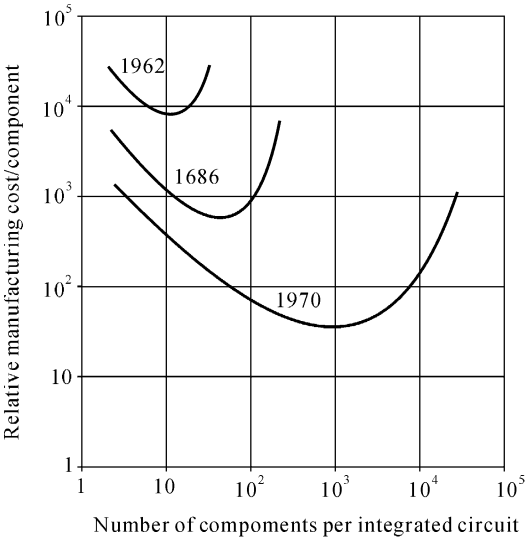


Fig. 1.6. Moore's published figure

The term "Moore's law" was coined around 1970 by Carver Mead, a professor in Caltech, VLSI pioneer. Almost every measure of the capabilities of digital electronic devices is linked to Moore's law: processing speed, memory capacity, even the resolution of digital cameras. All of these are improving at (roughly) exponential rates. In 1975, Moore altered his projection to a doubling every two years.

According to the Moore's law, in a semi-log plot, the relationship between the performance of a microprocessor and time should be a straight line (Fig. 1.7).

The unusually rapid speed of development in the micro-industry is quickly changing the face of many economic areas, including aerospace, ocean engineering, consumer electronics, manufacturing, and office management.

The DSP chip is a specific and important member of the microprocessor family. It is a signal-oriented microprocessor. Usually, microcomputers are extremely capable in two broad areas: (a) data manipulation, such as word processing and database management, and (b) mathematical calculation, used in science, engineering, and DSP. All microprocessors can perform both tasks but it is difficult (expensive) to make a device that is optimized for both. Fig. 1.8 illustrates the difference between the traditional architecture of a computer (the

so-called Von Neumann construction) and the Harvard construction, which is extensively used in the DSP chip. In the Von Neumann construction, the data bus has a bottleneck. This is not found in the Harvard construction because of its distributed data memory and algorithm unit.

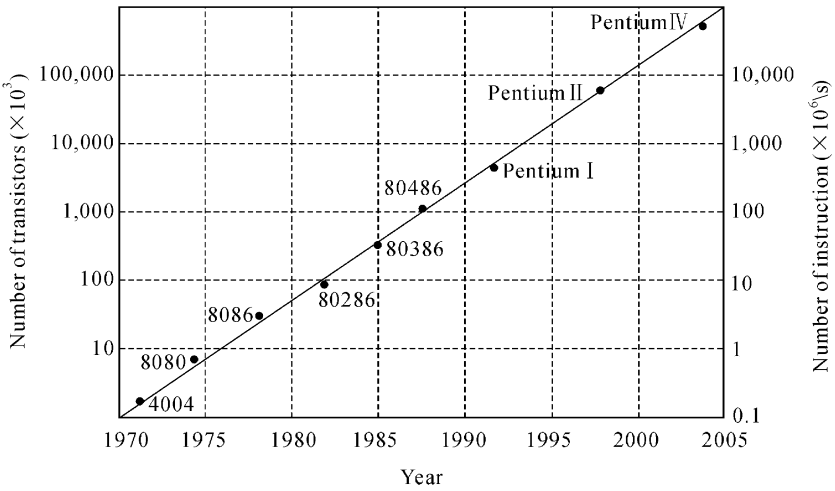


Fig. 1.7. Microcomputer of Intel Corp.

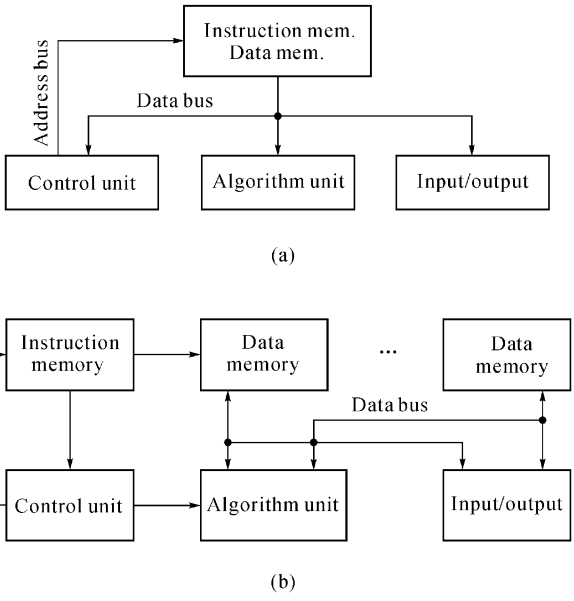


Fig. 1.8. The V. Neumann construction and Harvard construction of microcomputer. (a) V. Neumann construction; (b) Harvard construction

In 1978, Intel released the 2920 as an "analog signal processor". It had an on-chip ADC / DAC with an internal signal processor, but it did not have a hardware multiplier and was not successful in the market. In 1979, AMI released the S2811. It was designed as a microprocessor peripheral and had to be initialized by the host. The S2811 was likewise not successful in the market. In 1979, Bell Labs introduced the first single chip DSP, the Mac 4 microprocessor. Then, in 1980, the first successful stand-alone, complete DSPs (the NEC μPD7720 and AT&T DSP1, were presented at the IEEE International Solid-State Circuits Conference'80. The first DSP produced by Texas Instruments (TI), the TMS32010, presented in 1983, proved to be an even bigger success. It was based on the Harvard architecture, and so had separate instruction and data memory. It already had a special instruction set, with instructions like load-and-accumulate or multiply-and-accumulate. It could work on 16-bit numbers and needed 390 ns for a multiply-add operation. TI is now the market leader in general-purpose DSPs. Another successful design was the Motorola 56000. The second generation of DSPs had three memories for storing two operands simultaneously and included hardware to accelerate tight loops. They also had an addressing unit capable of loop-addressing. Some of them operated on 24-bit variables and a typical model required only about 21 ns for a Mac (multiply-accumulate) instruction. Members of this generation included, for example, the AT&T DSP16A and the Motorola DSP56001.

The main improvement in the third generation was the appearance of application-specific units and instructions in the data path, or sometimes as co-processors. These units allowed direct hardware acceleration of very specific but complex mathematical problems, like the Fourier-transform or matrix operations. Some chips, like the Motorola MC68356, even included more than one processor core to work in parallel. Other DSPs from 1995 include the TI TMS320C541 and the TMS 320C80. The fourth generation is best characterized by the changes in the instruction set and the instruction encoding. As always, the clock-speeds have increased, a 3-ns Mac now becoming possible. Another major signal processor manufacturer today is Analog Devices. This company provides a broad range of DSPs but its main portfolio is multi-media processors, such as codec, filters and digital-analog converters. Its SHARC-based processors range in performance from 66 MHz / 198 MFLOPS (million floating-point operations per second) to 400 MHz/2400 MFLOPS. Another product of the company is the Blackfin family of embedded digital signal processors, with models ranging from ADSP-BF531 to ADSP-BF536. These processors combine the features of a DSP with those of a general use processor. As a result, these processors can run simple operating systems like Linux, while operating relatively efficiently on real-time data.

The design of a DSP is based on the application requirements. In a digital sonar system, the requirements include operation speed, amount of memory, address access capability and the communication interface. There are already some benchmarks by which a DSP chip can be compared. BDTI (Berkeley Design Technology Inc.) in Oakland, California developed a general examination program,

DSP Kernel Benchmarks, as a DSP chip test. The main items include:

(1) Real block FIR: An FIR filter that operates on a block of real (not complex) data.
(2) Complex block FIR: An FIR filter that operates on a block of complex data.
(3) Real single-sample FIR: An FIR filter that operates on a single sample of real data.
(4) LMS adaptive FIR: Least-mean-square adaptive filters that operate on a single sample of real data.
(5) IIR (infinite impulse response): An IIR filter that operates on a single sample of real data.
(6) Vector dot product: The sum of the pointwise multiplication of two vectors.
(7) Vector add: The pointwise addition of two vectors, producing a third vector.
(8) Vector maximum: Finds the value and location of the maximum value in a vector.
(9) Viterbi decoder: Decodes a block of bits that has been convolutionally encoded.
(10) Control: A sequence of control operations (test, branch, push, pop, and bit manipulation).
(11) 256-point in-place FFT: Fast Fourier transform converts a timedomain signal to the frequency domain.
(12) Bit unpack: Unpacks variable-length data from a bit stream.

For digital sonar, different types of DSP chip of the same generation have the ability to meet the hardware requirements. So it is unnecessary to limit an application to a specific brand of DSP chip. The most important thing is to choose a chip that can satisfy the design specifications[52-54].

In the designing phase of digital sonar, the bus type need not be considered, as there are no reasons to restrict the data interchange at the printed board level. But, the type of data bus is an important factor we have to consider in hardware realization. There are two types of data bus in microprocessor applications: VME (versa module Europe) was introduced by Motorola Inc. in Europe in 1980 and has become the standard data bus of the US military. The other data bus used in sonar signal processing is the PCI (personal computer interface) data bus. This was introduced by a PCI Special Interest Group. Today, VME and PCI (or CPCI (compact PCI)) are extensively used in digital sonar. A comparison of the two types of data bus is shown in Table 1.3.

Table 1.3 Comparison of VME and CPCI

Item	VME	CPCI
Number of data bit	16 / 32 / 64-bit	32 / 64-bit
Transmission rate	80 Mbps	264 Mbps
Extendable ability	21 U	8 / 15 / arbitarary
JTAG simulator	None	Yes

References

[1] Bergamann, P. G.: Physics of Sound in Sea. National Defense Research Committee, US (1946)
[2] Urick, R. J.: Principles of Underwater Sound, 3rd Edn. McGraw-Hill, New York (1975)
[3] Horton, J. W.: Fundamentals of Sonar. US Navy Inst. Press, Annapolis (1959)
[4] Alber, V. O.: Underwater Acoustics Handbook. The Penn. State Univ. Press, Phila (1960)
[5] Brekhovskikh, L. M., Lysanov, Y. P.: Fundamentals of Ocean Acoustics, 2nd Edn. Springer-Verlag, New York (1991)
[6] http://www.wikipedia.com/underwater acoustics. Accessed 1 Jan 2011
[7] Jensen, F. B., Kuperman, W. A., Portor, M. B., Schmidt, H.: Computational Ocean Acoustics. AIP Press, Springer, New York (2000)
[8] Katsnelson, B., Petnikov, V.: Shallow Water Acoustics. Springer, New York (2002)
[9] Urick, R. J.: Sound Propagation in the Sea. Peninsula. Los Altos, CA., (1982)
[10] Tolstoy, I., Clay, C. S.: Ocean Acoustics. AIP Press, New York (1987)
[11] Munk, W., Worcester, P., Wunsch, C.: Ocean Acoustic Tomography. Cambridge University Press, New York (1995)
[12] Kinsler, L. E., Frey, A. R.: Fundamentals of Acoustics, 3rd Edn. Wiley, New York (1982)
[13] Medwin, H.: Fundamentals of Acoustical Oceanography. Academic Press, Boston (1998)
[14] Clay, C. S.: Acoustical Oceanography: Principles and Applications. Wiley, New York (1977)
[15] Tucker, D. G.: Applied Underwater Acoustics, 1st Edn. Oxford, New York (1966)
[16] Crocker, M. J.: Handbook of Acoustics. Wiley, New York (1998)
[17] Goodman, R.: A brief history of underwater acoustics. In: Proceedings of ASA at 75, Anniversary, USA, pp. 204-227 (2004)
[18] Candy, J. V.: Model-based Signal Processing. Wiley, New York (2005)
[19] Nielsen, R. O.: Sonar Signal Processing. Artech house, Norwood (1991)
[20] Etter, P. C.: Underwater Acoustic Modeling: Principles, Techniques and Applications, 2nd Edn. Spon Press, New York (1996)
[21] Cox, A. W.: Sonar and Underwater Sound. Lexington Books, Lexington, Mass. (1974)
[22] Horton, C. W.: Signal Processing of Underwater Acoustic Waves., U.S. Govt. Print. Off. Washington, Supt. of Docs. (1969)
[23] Blanhut, R. E. et al.: Radar and Sonar. Springer, New York (1991)
[24] Kock, W. E.: Radar, Sonar and Holography. Academic Press, Boston (1973)
[25] Introduction to Naval Weapons Engineering. Military Analysis Network. http://www.fas.org/man/dod-101/navy/docs/. Accessed 1 Jan 2011

[26] Jenkins, J. W.: Sonar principles and antisubmarine warfare. Sea Technology 34-2, pp. 61-71 (1993)
[27] Cebrowski, K., Garstka, J. J.: Network centric warfare: its origin and future. In: Naval Institute Proceedings, JCS (1998)
[28] Tobin, P. E.: Year of the ocean: a chance to consider our own role. Sea Technology 39-1, pp. 10-14 (1998)
[29] National Research Council (Eds): Technology for the United States Navy and Marine Corps., 2000-2035, Becoming a 21st-Century force. NA Press, USA (1997)
[30] Sherman, J.: Rumsfeld's new speed goals. Defense News, pp. 1-8 (2004)
[31] Gerken, L.: ASW versus submarine technology battle. American Scientific Corp. CA (1986)
[32] Burns, R. F.: Undersea warfare 1999-2000: Unprecedented times. Sea Technology 40-11, pp. 10-18 (1999)
[33] National Research Council of USA (Eds): Network Centric Naval Forces: A Transition Strategy for Enhancing Operational Capabilities. NA Press, USA (2000)
[34] Shannon, C.: A mathematical theory of communication. Bell System Tech. J. 27-4, pp. 623-656 (1948)
[35] Wiener, N.: Cybernetics, or Control and Communication in the Animal and the Machine, 2nd Edn. MIT Press, London (1962)
[36] Negroponte, N.: Being Digital. Alfred A. Knopf, New York (1995)
[37] Knight, W. C., Pridham, R. G., Kay, S. M.: Digital signal processing for sonar. In: Proc. of IEEE 69, pp. 1451-1506(1981)
[38] Lee, E. A.: Programmable DSPs: A brief overview. IEEE Micro. 10, pp. 24-26 (1990)
[39] Anderson, V. C.: Digital array phasing. J. Acoust. Soc. Amer. 32, pp. 867-870 (1960)
[40] Rudnick, P.: Small signal detection in the DIMUS array. J. Acoust. Soc. Amer. 32, pp. 871-876 (1960)
[41] Anderson, V. C.: DICANNE, a realizable adaptive process. J. Acoust. Soc. Amer. 45-63, pp. 398-405 (1969)
[42] Nichles, J. C.: Programmable DIMUS processor. J. Acoust. Soc. Amer. 51, pp. 134-138 (1972)
[43] Bartram, J. F., Ramseyer, R., Heines, J.: Fifth generation of digital sonar processing. IEEE Trans. OE-2, pp.337-343 (1977)
[44] Steane, A. M., Rieffel, E. G.: Beyond bits: the future of quantum conformation processing. Computer 33-1, pp. 38-45 (2000)
[45] Applewhite, A., Kumagai, J.: Technology trends 2004. IEEE Spectrum, pp. 8-13 (2004)
[46] Alta, R. H. V., Lippitz M. J., Lupo, J. C., Mahoney R., Nunn, J. H.: Transformation and transition: DARPA's role in fostering and emerging revolution in military affairs Vol.1 Overall assessment, IDA, USA (2003)
[47] Jeffrey, W.: DARPA technology transition: 2003. DARPA, USA (2004)
[48] Moore, G. E.: Cramming more components onto integrated circuits. Electronics

Magazine 38-B, pp. 114-117 (1965)
[49] http://wantitnow.ebay.com/ws/ebay. Accessed 1 Jan 2011
[50] http://www.intel.com. Accessed 1 Jan 2011
[51] http://www.bdti.com. Accessed 1 Jan 2011
[52] Price, W. J.: A benchmark tutorial. IEEE Micro. 9-5, pp. 28-43 (1989)
[53] Bier, J.: Selecting the right tools for DSP processor software development. DSP and Multimedia Technology 3-7, pp. 55-62 (1994)
[54] Lapsley, P., Bior, J.: Choosing DSP processor. DSP Applications 3-1, pp. 24-32 (1994)

2

Basic Theory of Digital Signal Processing

Digital signal processing theory and technique, as a very important area in science and engineering, has developed very rapidly in the last thirty years[1-15]. This development has benefited from progress in both computer technology and the micro-industry. In fact, 30 years ago, digital computers and the related hardware were expensive for industrial and commercial applications, and therefore their application was limited to the non realtime calculation and control of science and engineering, but the rapid growth of the micro-industry substantially changed that. The theory of digital signal processing is based on traditional signal processing theory; the early research work began in the late 1940s. Today, the theory and application of digital signal processing is deeply involved in many science and engineering areas, including civil and military fields[16-25].

The basic theory and techniques of digital signal processing are introduced in this chapter. In particular, the most important and widely used theoretical results from digital sonar design[26-41] are described, including digital filtering, random signal correlation, spectrum analysis, FFT (fast Fourier transform) algorithm, two-dimensional image processing and some other topics of current interest in the signal processing field which may have potential applications in digital sonar design. These new concepts include wavelet and fractal transform.

2.1 Digital Conversion of Analogue Signal: Quantization and Sampling

From a mathematical point of view, a one-dimensional analogue signal can be considered as the continue function of time t, in which the value of t can vary within a certain segment. Fig. 2.1 illustrates a record of a time wave in the real world. Fig. 2.1(a) shows the time wave of an adult (male) voice, speaking a consonant after a vowel; Fig. 2.1(b) is an earthquake wave; Fig. 2.1(c) is the wave of a human heart impulse. Although the sources are quite different, they have the same characteristics: that is the value of continuity. Of course the computer can

also generate various signals, from a deterministic signal to a random signal. Fig. 2.2 illustrates a simulated Gaussian random signal. The probability density function is shown in Fig. 2.2(b), indicating that this Gaussian random signal has a zero mean and a variance of 1.

Generally speaking, the digital processing of an analogue signal can be divided into two steps: quantization in time domain (sampling) and quantization in amplitude (layering). After completion of these two steps, the analogue signal becomes a digital signal. Fig. 2.3 shows a block diagram of this process. Before signal sampling, it is very important that the analogue signal is pre-filtered in some way. This kind of filtering is called anti-aliasing filtering; its main function is to avoid spectrum distortion due to sampling operation. This filtering must be done even though it may remove some useful components of the analogue signal. In practice, the best we can do is to minimize the signal distortion. We will explain the need for anti-aliasing filtering in more detail later.

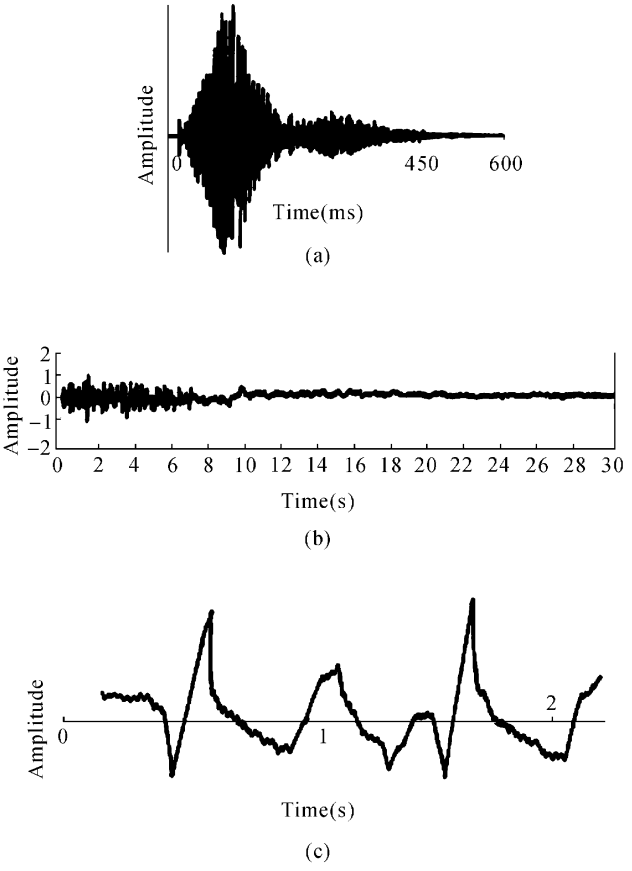


Fig. 2.1. Several wave shapes in time domain. (a) Adult voice; (b) Earthquake wave; (c) Adult heart impulse

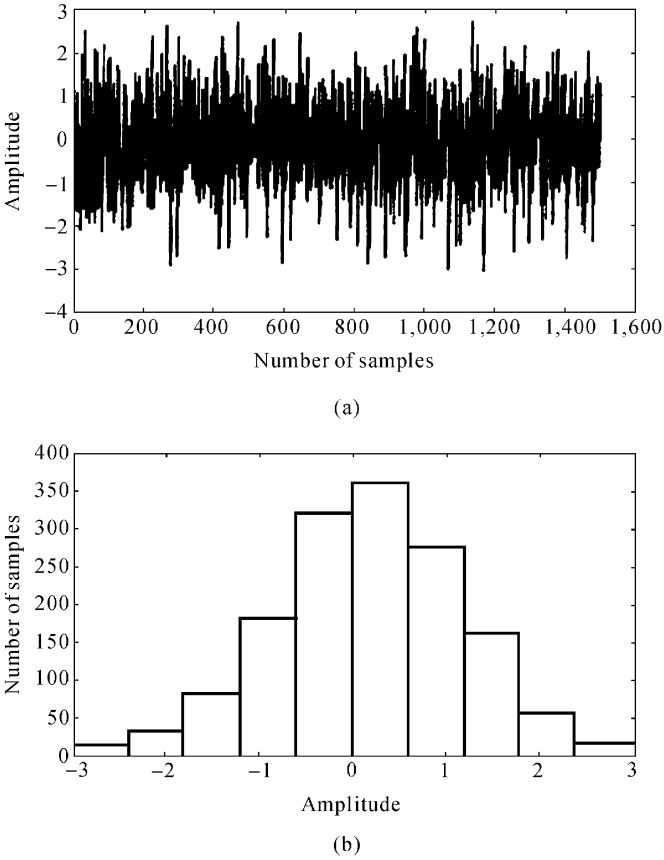


Fig. 2.2. Random Gaussian noise generated by computer. (a) Samples in time domain; (b) Probability distribution

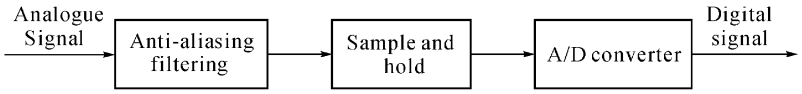


Fig. 2.3. Digitization of an analogue signal

2.1.1 Signal Sampling

First, let us examine signal quantization in time domain sampling. We are interested only in equal time duration sampling. Fig. 2.4 illustrates the example of sampling a sinusoid signal. The signal frequency is f and the sampling duration is T_s. Before sampling, the analogue sinusoid signal is

$$x(t) = A\cos(2\pi ft+\theta) \tag{2.1}$$

where A is the amplitude of the signal and θ represents the random phase. After sampling, the signal becomes a discrete time series

$$x(n) = x(nT_s) = A\cos(2\pi fnT_s + \theta) \tag{2.2}$$

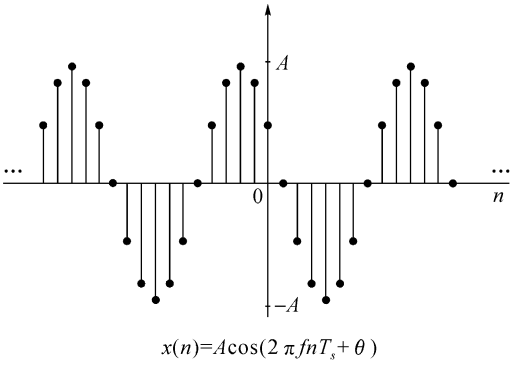


Fig. 2.4. Sampling of a single frequency signal

In the sampling operation, we neglect the signal value between sampling intervals and therefore it is possible that two different analogue signals may correspond to the same discrete time series. Fig. 2.5 shows an example: suppose the analogue signal to be sampled is a sinusoid signal with a frequency $f_1 = \frac{1}{8}$ Hz, and the sampling frequency is f_s = 1 Hz. The sample is shown in the figure in dot sign. But we can see that these samples are also the result of sampling a signal with a frequency $f_2 = -\frac{7}{8}$ Hz. Clearly, to prevent this kind of phenomenon occurring and to ensure that the results from a sampling operation are unique, we need some kinds of rules for the original signal to be sampled and / or the sampling frequency. The following sample theorem solves this problem.

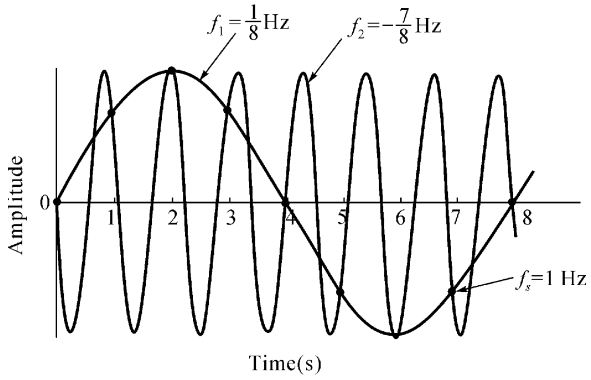


Fig. 2.5. Illustration of aliasing phenomena

2.1.1.1 Sampling Theorem for a Deterministic Signal

Assuming the deterministic signal $x(t)$ has limited bandwidth, i.e., exits a positive value W, such that the spectrum of $x(t)$, $x(f) = 0$ when $|f| \geq W$, then the signal $x(t)$ can be expressed as a sum of discrete samples of $x(kT_s)$, $k = 0, \pm 1, \pm 2, \ldots$, and the time duration of adjacent samples is $T_s = 1/(2W)$

$$x(t) = \sum_{k=-\infty}^{\infty} x(kT_s) \frac{\sin 2\pi W(t - kT_s)}{2\pi W(t - kT_s)} \quad (-\infty < t < \infty) \tag{2.3}$$

This theorem shows that the discrete sample set of $x(kT_s)$ includes all the information of the original signal if the sampling frequency $f_s = 1/T_s$ is equal to or two times greater than the bandwidth. $f_s = 2W$ is called the Nyquist sampling frequency.

To prove the sampling theorem, let us introduce a sampling function where $\delta(\cdot)$ is Dirac δ function. Eq. (2.4) presents an indefinite impulse series with time duration T_s. The original continue signal is $x(t)$. The sampling operation can be considered as the product of $x(t)$ and $s(t)$, i.e.

$$x_s(t) = x(t)s(t) = x(t)\sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{2.4}$$

The spectrum of the sampling function is

$$S(f) = \sum_{k=-\infty}^{\infty} \delta\left(f - k\frac{1}{T_s}\right) \tag{2.5}$$

Therefore, the spectrum $X_s(f)$ of $x_s(t)$ is equal to the convolution of $X(f)$ and $S(f)$, i.e.

$$X_s(f) = X(f) * S(f) \tag{2.6}$$

Here * represents the convolution operation. Fig. 2.6(a) is an explanation of this procedure. We can see that $X_s(f)$ is a periodic function. In each period the shape is the same as $X(f)$. When $x_s(t)$ passes a filter with a rectangular frequency response and bandwidth W, the spectrum of the output $y(t)$ of this filter should have the same spectrum as $x(t)$, that is

$$y(t) = x(t) \tag{2.7}$$

However, the impulse response of the rectangular low pass filter is

$$h(t) = \frac{\sin 2\pi Wt}{2\pi Wt} \tag{2.8}$$

From the convolution theorem of a linear system, we have

$$y(t) = x_s(t) * h(t) = (x(t)\sum_{n=-\infty}^{\infty}\delta(t-nT_s)) * h(t)$$
$$= \sum_{k=-\infty}^{\infty} x(kT_s)\frac{\sin 2\pi W(t-kT_s)}{2\pi W(t-kT_s)} \tag{2.9}$$

This proves the sampling theorem.

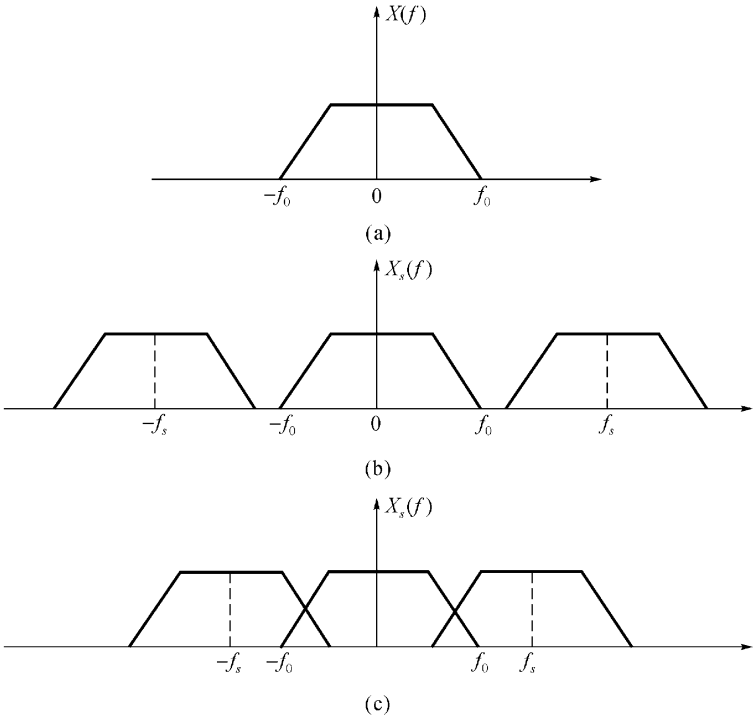


Fig. 2.6. Explanation of sampling theorem in frequency domain. (a) Spectrum of original signal; (b) $f_s \leq 2f_0$, no frequency aliasing; (c) $f_s \geq 2f_0$, aliasing in frequency domain

Fig. 2.6 is the explanation of sampling theorem in frequency domain.

The sampling theorem shows that the analogue signal can be expressed by the discrete sample value of the signal when the sampling frequency is equal to or two times greater than the signal bandwidth. Fig. 2.6(b) clearly shows the role of $f_s \geq 2W$.

If a digital system already has a chosen sampling frequency, then to prevent frequency aliasing phenomena, it is necessary to pre-filter the signal to be sampled. The cut-off frequency of this filter must be equal to or less than half the sampling frequency. This kind of filtering is called anti-aliasing. Without an anti-aliasing filter, the sampling procedure will result in unexpected signal distortion in

frequency domain. Fig. 2.6(c) shows the situation when $f_s < 2W$. The spectrum of the original signal appears to fold at the center of f_s, as a result of some kind of distortion.

Note that $f_s = 2W$ is only the minimum requirement for a digital sonar system. In practice many other factors will impose additional requirements on the sampling frequency f_s. However, it is difficult to make an anti-aliasing filter with a strictly rectangular shape, and some authors suggest that the sampling frequency in digital sonar should be at least equal to $2.66W$[14]. For example, if an underwater acoustic signal is in the 0 – 2,500 Hz frequency band, then the sampling frequency f_s should be at least equal to 2.66×2,500 = 6,650 (Hz).

In addition to being expressed in time domain (Eq. (2.3)), we can also express the sampling theorem in frequency domain, since the deterministic signal and its spectrum are a pair of Fourier transforms. Signal $x(t)$ and its spectrum $X(f)$ are interchangeable.

2.1.1.2 Sampling Theorem for a Deterministic Signal in Frequency Domain

Suppose signal $x(t)$ is time limited, i.e., there exits a positive value of T, such that for $|t| > T, x(t) = 0$, then the spectrum $x(f)$ of $x(t)$ can be expressed by the discrete samples series $X(kf_s)$ of $X(f)$, and the duration of adjacent samples is $f_2 = 1/(2T)$

$$X(f) = \sum_{k=-\infty}^{\infty} X(kf_s) \frac{\sin[2\pi T(f - kf_s)]}{2\pi T(f - kf_s)} \tag{2.10}$$

Eq. (2.10) is similar to Eq. (2.3), the only difference being the interchange of (t, f), (f_s, T_s) and (W, T).

2.1.1.3 Sampling Theorem for Stationary Random Signal

According to the basic mathematical definition of a random process, the direct Fourier transform of a stationary random signal does not exist, but the power density spectrum of such a signal can take a role similar to that of the spectrum of a deterministic signal. The sampling theorem for a stationary random signal has a slightly different form compared with that for a deterministic signal.

Assume $x(t)$ is a stationary random signal and its power density spectrum $K_x(f)$ is band limited, i.e., there exists a positive value of W, such that $K_x(f) = 0$, when $|f| > W$. Then, in the mean square sense, $x(t)$ can be expressed as the sum of discrete samples $x(kT_s)$ and the time duration of adjacent samples is $T_s = 1/(2W)$

$$x(t) = \lim_{N\to\infty} \sum_{k=-N}^{N} x(kT_s) \frac{\sin[2\pi W(t - kT_s)]}{2\pi W(t - kT_s)} \tag{2.12}$$

where symbol lim. expresses the limit in the sense of a mean square, i.e.

$$E\left\{x(t)-\sum_{k=-N}^{N}x(kT_s)\frac{\sin[2\pi W(t-kT_s)]}{2\pi W(t-kT_s)}\right\}^2\to 0,\ \text{when } N\to\infty \tag{2.12}$$

Clearly, the sampling theorem for a random signal is more complicated than that for a deterministic signal. The proof of this theorem has to use the Wiener-Khinchine theorem for the correlation function and the power density spectrum of a stationary random process[15]. The procedure is quite similar to that used in the proof of the sampling theorem for a deterministic signal.

Note that in practice we can take a truncated formula instead of the infinite summation in Eq. (2.11). Because the real observation of a random signal is always time limited, for a truncated random signal, the amplitude spectrum does exist so we can use the sampling theorem of a deterministic signal as an approximation, but we should note that, in this case, the spectrum cannot be band limited[42].

2.1.2 Signal Quantization

The second step in the digital processing of an analogue signal is amplitude quantization, that is signal layering or A / D conversion. Fig. 2.7 shows an example of a 3-bit layer. There are two kinds of layers: Fig. 2.7(a) shows a "0" state; in Fig. 2.7(b), "0" is only a transition state. The difference between these two kinds of layers has no significant effect, especially when the number of layers is large. In this book, we will adopt the second kind of layer.

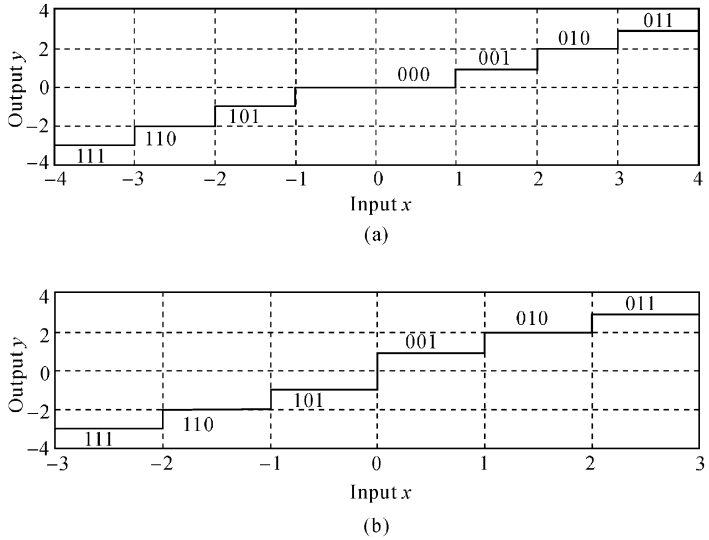


Fig. 2.7. Two types of amplitude quantization. (a) With "0" state; (b) Without "0" state

In general, the first bit of a digital expression is usually a sign bit: 0 indicates positive and 1 indicates negative. 3-bit quantization can express 7 numbers, i.e., 000, 001, 010, 011, 101, 110, 111. For an n-bit quantization, the total number of numbers that can be represented is $2^n - 1$ (including the sign bit). The layering operation that converts an analogue signal to digital introduces a type of error called the "quantization error". If we indicate Δ as the layer step, then the quantization error will fall into the segment of $[0, \Delta]$ or $[-\Delta/2, \Delta/2]$, and the variance of this error will be $\Delta^2/12$. If the power of the input signal is σ^2, then the signal-to-noise ratio of quantization is

$$SQNR = 10\lg\frac{\sigma^2}{\Delta^2/12} \tag{2.13}$$

Eq. (2.13) shows that, for an A/D converter, once the quantization step Δ is defined, then the $SQNR$ is proportional to the signal power σ^2. To enhance the SNR it is necessary to use the dynamic range of the A/D converter as correctly as possible. The selection of the dynamic range of an A/D converter depends on the value of the signal power σ^2.

The value of the signal power σ^2 depends on the class of signal. Generally, in designing A/D converters, it is assumed that the input signal is a Gaussian random signal. Based on the statistical characteristics of a Gaussian signal, we can say that the maximum amplitude of the input signal is $4\sigma^2$, because for a Gaussian signal $x(t)$ with variance σ^2, we have

$$P\{-4\sigma < x(t) - E[x(t)] < 4\sigma\} = 0.997 \tag{2.14}$$

Therefore, we can take it that there is no chance that the absolute value of $x(t) - E[x(t)]$ is greater than 4σ. If the quantization bit number of an A/D converter is n+1 (including the sign bit), then the step size of a layer should be $\Delta = 4\sigma/2^n$ and the $SQNR$ can reach the maximum value. Subsituting this step size into Eq. (2.13), we have

$$SQNR = 6n - 1.24 \tag{2.15}$$

This is a basic formula in digital signal processing, which shows the maximum value of $SQNR$ that can be obtained. For example, assume there is an 11-bit A/D converter, the first bit indicates the sign of the value, so n = 10. If the dynamic range used is satisfactory, then the quantization SNR is $SQNR$ = 60×10 – 1.24 = 58.76 (dB).

In practice, one problem arises: for any well designed A/D converter, the step size of layering is fixed, but the input signal may be small or large. If the amplitude of signal is too big, it will result in non-linear distortion, and if the signal is too small the SNR will decrease. We have to balance these situations. Suppose the signal power of a real signal in the input end of an A/D converter, designed in σ^2, is $\sigma^2 \le \sigma^2$, then the quantization SNR is

$$SQNR = 6n - 1.24 - 10\lg(\sigma^2 / \sigma_0^2) \tag{2.16}$$

where $10\lg(\sigma^2 / \sigma_0^2)$ is the loss arising from the use of an unsatisfactory dynamic range of the A/D converter. In the early stages of digital sonar design, this is a principle that must be taken into account.

For example, assume the sensitivity of a hydrophone is $M = -196$ dB (ref. 1 dB = 1 V/μPa). If the minimum ambient noise level in the range 100 – 10,000 Hz is 0.5×10^5 μPa and the dynamic range of the A/D converter is ±5 V, then how large should be the pre-amplifier? Based on the conditions described above, the ambient noise received by the hydrophone is about $0.5\times10^5\times1.6\times10^{-4}$ μV = 8 μV. If the dynamic range of the A/D converter is ±5 V, then its operating point should be at about 1.2 V. Therefore, the amplification value of the pre-amplifier is 1.2 V/8 μV = 1.5×10^5. It is very important to make this kind of calculation in pre-amplifier design, since it is the basis of subsequent digital signal processing.

The mathematical expression for the layering operation can be derived as follows. Assume the input of an A/D converter is x and the output is y, and Q is the layer number. For the quantization form of Fig. 2.7(a), we have

$$y = \frac{\Delta}{2}\sum_{k=1}^{Q-1}[\mathrm{sgn}(x - k\Delta) + \mathrm{sgn}(x + k\Delta)] \tag{2.17}$$

For the quantization form of Fig. 2.7(b), we have

$$y = \Delta\{\mathrm{sgn}\, x + \frac{1}{2}\sum_{k=1}^{Q-1}[\mathrm{sgn}(x - k\Delta) + \mathrm{sgn}(x + k\Delta)]\} \tag{2.18}$$

where Δ is the step size of quantization and sgn(·) is the sign function

$$\mathrm{sgn}\, x = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \tag{2.19}$$

2.1.3 Signal Companding

The quantization described above is equal step size layering and the performance of this kind of layer is relatively simple. But in practice, sometimes it is necessary to have a large dynamic range in a fixed quantization bit number. In this situation, the equal step size quantization, i.e., the linear quantization, cannot meet the requirements. The most useful kind of non-linear quantization in digital sonar is so-called "companding". The main feature of this kind of quantization is that it enhances a small signal and compresses a large signal. The data companding

technique is used extensively in communication systems and speech signal processing. Two kinds of companding are used in signal processing practice, i.e., A-law and μ-law[43]. In the μ-law, proposed by Smith in 1957, the input/output relationship is

$$y = \begin{cases} \dfrac{a\lg(1+\mu x/a)}{\lg(1+\mu)} & 0 \le x \le a \\ -\dfrac{a\lg(1-\mu x/a)}{\lg(1+\mu)} & -a \le x \le 0 \end{cases} \tag{2.20}$$

where a is a value defined by the dynamic range of the input signal. The typical value of μ is 255. In the A-law, proposed by Catlermola in 1962, the input/output relationship is

$$y = \begin{cases} \dfrac{Ax}{1+\lg A} & 0 \le x \le 1/A \\ \dfrac{1+\lg Ax}{1+\lg A} & 1/A \le x \le 1 \end{cases} \tag{2.21}$$

The typical value of A is 87.5.

Fig. 2.8 illustrates the companding function of the A-law and μ-law. There are only slight differences between these two kinds of companding when $A = 87.5$, $\mu = 255$. An input value near 0 is amplified and a value near 1 is compressed such that the number of quantization bits is reduced considerably. For example, using the A-law, when the $A = 87.5$, the conversion from 12 bits to 7 bits is completed.

$$y = \begin{cases} \dfrac{Ax}{1+\lg A}\times 2^7 & 0 \le x \le \dfrac{2^{12}}{A} \\ \dfrac{1+\lg Ax}{1+\lg A}\times 2^7 & \dfrac{2^{12}}{A} \le x \le 2^{12} \end{cases} \tag{2.22}$$

x varies in $0 - 2^{12}$ and y varies in $0 - 2^7$ (Table 2.1).

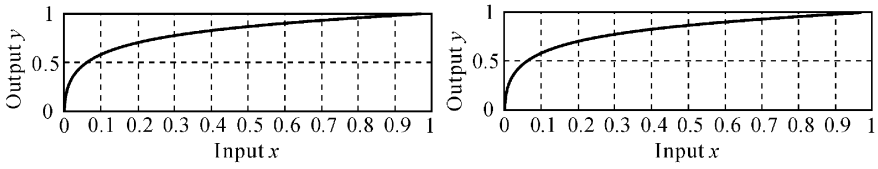


Fig. 2.8. Data companding. (a) μ-law ($\mu = 225$); (b) A-law ($A = 87.5$)

Table 2.1 The corresponding relationship from 12 bits to 7 bits ($A = 87.5$)

x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_A	x_B	x_C	x_D	y_2	y_3	y_4	y_5	y_6	y_7	y_8
0	0	0	0	0	0	0	K	L	M	N	*	0	0	0	K	L	M	N
0	0	0	0	0	0	1	K	L	M	N	*	0	0	1	K	L	M	N
0	0	0	0	0	1	K	L	M	N	*	*	0	1	0	K	L	M	N
0	0	0	0	1	K	L	M	N	*	*	*	0	1	1	K	L	M	N
0	0	0	1	K	L	M	N	*	*	*	*	1	0	0	K	L	M	N
0	0	1	K	L	M	N	*	*	*	*	*	1	0	1	K	L	M	N
0	1	K	L	M	N	*	*	*	*	*	*	1	1	0	K	L	M	N
1	K	L	M	N	*	*	*	*	*	*	*	1	1	1	K	L	M	N

* Letters indicate 0 or 1 digits in both 7 bits and 12 bits numbers, and the symbol indicates a digit that appeared in 12 bits number but which is neglected in the corresponding 7 bits number.

2.1.4 $\Sigma - \Delta$ Modulation

We are going to introduce briefly a special kind of 1 bit quantization technique called $\Sigma - \Delta$ modulation. Although the output data series is the 0 – 1 digit, it has the specifications of a multi-layer system. The outstanding advantage of $\Sigma - \Delta$ modulation is that the quantization noise depends on the sampling frequency. When the sampling frequency is far from Nyquist frequency, the quantization noise is near zero. Fig. 2.9 shows a block diagram of the $\Sigma - \Delta$ modulation technique.

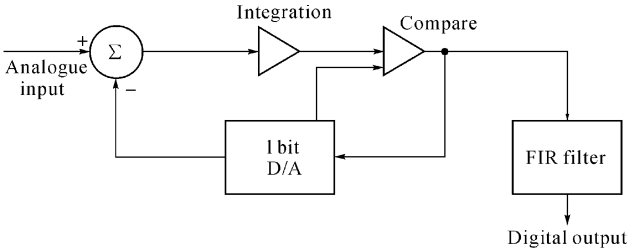


Fig. 2.9. Block diagram of $\Sigma - \Delta$ modulation

The input analogue signal is integrated and compared with the output data of the previous moment. The addition of 1 to the digit of the output signal depends on the result of the comparison (Fig. 2.10).

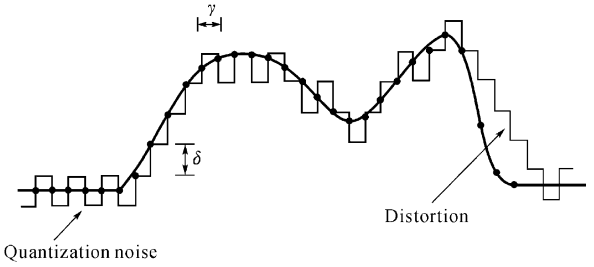


Fig. 2.10. Wave illustration of $\Sigma - \Delta$ modulation

The frequency response of integrator is

$$H(f) = 1/f \tag{2.23}$$

The relationship between input and output data is approximately

$$y = \frac{x - y}{f} + q \tag{2.24}$$

or

$$y = \frac{x}{f+1} + \frac{qf}{f+1} \tag{2.25}$$

At low frequency, $f \approx 0$, and therefore $y \approx x$, and there is almost no quantization noise. But at high frequency, $f \gg 0$, and therefore $y \approx q$, and there is a lot of quantization noise (Fig. 2.11).

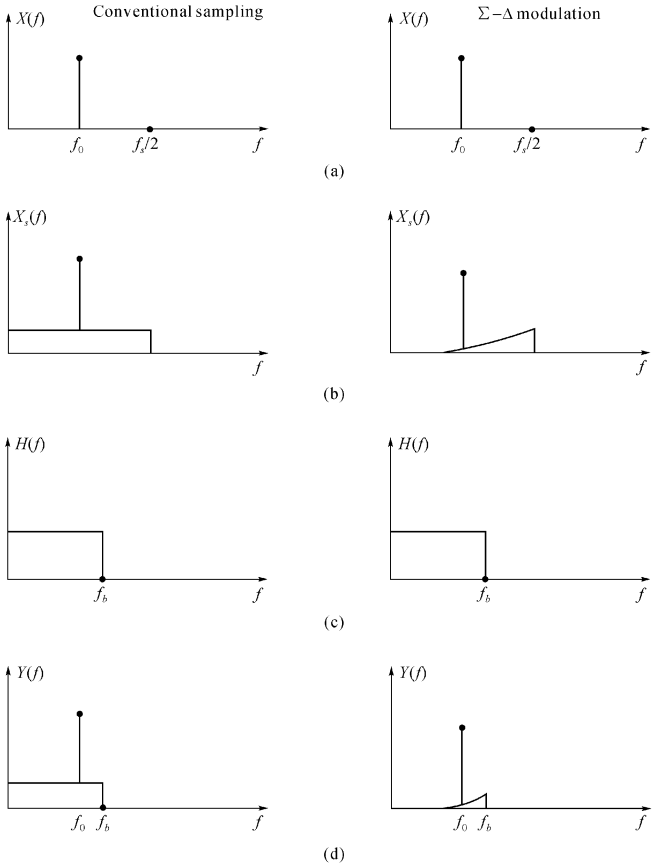


Fig. 2.11. Comparison of $\Sigma - \Delta$ modulation and conventional sampling in frequency domain. (a) Spectrum of original signal; (b) Spectrum after sampling; (c) Frequency response of low pass filter; (d) Spectrum of output signal

2.2 Digital Filtering of Signal

Digital processing of signal includes data storage, transformation, transmission, etc. Digital filtering is one of the most important operations, and linear filtering is the most popular digital filtering technique. It can be considered as a special kind of digital transformation.

2.2.1 Linear Digital Filtering

Fig. 2.12 is a diagram showing linear digital filtering in which, after a certain operation / transform, the input data $x(n)$ becomes the output data $y(n)$.

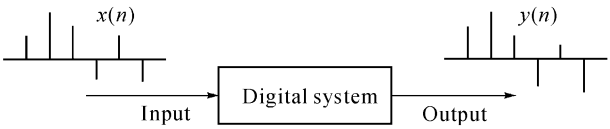


Fig. 2.12. Input / output of a digital system

In this example, the time duration of samples remains fixed. However, as $x(n)$ and $y(n)$ are pure data, in a sense, the time duration between data is meaningless. This point will play a part in down-sampling and up-sampling. Fig. 2.13 illustrates some examples of digital signals, where Fig. 2.13(a) is the unit impulse series

$$\delta(n)\begin{cases}1 & n=0\\ 0 & n\neq 0\end{cases} \tag{2.26}$$

and Fig. 2.13(b) is the unit step impulse series

$$u(n)=\begin{cases}1 & n\geq 0\\ 0 & n<0\end{cases} \tag{2.27}$$

Clearly, $\delta(n) = u(n) - u(n-1)$. Fig. 2.13(c) expresses a periodic series. The characteristic of a periodic series is that there exists a positive integer N_0, such that $x(n) = x(n+N_0)$. Note that a sampled discrete series of analogue periodic signals may not be periodic. For example, the analogue single frequency signal $x(t) = A\cos(2\pi ft+\theta)$ appears to be a periodic signal with period $T = 1/f$. From the sampled results of this signal with a sampling duration T_s, we get $x(kT_s) = A\cos(2\pi fkT_s+\theta)$ It is well known that fT_s is not a rational number, therefore $x(kT_s)$ is a non-periodic series.

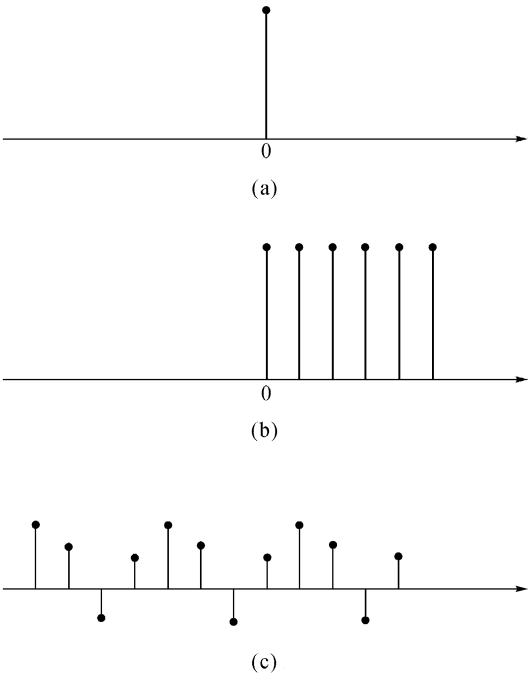


Fig. 2.13. Examples of digital signals. (a) Unit impulse series; (b) Step impulse series; (c) Periodic pulse series

Usually, $x(n)$ is expressed as the input of a digital system and $y(n)$ as the output.

Definition 2.1 *A digital system L is called a linear system, if* $y_1(n) = L[x_1(n)]$, $y_2(n) = L[x_2(n)]$, *then for real values a and b, we have*

$$L[ax_1(n) + bx_2(n)] = ay_1(n) + by_2(n) \tag{2.28}$$

Definition 2.2 *Suppose* $y(n) = L[x(n)]$, *then for any integer* N_0, *we have*

$$L[x(n - N_0)] = y(n - N_0) \tag{2.29}$$

This system is called time invariant.

The time invariant and linear digital system is the most useful digital filter system used in a sonar system. In this book, for simplification, unless otherwise specified, only this system is considered.

Similar to an analogue filter, a digital filter is uniquely defined by the unit impulse response. If the system input is $\delta(n)$ and the output is $h(n)$, then $h(n) = L[\delta(n)]$. Then, for any input $x(n)$, the output

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) \tag{2.30}$$

The proof of this equation can be obtained from the expression of

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)\delta(n-k) \tag{2.31}$$

By using the linear, invariant operator L on both sides, from Eqs. (2.28) and (2.29), we have

$$\begin{aligned} y(n) &= L[x(n)] = L[\sum_{k=-\infty}^{\infty} x(k)\delta(n-k)] = \sum_{k=-\infty}^{\infty} L[x(k)\delta(n-k)] \\ &= \sum_{k=-\infty}^{\infty} x(k)L[\delta(n-k)] = \sum_{k=-\infty}^{\infty} x(k)h(n-k) \end{aligned} \tag{2.32}$$

The relationship between $x(n)$ and $h(n)$ expressed in Eq. (2.30) is called discrete convolution, denoted as $y(n) = x(n) * h(n) = h(n) * x(n)$ for simplification.

As an example, assume $h(n) = (1/2)^n u(n)$, and

$$x(n) = u(n) - u(n-4) = \begin{cases} 1 & 0 \le n < 4 \\ 0 & \text{otherwise} \end{cases}$$

We want to calculate the value of $y(2)$, based on Eq. (2.30)

$$y(2) = x(2)h(0) + x(1)h(1) + x(0)h(2) = 1 + \frac{1}{2} + \frac{1}{4} = \frac{7}{4}$$

2.2.2 Transfer Function of a Linear System

It is well known that the transfer function (or frequency response) of a linear system, which is the Fourier transform of the system impulse response, plays a very important role in an analogue system. For the discrete system, a similar concept can be introduced, that is the z transform. The z transform of a discrete series $h(n)$ is defined as

$$H(z) = \sum_{n=-\infty}^{\infty} z^{-n} h(n) \tag{2.33}$$

If $h(n)$ is the system response of the unit impulse function, then its z transform is called the system function. An important fact is that $h(z)$ is actually the coefficient of the term z^n in the system output when the input signal is $x(n) = z^n$. This is because from Eq. (2.30), if $x(n) = z^n$, then

$$y(n) = \sum_{k=-\infty}^{\infty} z^{n-k} h(k) = z^n H(z)$$

However, if the z transform is taken into both sides of Eq. (2.30), we have

$$Y(z) = H(z)X(z) \quad \text{or} \quad H(z) = Y(z) / X(z) \tag{2.34}$$

Therefore, we have three equivalent definitions for the transfer function $H(z)$:
(1) $H(z)$ is the z transform of the unit impulse response function;
(2) $H(z)$ is the coefficient of the output term z^n when the input is z^n;
(3) $H(z)$ is the ratio of the z transforms for output and input discrete series.

In general, in the z transform expression, the argument z is a complex variable; therefore in discussion of function $H(z)$, the convergent zone is quite important. The following are some examples of z transforms for special series.

Finite Impulse Series

Suppose $n_1 < n_2$, for $n < n_1$ and $n > n_2$, $x(n) = 0$, we have

$$X(z) = \sum_{n=n_1}^{n_2} x(n) z^{-n}$$

If $n_1 < 0,\ n_2 < 0,$ then the convergent zone is $0 \le |z| < \infty$.
If $n_1 < 0,\ n_2 > 0,$ then the convergent zone is $0 < |z| < \infty$.
If $n_1 > 0,\ n_2 > 0,$ then the convergent zone is $0 < |z| < \infty$.

Right Series / Casual Series

Suppose $x(n) = 0,$ for $n < n_1$, then

$$X(z) = \sum_{n=n_1}^{\infty} x(n) z^{-n}$$

If r_1 is the minimum positive value such that

$$\sum_{n=n_1}^{\infty} |x(n)| r_1^{-n} < \infty$$

then the convergent zone is $|z| \ge r_1$.

Left Series

Suppose $x(n) = 0,$ for $n > n_1$, then

$$X(z) = \sum_{n=-\infty}^{n_1} x(n)z^{-n}$$

If r_2 is the maximum value such that

$$\sum_{n=-\infty}^{n_1} | x(n)r_2^{-n} < \infty$$

then the convergent zone is $| z | < r_2$.

General Case

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} = \sum_{n=-\infty}^{-1} x(n)z^{-n} + \sum_{n=0}^{\infty} x(n)z^{-n} \tag{2.35}$$

This is the summation of z transforms of two one-sided series: if the convergent zone of the right series is $| z | \geq r_1$ and the convergent zone of the left series is $| z | < r_2$ and $r_1 < r_2$, then the convergent zone of $X(z)$ is a ring in the z plane (Fig. 2.14). Otherwise, if $r_1 \geq r_2$, then $X(z)$ is divergent in the z plane.

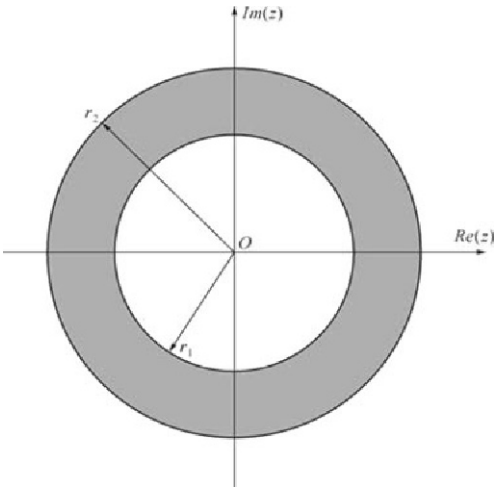


Fig. 2.14. Convergent zone of $X(z)$

2.2.3 Classification of Digital Filters

Similar to an analogue filter, a digital filter can be classified based on its application

field, for example, casual / non-casual, memory / non-memory, etc.

A digital filter has its own special characteristics, since it is uniquely defined by the convolution Eq. (2.30). From this point of view, it is a special operation in the discrete series. The operation procedure is tightly correlated to the response function $h(n)$ of the unit impulse. Some of the basic definitions and theorems for digital filters are described below.

Definition 2.3 *If the length of the response function $h(n)$ of a unit impulse is limited, i.e., the non-zero value of $h(n)$ is finite, then the filter is called an FIR system. If the non-zero value of $h(n)$ is infinite, the filter is called an IIR system.*

Definition 2.4 *If the output $y(n)$ of the n-th moment depends only on the input value and does not depend on the previous outputs before the n-th moment, then the filter is called non-recursive, and for this system*

$$y(n) = \sum_{k=0}^{n} a_k x(n-k) \tag{2.36}$$

If the output $y(n)$ of the n-th moment depends not only on the input value of the n-th moment, but also on the output value before the n-th moment, then the system is called recursive, and for this system

$$y(n) = \sum_{k=0}^{n} a_k x(n-k) - \sum_{k=1}^{n} b_k y(n-k) \tag{2.37}$$

Fig. 2.15 shows an example of a recursive and a non-recursive filter. Note that the recursive filter is actually a system with feedback operation.

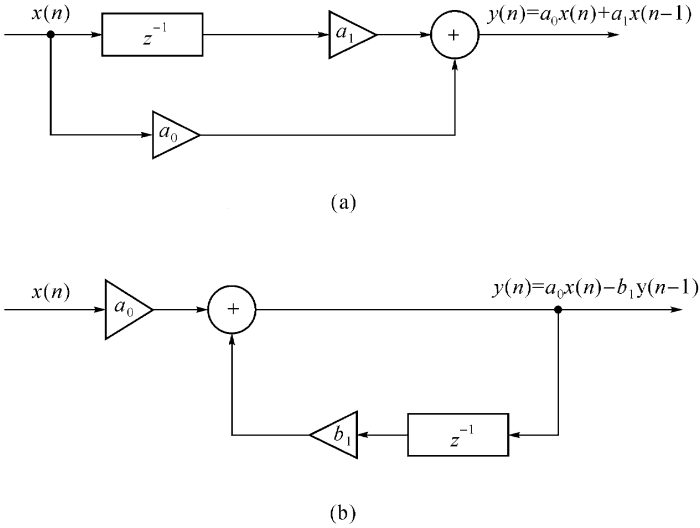


Fig. 2.15. Examples of recursive and non-recursive system. (a) Non-recursive $H(z) = a_0+a_1z^{-1}$; (b) Recursive $H(z) = a_0 / (1-b_1z^{-1})$

Definition 2.5 *If for any boundary input x(n), the output y(n) is boundary, the system is called reliable.*

In the design of a digital filter, being "reliable" is one of the most important requirements. If a boundary input series results in an un-boundary output, it means the system cannot work normally. Here is a theorem for identifying if a digital filter is reliable.

Theorem 2.1 (Reliability) *Suppose the system response function of a unit impulse is h(n), the system is reliable, if and only if*

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty \tag{2.38}$$

Proof:

(1) To prove the necessity, i.e., if the system is reliable, it should be

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty$$

Otherwise

$$\sum_{n=-\infty}^{\infty} |h(n)| \to \infty$$

Consider a boundary series

$$x(n) = \begin{cases} 1 & h(-n) \geq 0 \\ -1 & h(-n) < 0 \end{cases}$$

From Eq. (2.30)

$$y(0) = \sum_{n=-\infty}^{\infty} x(n)h(-n) = \sum_{n=-\infty}^{\infty} |h(-n)| \to \infty$$

That means the output series $y(n)$ is boundless, which is contrary to the reliable assumption.

(2) To prove the sufficiency. If the input $x(n)$ is a boundary series, i.e., there exists as a positive value M such that for each n, $|x(n)| \leq M$, then from Eq. (2.30)

$$|y(n)| = \left|\sum_{k=-\infty}^{\infty} h(k)x(n-k)\right| \leq \sum_{k=-\infty}^{\infty} |h(k)x(n-k)| \leq M\sum_{k=-\infty}^{\infty} |h(k)| < \infty$$

Thus, the series y(n) is boundary.

2.2.4 Cascade of Digital Filters

Now we consider the cascade problem of digital filters. In practice, it is often necessary to combine several digital filters to construct a new digital filter. A large system may be very complicated, but the basic cascade forms of a digital system are of two types, series connection and parallel connection (Fig. 2.16).

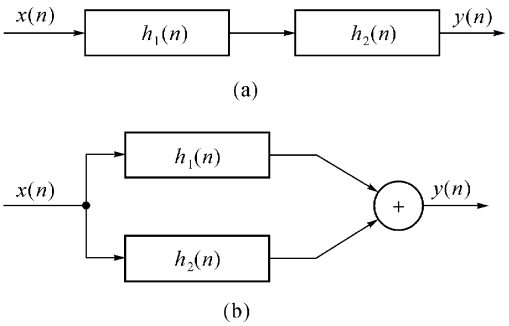


Fig. 2.16. Series and parallel connections of a digital filter. (a) Series connection; (b) Parallel connection

For the series connection (Fig. 2.16(a))

$$y_2(n) = (x(n) * h_1(n)) * h_2(n) \tag{2.39}$$

For the parallel connection (Fig. 2.16 (b))

$$y(n) = x(n) * h_1(n) + x(n) * h_2(n) \tag{2.40}$$

It is easy to prove that, for the series connection

$$\begin{aligned} h(n) &= h_1(n) * h_2(n) \\ H(z) &= H_1(z)H_2(z) \end{aligned} \tag{2.41}$$

and for the parallel connection

$$\begin{aligned} h(n) &= h_1(n) + h_2(n)\ , \\ H(z) &= H_1(z) + H_2(z) \end{aligned} \tag{2.42}$$

The proof of the above equation should follow the exchange law, distribute law and combine law in the convolution operation, which can be directly verified from the definitions of a linear digital system.

Exchange law

$$x(n) * h(n) = h(n) * x(n) \tag{2.43}$$

Combine law

$$(x(n) * h_1(n)) * h_2(n) = x(n) * (h_1(n) * h_2(n)) \tag{2.44}$$

Distribute law

$$x(n) * (h_1(n) + h_2(n)) = x(n) * h_1(n) + x(n) * h_2(n) \tag{2.45}$$

Some time invariant and variant digital systems are described below and will be discussed later.

2.2.5 Examples of Digital Filters

(1) Differential system (Fig. 2.17(a))

$$y(n) = x(n) - x(n-1) \tag{2.46}$$

This is a time invariant filter, since

$$y(n-k) = x(n-k) - x(n-k-1) \tag{2.47}$$

(2) Time multiplication system (Fig. 2.17(b))

$$y(n) = n\,x(n) \tag{2.48}$$

The multiplication value is proportional to the time, therefore this is a time variant system, because the response of $x(n-k)$ is $(n-k)x(n-k)$, but the delay of $y(n)$ is $y(n-k) = n\,x(n-k) \neq (n-k)x(n-k)$.

(3) Fold system (Fig. 2.17(c))

$$y(n) = x(-n) \tag{2.49}$$

This is a time variant system because $y(n-k) = x(-n+k)$, but the time delay of $y(n)$ is $x(-n-k) \neq x(-n+k)$.

(4) Modulation (Fig. 2.17(d))

$$y(n) = x(n)\cos(2\pi f_0 n) \tag{2.50}$$

This is clearly a time variant system.

In practical application, as well as the linear, time invariant system, there are many non-linear, time variant systems.

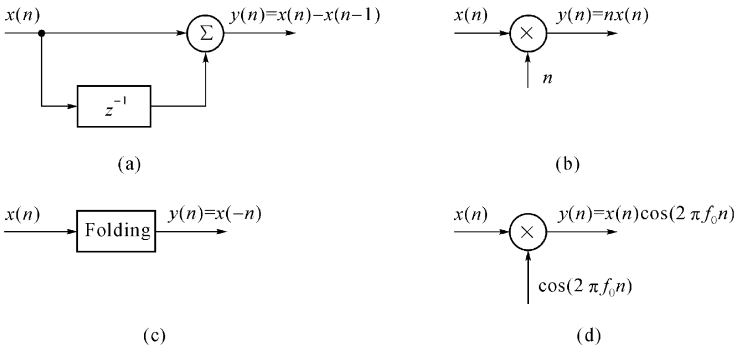


Fig. 2.17. Examples of time-varying and non-time-varying digital filters. (a) Differential system; (b) Time multiplication system; (c) Fold system; (d) Modulation

2.3 Characteristics of Digital Signals in Time Domain and Frequency Domain

The operation and processing of digital signal involves many characteristics of the signal including time domain (such as mean value, signal power, correlation function, etc.) and frequency domain (such as amplitude spectrum, power density spectrum, cross power spectrum, etc.). The study of the basic relationships of these characteristics is very important for signal processing. We begin by considering the basic concepts of analogue signals and then extend into digital signals.

2.3.1 Fourier Transform of Signal

Definition 2.6 *If a deterministic signal* $x(t)$ *satisfies* $\int_{-\infty}^{\infty} |x(t)|\,\mathrm{d}t < \infty$, *then its Fourier transform*

$$X(f) = \int_{-\infty}^{\infty} x(t)\exp(-2\pi \mathrm{j}ft)\mathrm{d}t \tag{2.51}$$

is called the amplitude spectrum of $x(t)$ *and is denoted as*

$$X(f) = F[x(t)] \quad \text{or} \quad x(t) \leftrightarrow X(f) \tag{2.52}$$

Upon some mathematical restriction on $x(t)$, we have

$$x(t) = F^{-1}[X(f)] = \int_{-\infty}^{\infty} X(f)\exp(2\pi \mathrm{j}ft)\mathrm{d}f \tag{2.53}$$

The following are some basic properties of Fourier transform.

(1) Linearity. The spectrum of signal summation is equal to the summation of the signal spectrum

$$F[x(t)+y(t)]=F[x(t)]+F[y(t)] \tag{2.54}$$

(2) Time scale. Suppose $x(t)\leftrightarrow X(f)$ when $c\neq 0$

$$x(ct)\leftrightarrow \frac{1}{|c|}X(f/c) \tag{2.55}$$

(3) Time shift. Suppose $x(t)\leftrightarrow X(f)$, then

$$x(t-t_0)\leftrightarrow X(f)\exp(-2\pi \mathrm{j}ft_0) \tag{2.56}$$

(4) Frequency shift. Suppose $x(t)\leftrightarrow X(f)$, then

$$x(t)\exp(2\pi \mathrm{j}f_0t)\leftrightarrow X(f-f_0) \tag{2.57}$$

(5) Convolution theorem in time domain. Suppose $y(t)=x(t)*h(t)=\int_{-\infty}^{\infty}h(\tau)x(t-\tau)\mathrm{d}\tau$ and $x(t)\leftrightarrow X(f)$, $h(t)\leftrightarrow H(f)$, then

$$y(t)\leftrightarrow Y(f)=H(f)X(f) \tag{2.58}$$

(6) Convolution theorem in frequency domain. Suppose $x(t)\leftrightarrow X(f)$, $h(t)\leftrightarrow H(f)$, then

$$x(t)h(t)\leftrightarrow X(f)*H(f) \tag{2.59}$$

(7) Parseval equality. The energy of signal $x(t)$ in time domain is equal to the integration of the spectrum $x(f)$ in frequency domain

$$\int_{-\infty}^{\infty}|x(t)|^2\mathrm{d}t=\int_{-\infty}^{\infty}|X(f)|^2\mathrm{d}f \tag{2.60}$$

Definition 2.7 *Assuming $x(t)$ is a stationary, ergodic random signal, let*

$$x_T(t)=\begin{cases}x(t) & |t|\leq T/2\\ 0 & otherwise\end{cases} \tag{2.61}$$

Denote $x_T(t)\leftrightarrow X_T(f)$, the limit of the mean square value

$$K_x(f)=\lim_{T\to\infty}\frac{1}{T}E(|X_T(f)|^2) \tag{2.62}$$

is called the mean power spectrum of $x(t)$, where E denotes the expectation

operation for a random signal. It is proved that the limit on the right side must exist. The following Wiener-Khinchine theorem explores the substantial relationship between the power spectrum and the auto-correlation function of a random signal.

2.3.2 Wiener-Khinchine Theorem

The Wiener-Khinchine theorem is the most important formula relating to the correlation function and power spectrum of a random signal. It is also one of the most important theoretical results in the field of random signal processing.

The following is the Wiener-Khinchine theorem of a stationary random signal. Suppose $x(t)$ is a stationary random signal, $E[x^2(t)] < \infty$ then the auto-correlation function $R_x(\tau) = E[x(t)x(t-\tau)]$ and the mean power spectrum $K_x(f)$ is a pair of Fourier transforms, i.e.

$$K_x(f) = \int_{-\infty}^{\infty} R_x(\tau)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau\,, \quad R_x(\tau) = \int_{-\infty}^{\infty} K_x(f)\exp(2\pi \mathrm{j} f\tau)\mathrm{d}f \tag{2.63}$$

2.3.3 Discrete Fourier Transform

Almost all of these important properties of analogue signals also apply to digital signals, but some modifications must be taken into account. Fig. 2.18 summarizes the main properties of digital signals in time and frequency domains.

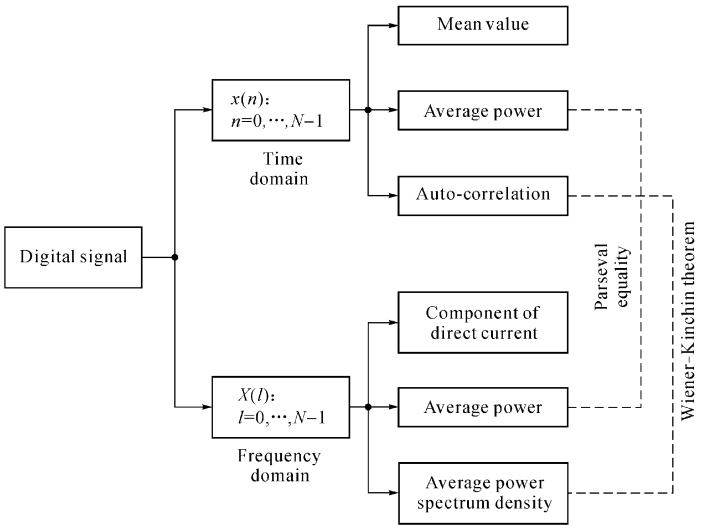


Fig. 2.18. Properties of digital signals in time and frequency domains

It can be shown that analogue signals for deterministic or random signals can be digitized if they satisfy some conditions. Denote the samples of signal $x(t)$, by $x(kT_s)$, $(k = 0, 1, 2, ...)$, where T_s is the sample duration.

Definition 2.8 *DFT (The discrete Fourier transform) of N point time series* $x(kT_s)$, $(k = 0, 1, ..., N-1)$ *is defined as*

$$X\left(\frac{l}{NT_s}\right) = \sum_{k=0}^{N-1} x(kT_s)\exp(-2\pi \mathrm{j}kl/N)\ ,\ (l = 0, 1, ..., N-1) \tag{2.64}$$

It is proved that

$$x(kT_s) = \frac{1}{N}\sum_{l=0}^{N-1} X\left(\frac{l}{NT_s}\right)\exp(2\pi \mathrm{j}kl/N),\ (k = 0, 1, ..., N-1) \tag{2.65}$$

This formula is called the IDFT (inverse discrete Fourier transform). $1/(NT_s)$ is sometimes called the frequency resolution bin and is actually the duration of the adjacent frequency in DFT. Since the sample duration T_s and frequency resolution bin $1/(NT_s)$ are clearly always present in DFT and IDFT, for simplicity, we denote $x(kT_s)$ as $x(k)$, and $X(l/NT_s)$ as $X(l)$. We can denote $X(l) = DFT[x(k)]$, and $x(k) = IDFT[X(l)]$.

Important Properties of DFT

(1) Symmetry. If the DFT of $x(k)$ is $X(l)$, then the DFT of its conjugate $X^*(l)$ is $x^*(k)$, because

$$x(k) = \frac{1}{N}\sum_{l=0}^{N-1} X(l)\exp(2\pi \mathrm{j}kl/N) = \left(\frac{1}{N}\sum_{l=0}^{N-1} X^*(l)\exp(-2\pi \mathrm{j}kl/N)\right)^*$$

Taking conjugate operation on both sides, therefore

$$x^*(k) = DFT[X^*(l)] \tag{2.66}$$

(2) Periodicity. If we extend the definition for $X(l)$ from $l = 0, 1, ..., N-1$ to any integer, then for any integer r, $X(l) = X(rN + l)$, $(l = 0, 1, ..., N-1)$. Thus, $X(l)$ is a periodic discrete series. This conclusion also applies to $x(k)$, if we extend the argument field from $k = 0, 1, ..., N-1$ to any integer, i.e., $x(k) = x(k + rN)$ $(k = 0, 1, ..., N-1)$. The data extension from $k = 0, 1, ..., N-1$ to any integer is called a periodic extension. This is a normal and inherent procedure in DFT, if there are no other notifications.

(3) Convolution. Suppose there are two time series $x(k)$ and $h(k)$ $(k = 0, 1, ..., N-1)$, after periodic extension, $x(k + rN) = x(k)$, $h(k + rN) = h(k)$, $k = 0, 1, ..., N-1$

for any integer r. The equation $y(k)=\sum_{i=0}^{N-1}x(i)h(k-i)=\sum_{i=0}^{N-1}h(i)x(k-i)$ is called the cycle convolution of $x(k)$ and $h(k)$. If the DFT of $x(k)$ and $h(k)$ is $X(l)$ and $H(l)$, then the DFT of $y(k)$ is

$$Y(l)=DFT[y(k)]=X(l)H(l)\ \ (l=0,\ 1,\ ...,\ N-1) \tag{2.67}$$

(4) Correlation. The correlation operation is the same as the convolution operation except that there is a difference in the order of the time delay. The correlation function of $x(k)$ and $y(k)$ is defined as

$$z_{xy}(k)=\sum_{i=0}^{N-1}x(i)y(i-k)\ (k=0,\ 1,\ ...,\ N-1) \tag{2.68}$$

The DFT of $z_{xy}(k)$ is

$$Z_{xy}(l)=X(l)Y^{*}(l)\ \ (l=0,\ 1,\ ...,\ N-1) \tag{2.69}$$

(5) Time delay. If the DFT of $x(k)$ is $X(l)$, i.e., $X(l)=DFT[x(k)]$, then the DFT of $x(k-i)$ is $X(l)\exp(-2\pi \mathrm{j}il/N)$, i.e.

$$X(l)\exp(-2\pi \mathrm{j}il/N)=DFT[x(k-i)] \tag{2.70}$$

(6) Modulation. Assume $X(l)=DFT[x(k)]$, then

$$X(l-i)=DFT[x(k)\exp(2\pi \mathrm{j}ki)] \tag{2.71}$$

Signal $x(k)\exp(2\pi \mathrm{j}ki)$ is the modulation of $x(k)$.

(7) Parseval equality. Assume $X(l)=DFT[x(k)]$, then

$$\sum_{k=0}^{N-1}|x(k)|^{2}=\frac{1}{N}\sum_{l=0}^{N-1}|X(l)|^{2} \tag{2.72}$$

Note that the Parseval equality of DFT differs from that of an analogue signal by a factor of $1/N$.

2.3.4 Digital Feature of Signal Represented by Discrete Samples

Note that when we compare digital and analogue processing of signal, we have to know that the DFT of the discrete series is the truncated signal and therefore the results of DFT must be periodic. Convolution and correlation are essential for a discrete series to be periodically extended. This fact must be dealt with carefully

because, in practice, we are facing a sequential time series, but the DFT is actually a block process. Therefore, a special algorithm is needed to guarantee that the digitizing process does not result in distortion.

The mean value of a discrete signal $x(k)$, $(k = 0, 1, ..., N-1)$ is defined as

$$\overline{x} = \frac{1}{N}\sum_{k=0}^{N-1} x(k) \tag{2.73}$$

This value is actually a component of direct current of DFT for $x(k)$. In fact

$$X(0) = \sum_{k=0}^{N-1} x(k) \tag{2.74}$$

i.e.

$$\overline{x} = X(0)/N \tag{2.75}$$

The mean square value (power) of $x(k)$ $(k = 0, 1, ..., N-1)$ is defined as

$$\overline{x}^2 = \frac{1}{N}\sum_{k=0}^{N-1} |\, x(k)\,|^2 \tag{2.76}$$

Based on the Parseval equality

$$\overline{x}^2 = \frac{1}{N^2}\sum_{l=0}^{N-1} |\, X(l)\,|^2 \tag{2.77}$$

Strictly speaking, to calculate the correlation function of a signal, it is necessary to have infinitely long samples, even for stationary, ergodic random signals, since according to the definition of the auto-correlation function, it is the limit of some kind of average value, i.e.

$$R_x(l) = E[x(k)x(k-l)] = \lim_{M\to\infty}\frac{1}{M}\sum_{k=0}^{M-1} x(k)x(k-l) \tag{2.78}$$

where $x(k)$ is supposed to be a stationary, ergodic random signal. The definition of the cross correlation function is similar: for two signals $x(k)$, $y(k)$ the cross correlation is defined by

$$R_{xy}(l) = E[x(k)y(k-l)] = \lim_{M\to\infty}\frac{1}{M}\sum_{k=0}^{M-1} x(k)y(k-l) \tag{2.79}$$

Clearly, to calculate the correlation function, it is necessary to have enough samples of signal. The cycle convolution operation discussed above cannot be used directly to calculate the correlation function, and therefore it cannot directly establish the relationship between the correlation function and the power spectrum

for discrete data. We will solve this problem later. Fig. 2.19 shows a block diagram for calculating the auto-correlation and cross-correlation functions. Generally, we should use sufficiently large M, and then use the approximation expression

$$R_x(l) \approx \frac{1}{M}\sum_{k=0}^{M-1} x(k)x(k-l) \tag{2.80}$$

$$R_{xy}(l) \approx \frac{1}{M}\sum_{k=0}^{M-1} x(k)y(k-l) \tag{2.81}$$

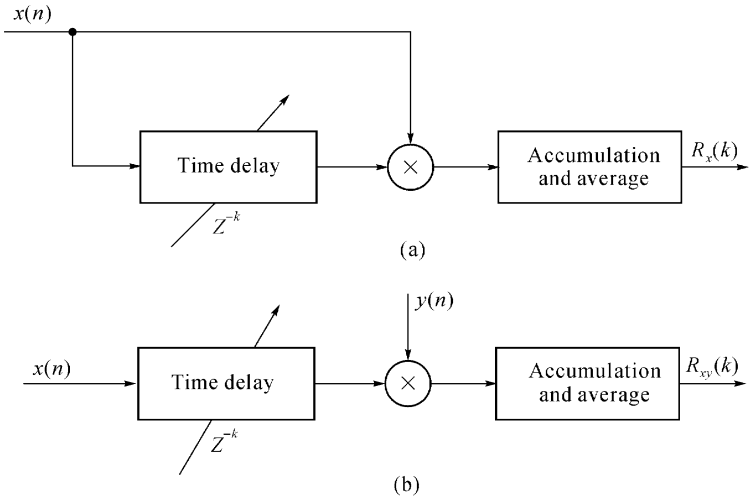


Fig. 2.19. Auto-correlation and cross-correlation algorithm of digital signal. (a) Auto-correlation; (b) Cross-correlation

The relationship between $R_x(l)$ and the power spectrum of $x(k)$ can be established using the Wiener-Khinchine theorem, but it is only an approximate relationship. We will discuss this problem as part of spectrum estimation.

2.3.5 *Algorithm of Fast Fourier Transform*

The fast Fourier transform (FFT) algorithm, proposed by Cooley and Tukey in 1965[31], has been of great benefit to the field of spectrum analysis of signal and, in particular, to digital signal processing. It reduces the DFT operation numbers from N^2 complex multiply to $N(\log_2 N) / 2$. When N is very large, the operation number reduction is considerable. Almost all scientists and engineers in the digital signal processing community have found the FFT algorithm very valuable, but why did it take until the 1960s to develop such an efficient algorithm? Dr. Heideman from

Rice University showed that[34]: as early as 1805, the famous mathematician Gauss had already proposed an algorithm similar to the modern FFT. The FFT algorithm meets the requirement of modern digital signal processing, and therefore has been used in many applications, including correlation analysis, spectrum estimation, beamforming in frequency domain and adaptive noise canceling. The fast calculation concept in digital signal processing has also been generalized in many other transformations, such as the fast Hartly transform, fast Walsh transform, etc.

Since the derivation of the FFT algorithm is a purely mathematical topic, we will not give a general proof of this algorithm. To illustrate the algorithm, taking $N = 4$ as an example, we will explain how the FFT reduces the operation number. The method used here is matrix decomposition. The decomposed matrix can be expressed as a signal flow chart. Based on this signal flow chart, a program of FFT can be assembled.

Consider the DFT of data series $x_0(k),\ (k = 0,\ 1,\ ...,\ N-1)$

$$X(l) = \sum_{k=0}^{N-1} x_0(k) \exp(-2\pi \mathrm{j} kl / N) \tag{2.82}$$

Here, the subscript 0 denotes the original data, and will correspond to the decomposition procedure. Let $N = 4,\ W = \exp(-2\pi \mathrm{j} / 4)$, we have

$$\begin{pmatrix} X(0) \\ X(1) \\ X(2) \\ X(3) \end{pmatrix} = \begin{pmatrix} W^0 & W^0 & W^0 & W^0 \\ W^0 & W^1 & W^2 & W^3 \\ W^0 & W^2 & W^4 & W^6 \\ W^0 & W^3 & W^6 & W^9 \end{pmatrix} \begin{pmatrix} x_0(0) \\ x_0(1) \\ x_0(2) \\ x_0(3) \end{pmatrix} \tag{2.83}$$

As the periodicity is W^p, Eq. (2.83) can be rewritten as

$$\begin{pmatrix} X(0) \\ X(1) \\ X(2) \\ X(3) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & W^1 & W^2 & W^3 \\ 1 & W^2 & W^0 & W^2 \\ 1 & W^3 & W^2 & W^1 \end{pmatrix} \begin{pmatrix} x_0(0) \\ x_0(1) \\ x_0(2) \\ x_0(3) \end{pmatrix} \tag{2.84}$$

Now decompose the matrix on the right side of Eq. (2.84)

$$\begin{pmatrix} X(0) \\ X(2) \\ X(1) \\ X(3) \end{pmatrix} = \begin{pmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^1 \\ 0 & 0 & 1 & W^3 \end{pmatrix} \begin{pmatrix} 1 & 0 & W^0 & 0 \\ 0 & 1 & 0 & W^0 \\ 1 & 0 & W^2 & 0 \\ 0 & 1 & 0 & W^2 \end{pmatrix} \begin{pmatrix} x_0(0) \\ x_0(1) \\ x_0(2) \\ x_0(3) \end{pmatrix}$$

Note that before decomposing the matrix, the second and third lines are already

replaced, i.e., the order of $X(0)$, $X(1)$, $X(2)$, $X(3)$ is changed to $X(0)$, $X(2)$, $X(1)$, $X(3)$. This kind of replacement is called "order modification" in the FFT algorithm. The matrix decomposition is the basis of the efficiency of the FFT algorithm. It actually divides the calculation into two steps: first

$$\begin{pmatrix} x_1(0) \\ x_1(1) \\ x_1(2) \\ x_1(3) \end{pmatrix} = \begin{pmatrix} 1 & 0 & W^0 & 0 \\ 0 & 1 & 0 & W^0 \\ 1 & 0 & W^2 & 0 \\ 0 & 1 & 0 & W^2 \end{pmatrix} \begin{pmatrix} x_0(0) \\ x_0(1) \\ x_0(2) \\ x_0(3) \end{pmatrix}$$

second

$$\begin{pmatrix} X(0) \\ X(2) \\ X(1) \\ X(3) \end{pmatrix} = \begin{pmatrix} x_2(0) \\ x_2(1) \\ x_2(2) \\ x_2(3) \end{pmatrix} = \begin{pmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^1 \\ 0 & 0 & 1 & W^3 \end{pmatrix} \begin{pmatrix} x_1(0) \\ x_1(1) \\ x_1(2) \\ x_1(3) \end{pmatrix}$$

Thus, we obtain the signal flow chart of FFT for $N = 4$ (Fig. 2.20). This kind of flow chart is called a butterfly graph. In this graph, each column corresponds to one decomposed matrix factor. If $N = 2^r$, then it is necessary to have r columns. Each column has N nodes. Two dashed lines entering each node represent the transfer paths of previous nodes and bring the calculated value to the next nodes by multiplying W^P. The coefficient of W^P is indicated near the arrow of each transfer path. If there is no coefficient, then $W^P = 1$. The final results $x_r(k)$, (k = 0, 1, ..., N-1) are simply the value of the spectrum $X(l)$, (l = 0, 1, ..., N-1).

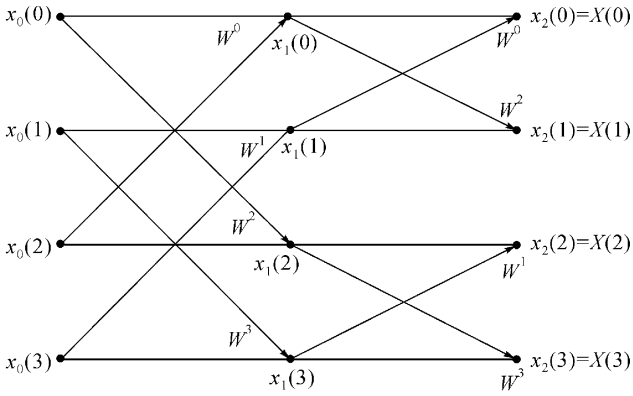


Fig. 2.20. Flow chart of FFT for $N = 4$

Because the FFT algorithm is used so extensively in digital signal processing, many studies have been published and many specific DSP chips have been

designed to implement the FFT algorithm. The normal operation time of 1,024 point FFT for a specific DSP chip is about 100μs.

Table 2.2 shows the reduction in operation number achieved by using FFT.

Table 2.2 Comparison of operation number of DFT and FFT

Number of samples(N)	Multiplication number of DFT (N^2)	Multiplication number of FFT($N/2$)$\log_2 N$	Reduction factor
4	16	4	4.0
16	256	32	8.0
64	4,096	192	21.3
256	65,356	1,024	64.0
512	262,144	2,304	113.8
1,024	1,048,576	5,120	204.8

2.3.6 *Calculation of DFT for Real Value Data*

In general, the FFT calculation program is designed for complex data, as it is necessary to use the same program for calculating both DFT and IDFT. If the input data of a program of FFT are real values, then half of the operation will be wasted. The following method can be used to improve the calculation efficiency.

(1) Simultaneously calculate the DFT of two real data series. Suppose $g(k)$ and $h(k)$ are two real data series, then construct N point complex data series $y(k) = h(k) + \mathrm{j}g(k) \quad (k = 0,\ 1,\ ...,\ N-1)$. Calculate

$$Y(l) = DFT[y(k)] = \sum_{k=0}^{N-1} y(k)\exp(-2\pi \mathrm{j}k / N) = R(l) + \mathrm{j}I(l) \quad (l = 0,\ 1,\ ...,\ N-1)$$

Then the DFT of $g(k)$ and $h(k)$ are as follows

$$\begin{aligned} G(l) &= DFT[g(k)] = [I(l) + I(N-l)]/2 - \mathrm{j}[R(l) - R(N-l)]/2 \\ H(l) &= DFT[h(k)] = [R(l) + R(N-l)]/2 + \mathrm{j}[I(l) - I(N-l)]/2 \end{aligned} \tag{2.85}$$

(2) Calculation of $2N$ point DFT of real data using N point FFT program of complex data. Suppose $x(k) \quad (k = 0,\ 1,\ ...,\ 2N-1)$ is a real data series. Let $h(k) = x(2k),\ g(k) = x(2k+1) \quad (k = 0,\ 1,\ ...,\ N-1)$

To construct complex data series

$$y(k) = h(k) + \mathrm{j}g(k) \quad (k = 0,\ 1,\ ...,\ N-1) \tag{2.86}$$

The DFT of $y(k)$ is

$$Y(l) = DFT[y(k)] = \sum_{k=0}^{N-1} y(k)\exp(-2\pi \mathrm{j}kl / N) = R(l) + \mathrm{j}I(l) \quad (l = 0,\ 1,\ ...,\ N-1)$$

From this we can obtain the real part $X_r(l)$ and image part $X_r(l)$ of the DFT of $x(k)$, i.e.

$$\begin{aligned} X_r(l) &= [R(l) + R(N-l)]/2 + \cos(l\pi / N)[I(l) + I(N-l)]/2 \\ &\quad - \sin(l\pi / N)[R(l) - R(N-l)]/2 \\ X_i(l) &= [I(l) - I(N-l)]/2 - \sin(l\pi / N)[I(l) + I(N-l)]/2 \\ &\quad - \cos(l\pi / N)[R(l) - R(N-l)]/2 \end{aligned} \tag{2.87}$$

Example

Suppose the bandwidth of the pre-filter of a sonar system is 4 – 10 kHz. Based on the sampling theorem, the sampling frequency f_s must be greater than 20 kHz. We choose f_s=25 kHz, therefore the sampling duration T_s = 40 μs. A 1,024 point FFT operation can obtain 1,024 spectrum values in frequency bins. The duration of a frequency bin is $\Delta f = f_s / N$ = 24.4 Hz. Since the spectrum of real data is symmetrical to the $N/2$ = 512 of the frequency bin, there are only 0 – 511 independent frequency values. According to the bandwidth of the signal, only 164 – 410 of the frequency bins are considered.

2.4 Basic Processing Technique for One-Dimensional Digital Signal

The linear, time invariant filter was introduced briefly in previous section. Note that this filter forms only a very small part of the filter family. Many filters applied in signal processing and engineering applications are not linear and time invariant. In this section, some filtering techniques used frequently for one-dimensional signal processing are discussed. We can consider these filters as components to be integrated into a complete sonar system.

The classification of cascade methods in digital systems is given in Fig. 2.21. There are four kinds of cascade: series, parallel, series / parallel and feedback. Many filtering techniques are used in one-dimensional signal processing[44-46]. The key point in distinguishing digital filters is to find the relationship between the input and output.

2.4.1 Local Average Filtering

The local average algorithm of digital data is a date smoothing operation. There are two kinds of local average algorithms. The first is a moving average with a

rectangular window (Fig. 2.22(a)). Assuming that the input data is $x(n)$ ($n = 1, 2, ...$), and the window length is M, the output signal of the moving average is

$$y(n) = \frac{1}{M} = \sum_{k=1}^{M} x(n-k) \quad (n \geq M) \tag{2.88}$$

The calculation of this kind of local average is simple, but it should be noted that $\sum_{k=1}^{M} x(n-k)$ may overflow when M is very large.

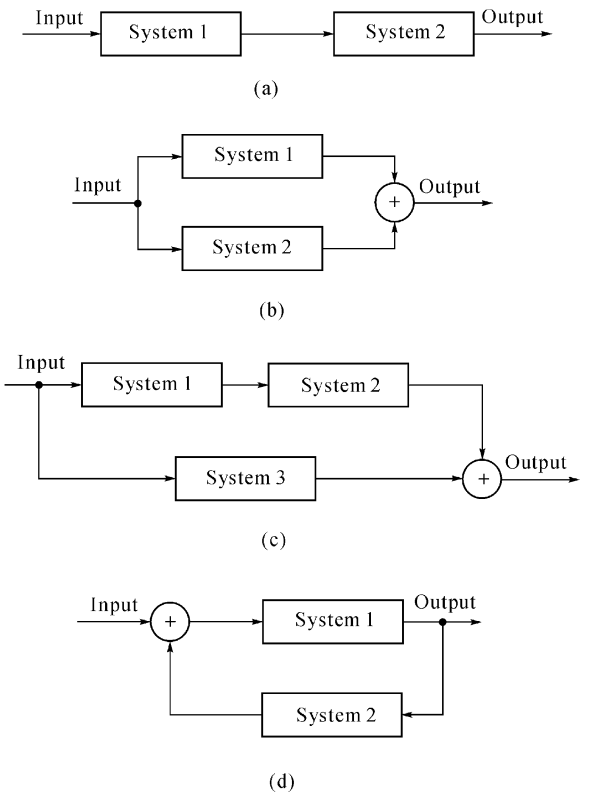


Fig. 2.21. Classification of cascades in digital systems. (a) Series; (b) Parallel; (c) Series/Parallel; (d) Feedback

Another local average algorithm, which can avoid the overflow problem, is called the recursive moving average (Fig. 2.22(b)). The input/output relationship is

$$y(n) = \left(1 - \frac{1}{M}\right) y(n-1) + \frac{x(n)}{M} \tag{2.89}$$

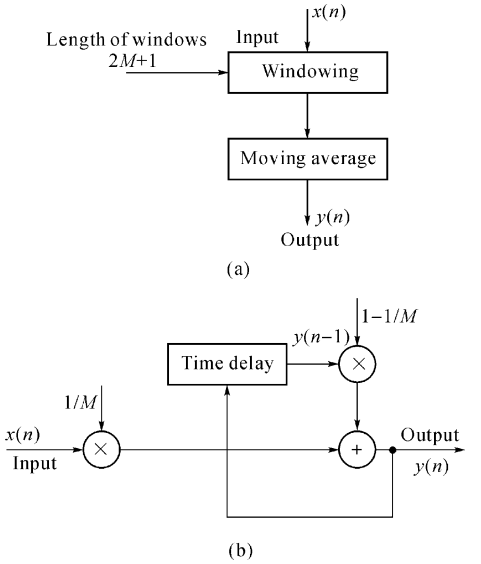


Fig. 2.22. Two algorithms of local average. (a) Rectangle windowing moving average; (b) Recursive moving average

The output value $y(n)$ depends not only on the input $x(n)$, but also on the output value $y(n–1)$ of the last moment. This kind of smoothing average has very good input / output performance. When the window length has the form of $M = 2^N$, the filtering algorithm can be implemented by data shifting and there are no overflow problems. Fig. 2.23 plots the results of two local average algorithms with a window length of 32. The filtering results are basically the same.

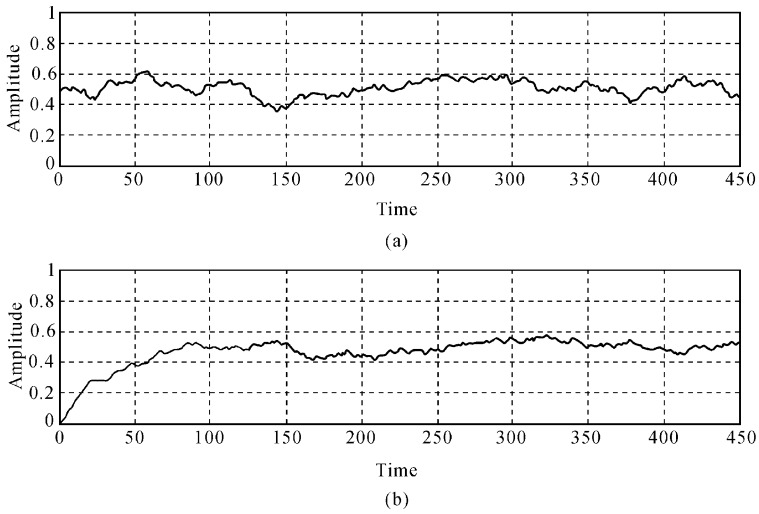


Fig. 2.23. Average of random number. (a) Moving average of rectangle window ($M = 32$); (b) Recursive moving average ($M = 32$)

2.4.2 Median Value Filtering

Median filtering is a non-linear process, used mainly when a stationary random signal is corrupted by impulse interference. By using a median filter, the impulse interference, which appears randomly in time moment, can be effectively canceled and the stationary random signal remains unchanged. Fig. 2.24 shows a block diagram of a median value filter.

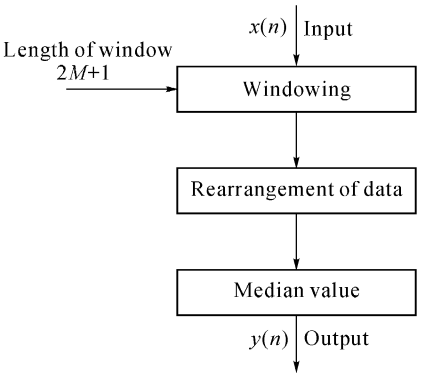


Fig. 2.24. Flow chart of a median value filtering algorithm

Suppose the input signal is $x(n)$ (n = 1, 2, …), and the window length of the median filter is $2M$+1. The median filter is a point-to-point process, but each time it is necessary to process a data series of length $2M$+1. For the moment n, choose the data $x(n–M)$, ..., $x(n–1)$, $x(n)$, $x(n+1)$, ..., $x(n+M)$, Rearranging this sequence based on value magnitude from small to large, we obtain a new sequence $v(1)$, $v(2)$, ..., $v(2M+1)$. The value in the middle is called the median value, that is $v(M+1)$.

Now, denote

$$y(n) = v(M+1) = \text{Median value}\{x(n-M),\ ...,\ x(n+M)\} \qquad (2.90)$$

Figs. 2.25 – 2.27 shows some results from system simulations. In Fig. 2.25 we can see that for a stationary random signal the filtering results of a moving average and a median value filter are about the same. If the window length becomes large, the median value filter will retain more fluctuating values compared with the moving average filter. The effect of impulse suppression is very clear when using a median value filter (Fig. 2.27).

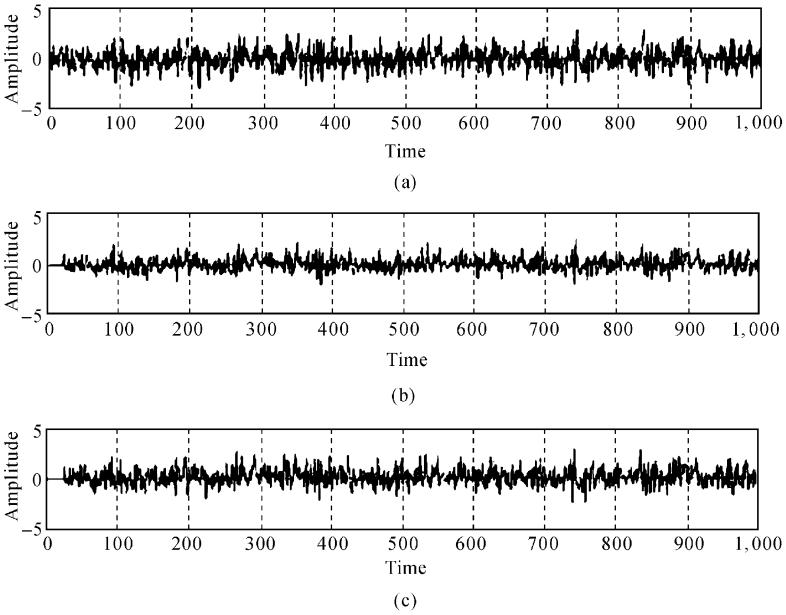


Fig. 2.25. Mean and median value filtering ($M = 25$). (a) Gaussian noise; (b) Mean value filtering; (c) Median value filtering

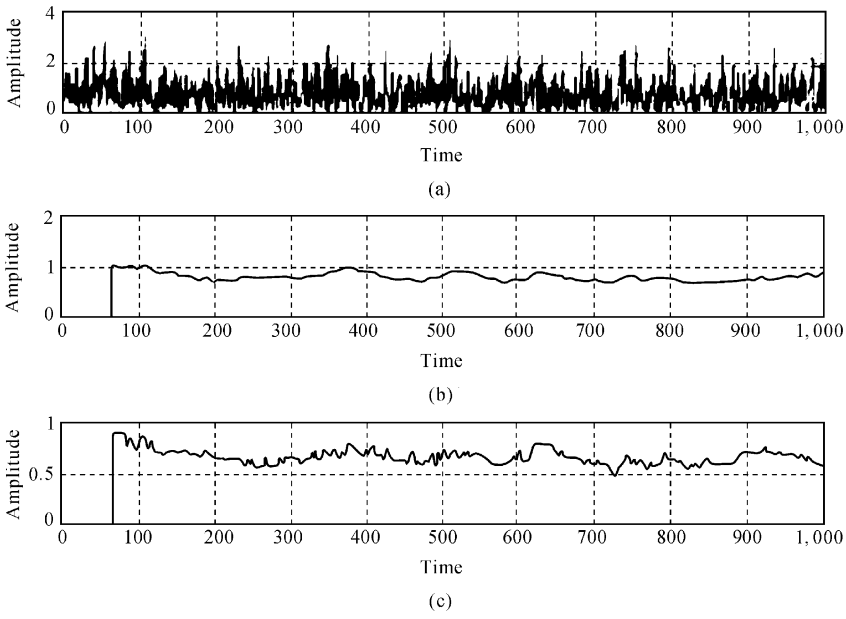


Fig. 2.26. Mean and median value filtering ($M = 65$). (a) Input signal: Gaussian random noise; (b) Mean value filtering; (c) Median value filtering

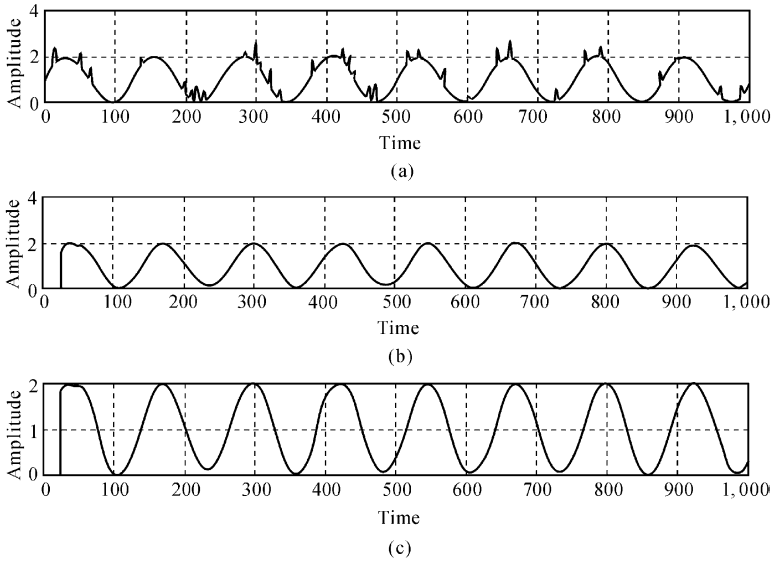


Fig. 2.27. Suppressing interference by using mean or median value filtering. (a) Input sine wave plus random interference; (b) Mean value filtering; (c) Median value filtering ($M = 25$)

2.4.3 *Threshold Filtering and Truncate Filtering*

The threshold and truncate process is a kind of filter which rearranges data in a certain range and sets a part of the data to be zero. Only those data that change quickly and suddenly are extracted. Fig. 2.28 shows a block diagram of these two algorithms.

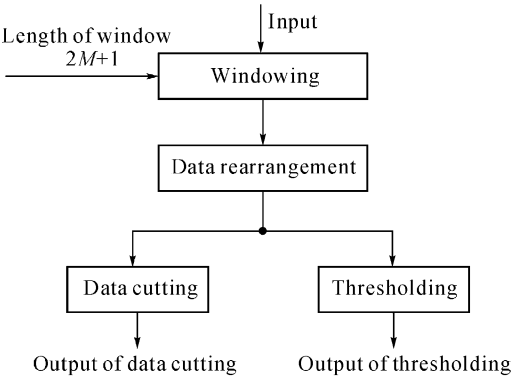


Fig. 2.28. Data cutting and threshold filtering

Suppose the input signal is $x(n)$ $(n = 1, 2, ...)$ and the window length is $2M+1$. In this window, the data are

$$x(n-M),\ ...,\ x(n-1),\ x(n),\ x(n+1),\ ...,\ x(n+M) \tag{2.91}$$

Taking the average

$$\overline{x}_n = \frac{1}{2M+1}\sum_{k=-M}^{M} x(n+k) \tag{2.92}$$

According to the results of comparison of $x(n)$ and $\overline{x}_n$, the threshold filtering output is

$$y(n) = \begin{cases} 0 & x(n) \le \overline{x}_n \\ x(n) - \overline{x}_n & x(n) > \overline{x}_n \end{cases} \tag{2.93}$$

Note that the threshold value of $\overline{x}_n$ can also have the form of $\alpha\overline{x}_n$ ($\alpha > 0$). For different α, the effect of threshold filtering is different, therefore α can be used to adjust the filtering performance.

For truncate filtering, the input data in the selected window should be rearranged based on the magnitude of the values, from small to large

$$v(1),\ v(2),\ ...,\ v(2M+1) \tag{2.94}$$

Taking the average

$$\tilde{x}_n = \sum_{k=-M}^{0} v(M+k+1) \tag{2.95}$$

The output of truncate filtering is

$$y(n) = \begin{cases} 0 & x(n) \le \tilde{x}_n \\ x(n) - \tilde{x}_n & x(n) > \tilde{x}_n \end{cases} \tag{2.96}$$

Here, the truncate value can be chosen also as $\alpha\tilde{x}_n$. Fig. 2.29 shows the results of system simulation. These two kinds of filtering are quite good for extracting sudden change values. It is necessary to show that threshold filtering and truncate filtering are useful in post processing of digital sonar. They have different application areas that will be discussed later. The threshold and truncate values are not preset before filtering, but are defined in the procedure of data manipulation. In particular, the window length will play an important role.

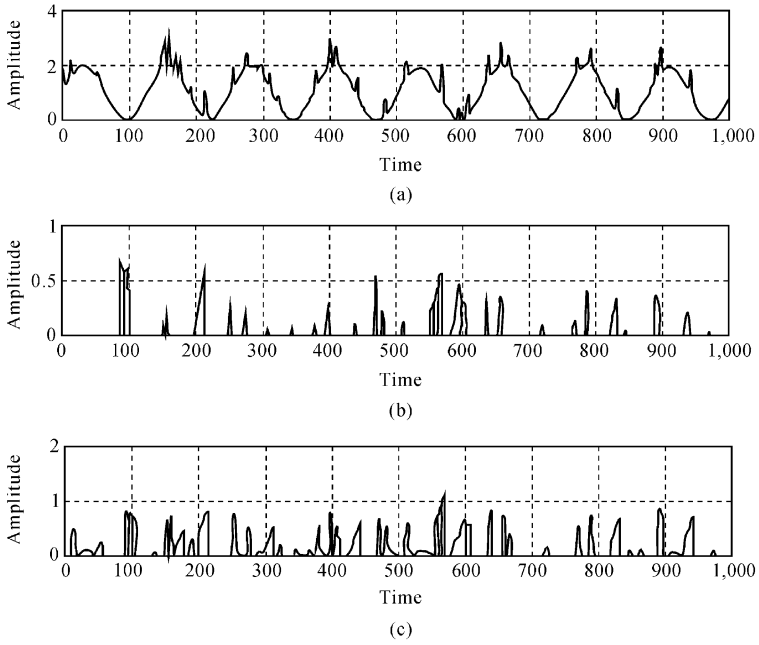


Fig. 2.29. Threshold and truncate filtering. (a) Input: sine wave plus random interference; (b) Truncate filtering ($N = 25$); (c) Threshold filtering

2.5 Two-Dimensional Digital Image Signal Processing

In the design of digital sonar, the two-dimensional signal processing technique is used mainly in the display system, i.e., the post-processing of the sonar signal. It is well known that the contents of two-dimensional signal processing are greatly enriched. In this section, only the basic theoretical results and techniques which are useful in digital sonar application are discussed, including the time / bearing waterfall display technique, target trace extraction technique and image enhancement, etc. Some two-dimensional image processing techniques used in sonar systems have their own characteristics because, from a detection point of view, the requirements of sonar display are different from those of pure image processing.

2.5.1 Definition of Digital Images

For a two-dimensional continuous value image, in position (x, y) there is a color / brightness value $c(x, y, t)$ where t is the present time moment. Usually, for

simplification, $c(x, y, t)$ is limited to positive and boundary, i.e.

$$0 \le c(x, y, t) \le A \tag{2.97}$$

For a discrete digital image, the position coordinate and color value are discrete. Neglecting the time moment, the value in each pixel (k, l) is $F(k, l)$ $(1 \le k \le K, 1 \le l \le L)$. The color function F can be considered as a three-dimensional vector, for example, [0, 0, 0] denoting white, [1, 0, 0] denoting pure red and [0.5, 0.5, 0.5] denoting middle gray, etc. If F is a single color and varies only in brightness, then F is a scalar, e.g., gray color $[\alpha, \alpha, \alpha]$ $(0 \le \alpha \le 1)$.

The upper boundary K, L of (k, l), sometimes written as $K \times L$, is called the resolution. K is the number of columns and L is the number of rows. For a fixed screen, the larger $K \times L$ is, the higher the resolution is.

For image processing, the general expression can be written as

$$P(m, n) = \sum_{k=1}^{K} \sum_{l=1}^{L} F(k, l) Q(k, l; m, n) \tag{2.98}$$

where $Q(k, l; m, n)$ represents the weighting operation on the image and $P(m, n)$ is the image after processing. The range of argument $1 \le m \le M$, $1 \le n \le N$ can be different with K, L.

2.5.2 2D Fourier Transform

It is well known that the Fourier transform of a one-dimensional signal is a very important means of analysis in signal processing. The Fourier transform of the time function is the frequency spectrum and the physical meaning is clear. The magnitude of the spectrum value in the frequency domain represents the distribution of signal variation, quickly or slowly. Therefore, the filtering process can directly change the signal behavior.

For a two-dimensional signal, the Fourier transform can be formally defined according to the one-dimensional situation. But, there is no direct, comprehensive meaning as in the case of a one-dimensional signal.

Suppose $x(k, l)$ $(k = 0, ..., K-1; l = 0, 1, ..., L-1)$ is a two-dimension signal. The Fourier transform is defined by

$$X(u, v) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} x(k, l) \exp(-2\pi \mathrm{j} ku / K) \exp(-2\pi \mathrm{j} lv / L) \tag{2.99}$$

It is proved that the inverse Fourier transform is

$$x(k, l) = \frac{1}{KL} \sum_{u=0}^{K-1} \sum_{v=0}^{L-1} X(u, v) \exp(2\pi \mathrm{j} ku / K) \exp(2\pi \mathrm{j} lv / L) \tag{2.100}$$

As we see in Eq. (2.100), there is a factor of $(1/KL)$.

There is also a fast calculation algorithm for the two-dimensional Fourier transform, that is 2D FFT. Similar to 1-D FFT, to perform 2D FFT it is necessary to periodically extend $x(k,l)$. For any integers m, n, let

$$x_p(k+mK,\ l+nL) = x(k,\ l) \tag{2.101}$$

Perform 2D DFT on the $x_p(k, l)$. We can obtain results similar to those from 1-D DFT, but we will not discuss the details here. As we can see in Eq. (2.99), the 2D DFT can be considered as a two-step 1-D DFT.

2.5.3 2D Cosine Discrere Transforms

The 2D discrete cosine transform (DCT) is extensively used in image processing because its results for real data are real. We will derive the DCT from the 1-D case and then generalize to 2D.

Suppose $x(n)$ is an N point real data series and $x(n)$ is zero when n≥0. Firstly, we construct a $2N$ point series $y(n)$, and then perform DFT on $y(n)$ and obtain $Y(k)$. From $Y(k)$ we find the DCT $C_x(k)$ of $x(n)$. The operation procedure is shown below

$$\begin{array}{ccccccc} x(n) & \leftrightarrow & y(n) & \leftrightarrow & Y(k) & \leftrightarrow & C_x(k) \\ \Updownarrow & & \Updownarrow & & \Updownarrow & & \Updownarrow \\ N \text{ point} & & 2N \text{ point DFT} & & 2N \text{ point} & & N \text{ point} \end{array}$$

Construct $y(n)$ from $x(n)$

$$y(n) = \begin{cases} x(n) & 0 \le n \le N-1 \\ x(2N-1-n) & N \le n \le 2N-1 \end{cases} \tag{2.102}$$

Data $y(n)$ is symmetrical to $n = N-1/2$. Taking $2N$ point DFT for $y(n)$, we get

$$Y(k) = \sum_{n=0}^{2N-1} y(n) \exp(-2\pi \mathrm{j} kn/(2N)), \quad k = 0,\ ...,\ 2N-1 \tag{2.103}$$

It is straightforward to calculate that

$$Y(k) = \exp(\pi \mathrm{j} k/(2N)) \sum_{n=0}^{N-1} 2x(n) \cos[\pi k(2n+1)/(2N)], \quad k = 0,\ ...,\ 2N-1 \tag{2.104}$$

Thus, we obtain the N point DCT $C_x(k)$ of $x(n)$

$$C_x(k)=\begin{cases}\exp(-\pi \mathrm{j}k/(2N))Y(k)\\ \qquad 0\end{cases}$$
$$=\begin{cases}\sum_{n=0}^{N-1}2x(n)\cos[\pi k(2n+1)/(2N)] & 0\le k\le N-1\\ \qquad 0 & \text{otherwise}\end{cases} \tag{2.105}$$

We can see that for the real $x(n)$, $C_x(k)$ is also real, and the inverse transform of DCT is

$$x(n)=\begin{cases}\frac{1}{N}\sum_{k=0}^{N-1}w(k)C_x(k)\cos[\pi k(2n+1)/(2N)] & 0\le k\le N-1\\ \qquad 0 & \text{otherwise}\end{cases} \tag{2.106}$$

where

$$w(k)=\begin{cases}1/2 & k=0\\ 1 & 1\le k\le N-1\end{cases} \tag{2.107}$$

The following is the proof of Eq. (2.106), since

$$y(n)=\frac{1}{2N}\sum_{k=0}^{2N-1}Y(k)\exp(2\pi \mathrm{j}kn/(2N))\quad (n=0,\ ...,\ 2N-1)$$

and

$$x(n)=\begin{cases}y(n) & 0\le n\le N-1\\ 0 & \text{otherwise}\end{cases}$$

Therefore

$$x(n)=\begin{cases}\frac{1}{N}[C_x(0)/2+\sum_{k=1}^{N-1}C_x(k)\cos[\pi k(2n+1)/(2N)]\\ \qquad 0\end{cases}$$
$$=\begin{cases}\frac{1}{\mathrm{N}}\sum_{k=0}^{N-1}w(k)C_x(k)\cos[\pi k(2n+1)/(2N)] & 0\le n\le N-1\\ \qquad 0 & \text{otherwise}\end{cases}$$

The 1D DCT can be directly generalized to a 2D situation. Suppose $x(n_1, n_2)$ is an $N_1 \times N_2$ data series, for $n_1 \geq N_1$ or $n_2 \geq N_2$, $x(n_1, n_2) = 0$. It is proved that 2D DCT pairs are

$$C_x(k_1, k_2) = \begin{cases} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} 4x(n_1, n_2) \cos\left[\frac{\pi k_1(2n_1+1)}{2N_1}\right] \cos\left[\frac{\pi k_2(2n_2+1)}{2N_2}\right] & k_1 = 0, \ldots, N_1-1;\ k_2 = 0, \ldots, N_2 \\ 0 & \text{otherwise} \end{cases} \tag{2.108}$$

$$x(n_1, n_2) = \begin{cases} \frac{1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} w_1(k_1) w_2(k_2) C_x(k_1, k_2) \cos\left[\frac{\pi k_1(2n_1+1)}{2N_1}\right] \cos\left[\frac{\pi k_2(2n_2+1)}{2N_2}\right] & k_1 = 0, \ldots, N_1-1;\ k_2 = 0, \ldots, N_2 \\ 0 & \text{otherwise} \end{cases} \tag{2.109}$$

2.5.4 Typical Image Processing Techniques

Typical image processing techniques are point to point processing, area processing and frame processing. As image processing is used extensively in many areas, a standard is needed to evaluate the quality of an image so that we can determine which kind of image processing is suitable for improving image quality. The IEEE Image Processing Society chose 31 frames of an image as subject and a standard picture to be used for testing image processing algorithms. Fig. 2.30 shows two of these pictures, the famous "Lenna" and "Baboon" photographs.



Fig. 2.30. Standard images for signal processing. (a) 256×256 pixels Lenna; (b) 256×256 pixels Baboon

Point to point processing is the transformation from pixels to pixels. The most convenient method is to use a look-up-table. This is a pixel-to-pixel color corresponding table. The implementation of a look-up-table includes image negative processing, threshold processing, contrast adjust processing and quasi-color processing, etc.

Image area processing is also called group processing. It is usual to choose a rectangular area around the pixel of interest and then to perform the convolution operation. The variation in brightness or color distribution between areas is usually called the spatial frequency. For example, for a 3×3 window, with a pixel to be processed denoted as P_5, the area consists of 9 points

$$\begin{matrix} P_1 & P_2 & P_3 \\ P_4 & P_5 & P_6 \\ P_7 & P_8 & P_9 \end{matrix}$$

Given a weighting matrix

$$\begin{matrix} K_1 & K_2 & K_3 \\ K_4 & K_5 & K_6 \\ K_7 & K_8 & K_9 \end{matrix}$$

we can obtain a new pixel value

$$Q_5 = K_1 \times P_1 + \ldots + K_9 \times P_9$$

Of course, there are many other methods for area processing, but the convolution type of processing is the one most frequently used. Here are a 3×3 low pass 2D filter and a high pass 2D filter

Low pass filter			High pass filter		
1/10	1/10	1/10	0	−1	0
1/10	1/5	1/10	−1	5	−1
1/10	1/10	1/10	0	−1	0

The technique for enlarging a picture can be considered as area processing. Fig. 2.31 is the result of local enlargement of Lenna's photo. Fig. 2.32 is a 3×3 window threshold processing and gray scale conversion.

2D median value filtering is a method of image processing that is effective for suppressing impulse type interferences. Fig. 2.33 is a result of system simulation; the white noise is added in Lenna's original photo.

Frame processing is a global image processing technique, and includes image rotation, inversion, scale conversion, pixel interpolation, decentralization, etc.

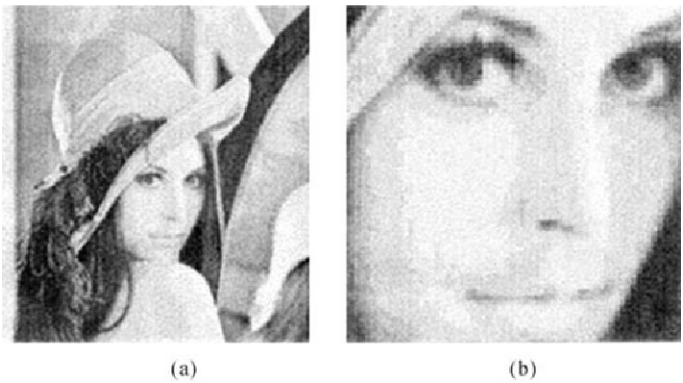


Fig. 2.31. Local enlargement. (a) Original; (b) Face enlargement

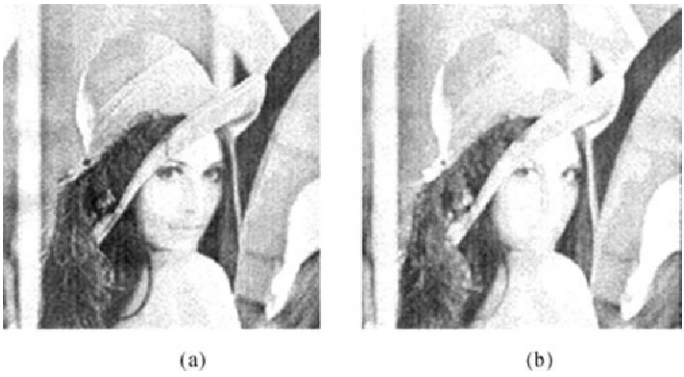


Fig. 2.32. Gray scale conversion. (a) Original; (b) Windowing 3×3

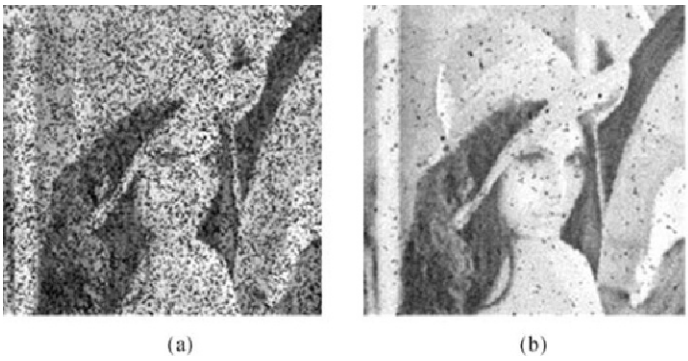


Fig. 2.33. Suppressing interference by median value filtering. (a) Interfered image; (b) After median value filtering

2.5.5 Time / Bearing Display in Digital Sonar

In digital sonar applications, the most frequently used color is gray. Theoretical study and practical application have proved that the gray color display is the most effective display method for a sonar screen. It is recognized that a $2^6 - 2^8$ level gray color display is sufficient to cover all necessary brightness information provided an appropriate dynamic display range is used.

Fig. 2.34 shows how to print the gray color display on white paper (i.e., hard copy), using the halftone technique. The gray scale is expressed in print by the density of print dots whether in an ink-jet or a laser printer.

The data compress technique is used extensively in image processing, especially in underwater image transmission. This is because the high transmission rate needs high frequency signal as its media, but the high frequency signal has more attenuation in underwater propagation. Therefore, decreasing the transmission frequency is a very important task and the data compress technique can make it possible.

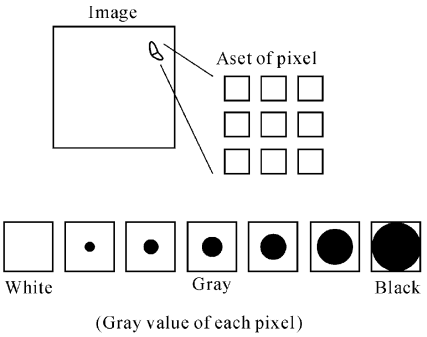


Fig. 2.34. Hardcopy: generation of gray scale values

The IEEE has expert groups for establishing image compress standards. These groups are joint photographic experts group (JPEG) and moving pictures experts group (MPEG).

Before we introduce the various standards for static and moving pictures, we are going to briefly estimate how many pixels there are in one file or picture. For example, if a one-page text file has 25 lines, each line has an average of 10 words, and each words has 8 letters, then the total amount of pixels is 8×10×25 = 2,000 B / page. For a one-frame picture of 256×256 pixels, the total amount of pixels is 256×256 = 65 KB / frame for gray color, and 256×256×3 = 196 KB / frame for full color. Thus, the memory required for storing a picture is much greater than that required for storing a text file.

The standards of JPEG and MPEG are as follows:

(1) JPEG: compression ratio 15:1, DCT coding.
(2) MPEG-1: 320×240 moving image compression ratio 200:1, transmission rate 1.5 Mbps.
(3) MPEG-2: 720×480 moving image compression, transmission rates 4 – 10 Mbps.
(4) MPEG-3: High definition TV moving picture, transmission rate 5 – 20 Mbps.

(5) MPEG-4: Video telephone, transmission rate 9 – 40 Kbps.

The basic display format of digital sonar for signal detection is a time / bearing display, sometimes called a waterfall display. Based on this display the sonar operator can search the target in a particular history plot. This gray color display is suitable for time / bearing displays and is also used for time / frequency or other waterfall type display formats. The number of brightness layers varies from 16 to 256, depending on the application situation and hardware ability. As discussed above, image processing in digital sonar is different from conventional image processing. Sonar displays have some special requirements. Therefore, we need a standard image, something similar to Lenna's photo, to test and evaluate the effect of a digital image processing technique / algorithm to be used in digital sonar. Fig. 2.35(a) shows a basic gray color display in which the layer number is 128. Fig. 2.35(b) is the random noise display. Fig. 2.36 shows a typical time / bearing plot in sonar detection. The horizontal axis is the bearing angle of the sonar system and the vertical axis is time, usually of the magnitude of 30 min.

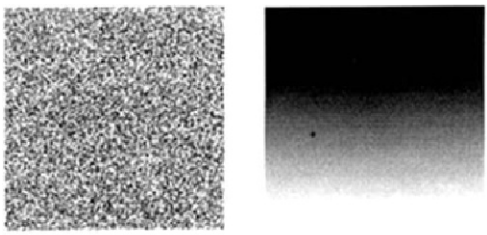


Fig. 2.35. Gray scale display for digital sonar. (a) Gray color display of random value; (b) 128 level gray scale

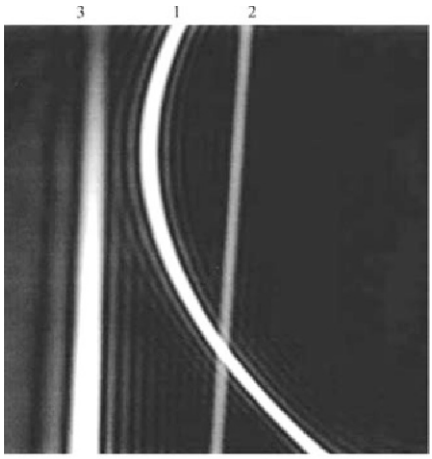


Fig. 2.36. A typical display image for digital sonar

In Fig. 2.36, there are three targets. target 1 is a fast moving contact with a slowly changing incidental angle but fast changing strength. Target 2 is a conventional contact with the bearing angle changing slowly. Target 3 is also a fast moving contact with a fast changing incidental angle but with constant strength. We will use this plot as a standard figure to evaluate whether an image processing technique / algorithm is effective for improving display quality.

2.6 New Topics of Digital Signal Processing: Wavelet Transform and Fractal Transform

Because the application areas of digital signal processing are very extensive and such a wide variety of problems appear in practice, practical requirements drive the development of digital signal processing. A typical example is adaptive filtering theory and artificial neural network theory. Recently, the fractal transform and wavelet transform theories have attracted much interest among researchers and have become popular topics in digital signal processing. Many researches in digital sonar design focus on the application of new developments in fractal transform and wavelet transform and some preliminary results have already been obtained.

In this section, we will introduce the basic concepts and theoretical framework of fractal transform and wavelet transform. Some of these techniques are already playing an important role in many research areas, including speech signal processing, image data compression, etc. We look forward to exploring the potential for applying these transforms to digital sonar design[38-41].

2.6.1 Fractal Transform

The concept of fractal transform was proposed by the American scientist Mandelbrot in 1975. He coined the term "fractal" based on the Latin "fractus", meaning "broken" or "fractured". Mandelbrot studied a class of very common phenomena in the natural world, which are irregular, piecewise but complicated, and have self-similarity and self-affiliation. When he described the concept, he defined a fractal as a shape in which a part of it is similar to the whole in a certain format.

These kind of phenomena are found everywhere, for example, the seashore line, clouds in the sky, a map of human blood vessels, etc. The concept of a fractal is also relevant to chaos theory and has become a powerful mathematical method in chaos study. It is recognized that chaos theory was one of most important new theories developed in the 20th century alongside the Theory of Relativity and the Quantum Theory. The concept of chaos theory describes phenomena that appear to

be random but are not really random. The fractal theory is well suited to solving this kind of theoretical problem.

The theoretical basis of fractal theory is set theory, measure theory and real variable function theory. Before discussing the formal definition and concept of the fractal transform, it is necessary to introduce some preliminary knowledge.

Definition 2.9 (Metric Space) *A metric space* (X, d) *is defined by an abstract space and a function d on X, by which the distance d between two arbitrary points* (*elements*) $x, y \in X$ *satisfies the following three axioms*

Symmetry:

$$d(x, y) = d(y, x), \quad d(x, x) = 0 \tag{2.110}$$

Non-negative:

$$0 < d(x, y) < \infty \tag{2.111}$$

Triangular inequality:

$$d(x, y) \le d(x, z) + d(z, y) \tag{2.112}$$

Example: If X is the set of a real number, $X = R$ = {all real number}, then the following function of d is the metric: $d(x, y) = |x-y|$, $d(x, y) = |x^3-y^3|$.

Definition 2.10 (Equivalent Metric) *If there exist two constants* c_1 and c_2 ($0 < c_1 < c_2 < \infty$), *such that the two metrics* d_1 and d_2 *on X, for any* $x, y \in X$, *we have*

$$c_1 d_1(x, y) \le d_2(x, y) \le c_2 d_1(x, y)$$

then d_1 *and* d_2 *are called equivalent.*

Definition 2.11 (Equivalent Metric Space) *Suppose there are two metric spaces* (X_1, d_1) and (X_2, d_2), *and if there is a one-to-one transform h from* X_1 to X_2 *such that the metrics* $\tilde{d}_1$ *and* d_2 *are defined by*

$$\tilde{d}_1(x, y) = d_2(h(x), h(y)), \qquad x, y \in X_1$$

then (X_1, d_1) *and* (X_2, d_2) *are called equivalent metric spaces.*

Definition 2.12 (Continuous Metric Space) *Suppose there is a transform f defined on metric spaces* (X_1, d_1) and (X_2, d_2), *and if for any given* $\varepsilon > 0$ and $x \in X_1$, *there exists* $\delta > 0$, *such that* $d_2(f(x), f(y)) < \varepsilon$, *when* $d_1(x, y) < \delta$, *then the transform f is called continuous.*

Definition 2.13 (Cauchy Series) *The series* x_n, $(n=1,2, \ldots)$ *in metric space* (X, d) *is called a Cauchy series if for any given* $\varepsilon > 0$, *there exists a positive integer N, such that for any* $n, m > N$, *we have* $d(x_n, x_m) < \varepsilon$.

Definition 2.14 (Limit) *Suppose* x_n, $(n=1,2, \ldots)$ *is a series in metric space*

(X, d), *and if for any given* $\varepsilon > 0$, *there exists a positive integer N, such that for any* $n > N$, *we have* $d(x, x_n) < \varepsilon$, *and* $x \in X$, *then it is said that series* x_n *is convergent to x and x is called the limit of* x_n, *denoted by* $x = \lim_{n\to\infty} x_n$.

Definition 2.15 (Complete Space) *In metric space* (X, d), *if for any Cauchy series* x_n, $(n=1,2, \ldots)$, *there exists a limit in X, then the metric space* (X, d) *is called complete.*

Definition 2.16 (Limit Point) *Suppose S is a subspace in metric space* (X, d), $x \in X$, *if There exists* x_n, $(n=1,2, \ldots)$ *and* $x_n \in S - \{x\}$, *such that* $x = \lim_{n\to\infty} x_n$, *then x is called the limit point of S.*

Definition 2.17 (Closed Set) *Assume* $S \subset X$, *and denote all limit points of* S *by* $\overline{S}$. *If* $S = \overline{S}$, *then* $\overline{S}$ *is called a closed set.*

Example: *Assume X is a real number, then* $S=(0,1)$ *is not a closed set, but* $S=[0,1]$ *is a closed set.*

Definition 2.18 (Compact Set) *Assuming* $S \subset X$, *if any infinite series* x_n $(n=1,2, \ldots)$ *in S has a sub-series, the limit point of which is in S, then S is called compact.*

Definition 2.19 (Boundary Set) *A set* $S \subset X$ *is called boundary, if there exists a point* $a \in X$ *and* $R > 0$, *such that for any* $x \in S$, *we have* $d(x, a) < R$.

Definition 2.20 (Overall Boundary Set) *A set* $S \subset X$ *is called an overall boundary, if for any* $\varepsilon > 0$, *there exists a finite point set* $\{y_1, y_2, \ldots, y_n\} \subset S$, *such that for any* $x \in S$ *there exist some* y_i $(1 \le i \le n)$ *and* $d(x, y_i) < \varepsilon$. *The subset* $\{y_1, y_2, \ldots, y_n\}$ *is called* ε *- net.*

Theorem 2.2 *If* (X, d) *is a complete metric space and* $S \subset X$, *then S is a compact set if and only if S is a closed space and an overall boundary.*

Definition 2.21 (Distance from Point to Set) *Suppose* (X, d) *is a complete metric space, and* $H(X)$ *is the set of all compact subsets of X. For* $B \subset X$, *define the distance from x to B*

$$d(x, B) = \min\{d(x, y) \mid y \in B\} \tag{2.113}$$

Definition 2.22 (Distance between Sets) *Suppose* (X, d) *is a complete metric space,* $A, B \in H(X)$, *the distance between A, B is defined as*

$$d(A, B) = \max\{d(x, B) \mid x \in A\} \tag{2.114}$$

Based on the assumption of a compact set, clearly we can find $\tilde{x} \in A, \tilde{y} \in B$ such that

$$d(A, B) = d(\tilde{x}, \tilde{y}) \tag{2.115}$$

Definition 2.23 (Hausdorff Distance) *Supposing* (X, d) *is a complete metric space,* $A, B \in H(X)$, *then the Hausdorff distance between A, B is defined as*

$$h(A, B) = \max\{d(A, B), d(B, A)\} \tag{2.116}$$

It is proved that h is the metric of $H(X)$, i.e., it satisfies the three axioms for a metric.

Definition 2.24 (Fractal Space) *Suppose* (X, d) *is a complete metric space and h is the Hausdorff metric on* $H(X)$, *then* $(H(x), h)$ *is called the fractal space. A transform in fractal space is called a fractal transform.*

The metric space $(H(X), h)$ is in a sense the natural field in which fractals are found.

Definition 2.25 (Contract Mapping) *A transform f of metric space* (X, d) *is called contract mapping, if there exists* s $(0 \le s < 1)$ *such that for any* $x, y \in X$, *we have*

$$d(f(x), f(y)) \le sd(x, y) \tag{2.117}$$

s *is called the contract factor.*

Theorem 2.3 *Suppose f is the contract mapping factor of metric space* (X, d), *then there is a unique point* $x_f \in X$, *for any other point* $x \in X$, *series* $f^{(n)}(x)$ $(n = 0, 1, 2, ...)$ *convergent to* x_f, *i.e.*

$$\lim_{n\to\infty} f^{(n)}(x) = x_f \tag{2.118}$$

An important result of fractal space is related to contract mapping. Assume there is a group of contract transforms w_n $(n = 1, ..., N)$ on complete metric space (X, d) with contract factor S_n. For any $D \in H(X)$, define

$$w(D) = \bigcup_{n=1}^{N} w_n(D)$$

Where $w_n(D) = \{w_n(x) \mid x \in D\}$ and $s = \max\{s_n | n = 1, \ldots, N|\}$. It is proved that metric space $H(X)$, $h(d)$ is with contract factor s, i.e.

$$h(w(B), w(C)) \le sh(B, C), \quad B, C \in H(X) \tag{2.119}$$

The unique fixed point in $H(X)$ is

$$A = w(A) = \lim_{n\to\infty} w^n(B), \quad \text{for any } B \in H(X) \tag{2.120}$$

w_n is called the IFS (iterated function system) and A is called the attractor.

We have introduced the concept of fractal transform in its abstract form. In the following, we will give some examples of IFS to show the applications of fractal transform.

In traditional computer graphics, the graphic is expressed by geometry components such as points, lines, squares and circles, etc. The more complicated graphics can be expressed by three-dimensional objects such as spheres, cubes, cylinders and shadows, etc. These kinds of components are effective in creating some artificial objects like wheels, roads, buildings, etc. But when facing natural phenomena such as clouds, forests, rainbows and sunsets, it is difficult to simulate the graphics precisely. If we want to store an image of cloud, we have to tell the computer the information about almost every pixel, and the data compression is almost impossible.

The fractal transform can solve this kind of problem easily. We need a basic database consisting of many simple geometric components. Performing the fractal transform on these components will create millions of complicated graphics. The IFS to be used is very limited.

Example

The affine transform

$$W\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

is a contract mapping transform, the contractor is $s = 0.75$. Fig. 2.37 shows how it transforms an upright face into a relatively small face on its side.

Fig. 2.38 is an example of IFS: it moves from a simple geometric graphic to a fern plant image via 8 iterations.

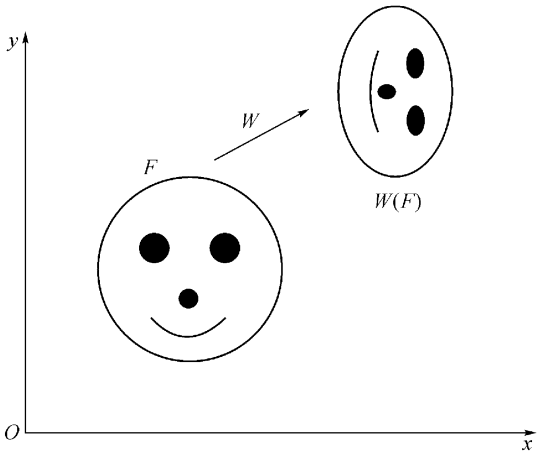


Fig. 2.37. An example of image mapping

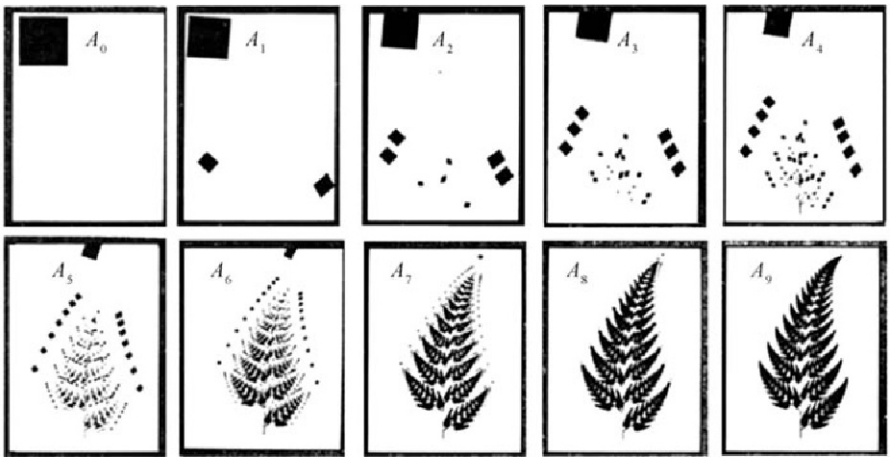


Fig. 2.38. Example of fractal transform

2.6.2 *Wavelet Transform*

The wavelet transform can be traced back to Haar's work in 1909. Mathematicians at that time were looking for the condition in which the Fourier series is convergent. Since as early as 1873, Bois-Reymond had constructed a continuous function which was divergent at a given point[38, 39]. Since then, mathematicians have studied the Fourier series theory in various ways: first, to find the special function classes which can be expressed as Fourier series; second, to try to modify the definition of convergence of Fourier series; third, to search the new orthogonal function class such that the phenomena explored by Bois-Reymond would not occur.

Research in all of the three directions has been beneficial. Research work in the third direction resulted the wavelet transform. However, until the 1970s the development of wavelet theory was limited to the theoretical field and there were no notable results in the application area. In 1986, Meyer and Mallat introduced the wavelet transform in digital image processing and data compression and obtained very encouraging results. As a result, many scientists and engineers in the digital signal processing area were attracted to wavelet transform research. Now, the wavelet transform has become a powerful technique in speech signal processing, feature extraction, image processing, etc. In some ways, the wavelet transform plays a role that Fourier transform cannot perform.

As we know, one of the most important components of signal processing is the signal expression in time and frequency domains. Of course, the Fourier transform is a very powerful method, but cannot tell us at which point in the time axis a signal frequency characteristic will show a specific behavior. Thus, we cannot identify in which segment in the time axis, a signal will have a specific behavior. In this situation, performing a window function is a good choice. Therefore, it is

necessary to use a short time Fourier transform (STFT).

Assume $x(t)$ is an energy limited signal

$$\int |x(t)|^2 \mathrm{d}t < \infty \tag{2.121}$$

The Fourier transform of $x(t)$ is

$$X(f) = \int x(t) \exp(-2\pi \mathrm{j} f t) \mathrm{d}t \tag{2.122}$$

To choose a window function $g(t)$ and a given point τ, the new STFT in this window function is

$$X(f, \tau) = \int x(t) g(t-\tau) \exp(-2\pi \mathrm{j} f t) \mathrm{d}t \tag{2.123}$$

The most popular window function is the Gaussian function

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-t/(2\sigma^2)) \tag{2.124}$$

The STFT with a Gaussian window is called the Gabor transform.

Fig. 2.39 shows a Gaussian function and its spectrum. The time and frequency width σ_t, σ_f of $g(t)$ and $G(f)$ is plotted, where

$$\sigma_t^2 = \frac{\int t^2 |g(t)|^2 \mathrm{d}t}{\int |g(t)|^2 \mathrm{d}t}, \quad \sigma_f^2 = \frac{4\pi \int f^2 |G(f)|^2 \mathrm{d}f}{\int |G(f)|^2 \mathrm{d}f} \tag{2.125}$$

It is proved that

$$\sigma_t \sigma_f \geq 1/2 \tag{2.126}$$

This inequality is called the Heisenberg principle.

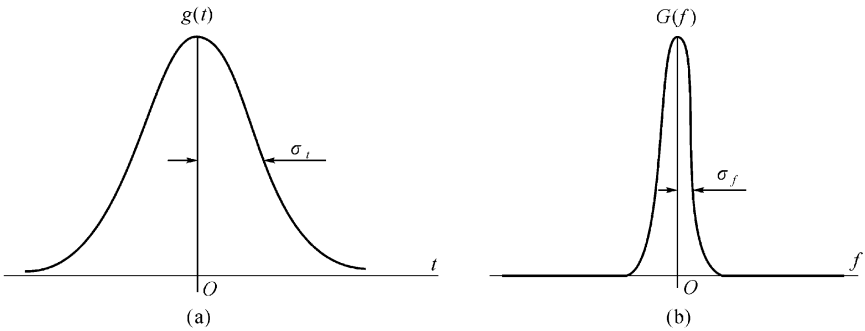


Fig. 2.39. Gaussian window and its Fourier transform

The general expression of the Heisenberg uncertainty principle is very important in the waveform design of radar and sonar and is described below.

Assume $x(t)$ is a square integrated function

$$\int |x(t)|^2 \mathrm{d}t < \infty$$

Let

$$X(f) = \int x(t) \exp(-2\pi \mathrm{j} f t) \mathrm{d}t, \quad \| x \|^2 = \int |x(t)|^2 \mathrm{d}t$$

Define

$$\mu = \int t \, |x(t)|^2 \, \mathrm{d}t / \| x \|^2, \quad \sigma = \int f \, |X(f)|^2 \, \mathrm{d}f / \| x \|^2,$$

$$\sigma_t^2 = \int (t-\mu)^2 \, |x(t)|^2 \, \mathrm{d}t / \| x \|^2, \quad \sigma_f^2 = \int (f-\sigma)^2 \, |X(f)|^2 \, \mathrm{d}f / \| x \|^2$$

Then

$$\sigma_t^2 \sigma_f^2 \geq 1/4 \tag{2.127}$$

Fig. 2.40 shows the results of a Gaussian window function modulated by different high frequency signals. The spectrum shifts in the frequency axis, but the time-frequency resolution of STFT stays the same (Fig. 2.41), i.e., for the Fourier transform the time-frequency resolution is the same regardless of the frequency in which the Gaussian window function is modulated.

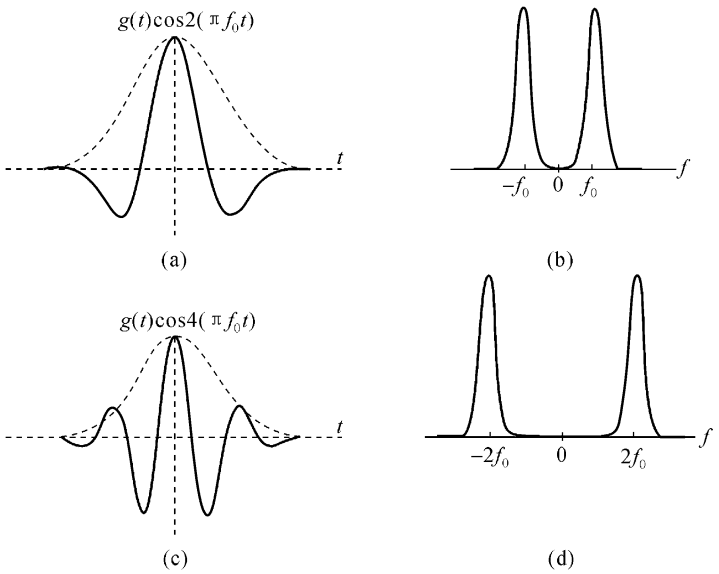


Fig. 2.40. Gaussian window modulation and its spectrum

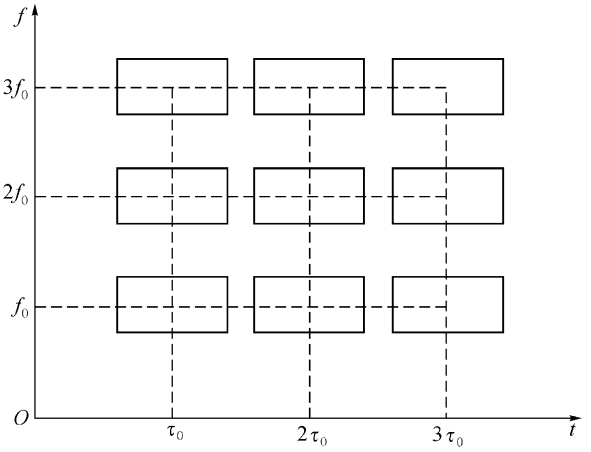


Fig. 2.41. Plane of time-frequency resolution ability of Fourier transform

As we know, the frequency resolution of DFT is uniform in the frequency axis. It depends on the sampling frequency f_s and the number of samples N, i.e., $\Delta f = f_s / N$. The wavelet transform provides flexibility for frequency resolution. In the low frequency region, the time resolution is low but the frequency resolution is high; in the high frequency region, the time resolution is high but the frequency resolution is relative low. This kind of behavior is useful for the analysis of short time duration signals.

The basic function class of a wavelet $v(t)$ is called the original function or generation function. To take time shift and scale transform on $v(t)$ obtains

$$v_{ab}(t) = \frac{1}{a} v((t-b)/a),\ a > 0$$

The wavelet of any function is defined by

$$X(a,b) = \int v_{ab}(t) x(t) \mathrm{d}t \tag{2.128}$$

For a large a the generation function is enlarge type and for the small a, the generation function is compression type. By adjusting the value of a, it is easy to control the enlargement or compression of the generation function in the time axis, which will result in a change of spectrum in the frequency axis (Fig. 2.42).

Suppose the Fourier transform of function $v(t)$ is $V(f)$, therefore the Fourier transform of $v_{ab}(t)$ is

$$V_{ab}(f) = \sqrt{a} V(f) \exp(-2\pi \mathrm{j} b f) \tag{2.129}$$

By using the Fourier transform of a wavelet, any signal $x(t)$ can be reconstructed

$$x(t) = \frac{1}{c_v} \int\int \frac{1}{a^2} X(a,b) v_{ab}(t) \mathrm{d}a\mathrm{d}b \tag{2.130}$$

where

$$c_v = \int_0^\infty \frac{|V(f)|^2}{f} \mathrm{d}f \tag{2.131}$$

The most popular wavelet function in digital signal processing is

$$\begin{aligned} v(t) &= (1-t^2)\exp(-t^2/2) \\ V(f) &= 2\pi\sqrt{2\pi} f \exp(-(2\pi f)^2/2) \end{aligned} \tag{2.132}$$

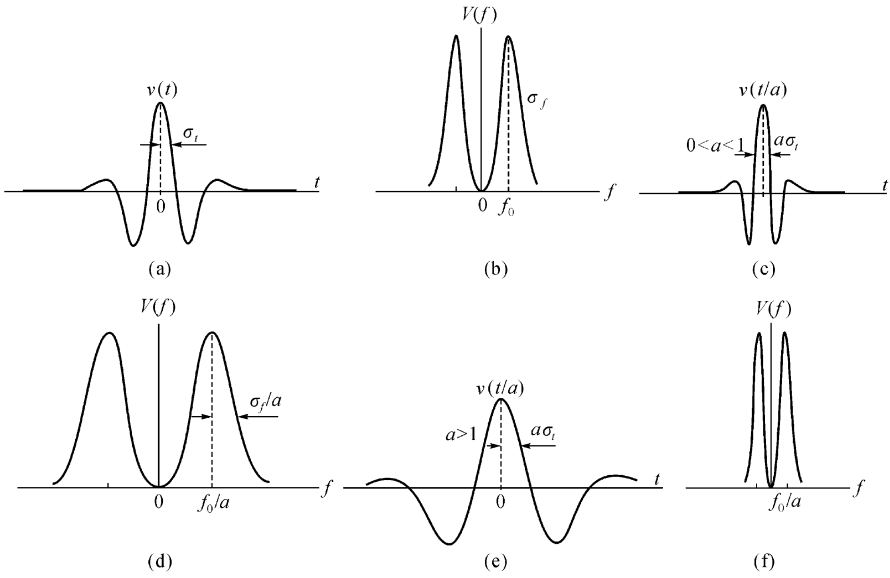


Fig. 2.42. Wavelet function and its Fourier transform

Fig. 2.43 illustrates a wavelet transform in the time-frequency plane. The time-frequency resolution unit is not a fixed shape: although the area value stays the same, the length and width of the rectangle can be adjusted.

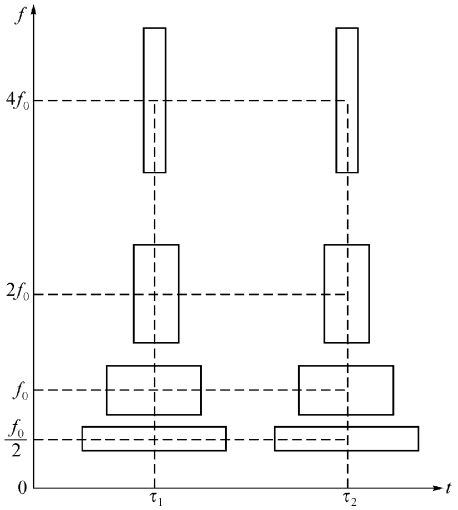


Fig. 2.43. Plane of time frequency resolution ability of wavelet transform

References

[1] Oppenheim, A. V. (Eds): Applications of Digital Signal Processing. Prentice Hall, Englewood (1978)
[2] Proakis, J. G.: Advanced Digital Signal Processing. Maxwell Macmillan Intl., New York (1992)
[3] Elliott, D. F. (Eds): Handbook of Digital Signal Processing, Engineering Applications. Rockwell Intl. Corp., Costa Mesa (1986)
[4] Chen, C. H. (Eds): Signal Processing Handbook. Marcel Dekker Inc., New York (1988)
[5] Poularikas, A.: Handbook of Formulas and Tables for Signal Processing. CRC Press, Boca Raton (1999)
[6] Poularikas, A. (Eds): The Transform and Applications Handbook. CRC Press, Boca Raton (1996)
[7] Srinath, M. D., Rajasekran, P. K.: An Introduction of Signal Processing with Applications. Wiley, New York(1979)
[8] Madisetti, V. K., Willians, D. B. (Eds): The Digital Signal Processing Handbook. CRC Press, Boca Raton (1999)
[9] Wiener, N.: Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications. Wiley, New York (1949)
[10] Franks, L. E.: Signal Theory. Prentice Hall, New Jersey (1969)
[11] Turin, G. L.: An introduction to matched filters. IRE Trans. IT-6, pp. 311-330 (1960)
[12] Papoulis, A.: The Fourier Tntegral and its Applications. McGraw-Hill, New

York (1963)
[13] Bracewell, R. N.: Aspects of the Hartley transform. In: Proc. of IEEE, 82-3, pp. 381-387 (1994)
[14] Knight, W. C., Pridham, R. G., Kay, S. M.: Digital signal processing for sonar. In: Proc. of IEEE 69, pp. 1451-1506(1981)
[15] Li, Q. H.: An Introduction to Sonar Signal Processing. Ocean Publ., Beijing (1985)
[16] Papamichalis, P. E. (Eds): Digital Signal Processing Applications with TMS320 family, Vol.3. Prentice Hall, Englewood (1990)
[17] Higgins, R. J.: Digital Signal Processing in VLSI. Prentice Hall, Englewood (1990)
[18] Abarbanel, H.: Obtaining order in a world of chaos. In: IEEE Signal Processing, 15-3 pp.49-66 (1998)
[19] Peled, A., Liu, B.: A new hardware realization of digital filters. IEEE Trans. ASSP-22, pp. 456-462 (1976)
[20] Mazor, S.: The history of the microcomputer–invention and evolution. In: Proc. of IEEE 83-12, pp. 1601-1608 (1995)
[21] Bier, J.: Selecting the right tools for DSP processor software development. DSP and Multimedia Technology 3-7, pp. 55-62 (1994)
[22] Price, W. J.: A benchmark tutorial. IEEE Micro. 9-5, pp. 28-43 (1989)
[23] Davis, A.: DSPs and the age of specification. DSP and Multimedia Technology 3-7, pp. 20-32 (1994)
[24] Kung, S. Y. (Eds): VLSI and Modern Signal Processing. Prentice Hall, Englewood (1985)
[25] Lapsley, P., Bior, J.: Choosing DSP processor. DSP Applications 3-1, pp. 24-32 (1994)
[26] Murphy, B. T.: Microcomputer: trends, technologies, and design strategies. IEEE J. Solid State Circuit SC-SC-18-3, pp. 236-244 (1983)
[27] Ralston, A.: Encyclopedia of Computer Science Engineering, 2nd Edn. Van Nostrand, New Jersey (1985)
[28] Bendat, J. L., Piersol, A. G.: Engineering Applications of Correlation and Spectral Analysis. Wiley, New York (1980)
[29] Proakis, J. G.: Digital Communications, 3rd Edn. McGraw-Hill, New York (1997)
[30] Pandharipande, A.: Information, uncertainty and randomness. IEEE Potentials Oct./Nov., pp. 32-34 (2002)
[31] Cooley, J. W., Tukey, W.: An algorithm for machine calculation of complex fourier series. Math. Comput. 19-2, pp. 297-301 (1965)
[32] Bracewell, R. N.: The Fourier Transform and its Applications. Mc-Graw Hill, New York (1986)
[33] Brigham, E. O.: The Fast Fourier Transform. Prentice Hall, Englewood (1974)
[34] Heideman, M. T. et al.: Gauss and history of the fast fourier transform. IEEE ASSP Magazine Oct., pp. 14-21 (1984)

[35] Lindley, C. A.: Practical Image Processing in C. Wiley, New York (1991)
[36] Pratt, W. K.: Digital Image Processing. Wiley, New York (1978)
[37] Lim, J. S.: Two Dimensional Signal and Image Processing. Prentice Hall, Englewood (1990)
[38] Rion, O., Vetterli, M.: Wavelets and signal processing. IEEE SP Magazine 10, pp. 14-38 (1991)
[39] Shensa, M. J.: The discrete wavelet transform: wedding the Trous and Mallat algorithm. IEEE Trans. SP-40, pp. 2462-2482 (1992)
[40] Fanily, F.: Fractals. DSP Applications 2-12, pp. 50-58 (1993)
[41] Barnsley, M. F.: Fractals Everywhere. Wiley, New York (1990)
[42] Papoulis, A.: Signal Analysis, 2nd Edn. McGraw-Hill, New York (1993)
[43] Members of the technical Staff at Bell Labs (Eds): Transmission Systems for Communications, 5th Edn. Bell Laboratory, Winston-Salem (1982)
[44] Poularikas, A.: Handbook of Formulas and Tables for Signal Processing. CRC Press, Boca Raton (1999)
[45] Gallagher, N. C. Jr.: Median filter: a tutorial. In: IEEE Proc. of Intl. Symp. Circuit and System, pp. 1737-1744 (1988)
[46] Struzinski, W. A.: A performance comparison of four noise background normalization schemes proposed for signal detection. J.Acoust. Soc. Amer. 76-6, pp. 1738-1742 (1984)

3

Detection and Estimation Theory of Digital Signals

Signal / system and detection / estimation theory is the most important theoretical and applied basis of modern information theory.

Distinguishing signal from background noise and estimating the parameters of detected signal are among the most important objectives of radar, sonar and communication systems[1-25]. In particular, interference suppression in the detection of weak signal is the main task of many applications.

In sonar signal processing, the first objective of active or passive sonar is to identify the existence of the target. For active sonar, after the transmission of an active sonar signal, the receiver waits to receive the echo signal from the target. But the echo signal is usually distorted and suffers interference from various noises. For passive sonar, because the radiated noise from the target is unpredictable, the detection of signal is more uncertain[11-20,26-32].

The estimation of signal parameters is also an important task in sonar signal processing. The parameters include information about signal amplitude, phase, frequency and time delay. The randomness of signal and background noise make parameter estimation more difficult. It is possible to estimate the signal parameters with a certain level of accuracy. Therefore, probability theory and mathematical statistics have become the theoretical basis of signal detection and estimation theory[33-38].

In this chapter, the traditional theories of signal / system and detection / estimation are described with some examples of applications. The basic framework is "detection before estimation" rather than "joint detection and estimation" or so-called "estimation before detection"[39-42].

3.1 Some Basic Results from Probability Theory and Mathematical Statistics

Probability theory and mathematical statistics are basic branches of mathematics: the contents of these two branches are extensive and profound. It is impossible to

introduce the framework of this theory in this section, or even the most important results. The only aspects that we are going to present are the basic conclusions from some study areas that are already used in signal detection and estimation. The reader can then understand how to use these results in sonar signal processing and sonar design. Readers who are interested in the detailed theoretical background of probability theory and mathematical statistics can read a relevant monograph in this area. Probability theory and mathematical statistics are essential courses for undergraduate or graduate students who wish to study signal processing and its applications.

3.1.1 Basic Definition of Probability

Probability theory is a branch of mathematics concerned with the analysis of random phenomena. The central objects of probability theory are random variables, stochastic processes, and events: mathematical abstractions of non-deterministic events or measured quantities that may either be single occurrences or evolve over time in an apparently random fashion. As a mathematical foundation for statistics, probability theory is essential to many human activities that involve quantitative analysis of large sets of data. Methods of probability theory are also applied in the description of complex systems given only partial knowledge of their state, as in statistical mechanics. A great discovery of 20th century physics was the probabilistic nature of physical phenomena at atomic scales, as described in quantum mechanics[43]. Mathematical statistics is the study of statistics from a purely mathematical standpoint, using probability theory as well as other branches of mathematics such as linear algebra and analysis. It deals with gaining information from data. In practice, data often contain some randomness or uncertainty. Statistics handles such data using methods of probability theory.

In the natural world, under certain conditions the outcomes of many events can happen or not happen. This kind of event is called a random event. For example, a coin with two faces: value and image. In each throw, the value face and image face have an equal chance of facing upwards but it is impossible to predict which face will face upwards in a single throw. Therefore the event "value face upward in one throw" is a random event.

Denote the all possible outcomes of an experiment by X: this is called the sample space. Each point in X, i.e., one possible outcome, is called a basic event. A random event is a set consisting of some of the points. Therefore, set theory in mathematics can be used to describe and discuss random events.

Probability of an Event

Probability is a quantity used to describe the value of the possibility of occurrence of an event. The precise definition of probability depends on the application area.

To introduce different definitions of probability, it is necessary to describe three basic axioms about the probability of an event.

Definition 3.1 *A measure $P(\cdot)$ in sample space X is a real value function of a sub-set, and satisfies the following three axioms*:

(1) *For each set*

$$A \subset X,\ P(A) \geq 0 \tag{3.1}$$

(2) *If* $A, B \subset X$, *and* $A \cap B = \phi$ *(i.e., the intersection of A and B is an empty set), then*

$$P(A + B) = P(A) + P(B) \tag{3.2}$$

(3) *The measure of sample space X is equal to 1*, i.e., $P(X) = 1$.

The goal in considering the probability of an event as a measure in an abstract sample space is to establish the serious theory architecture for the probability theory. In the history of the development of probability theory, several different definitions were proposed for use in calculating the probability of an event. These definitions meet the specific requirements of the application situation, which still play an important role in probability calculation.

Classic Definition of Probability

Suppose the sample space X consists of N points, event A contains n points, and the probability of each point is equal, then the probability of event A is

$$P(A) = n / N \tag{3.3}$$

Example

Suppose the number of a set of hydrophones is $N = 550$ and there are n=28 unqualified hydrophones. Taking one hydrophone arbitrarily, the probability of choosing an unqualified product is $28 / 550 = 0.0509$.

Geometry Definition of Probability

If the sample space consists of infinite points, it can be expressed as a two-dimensional graph (Fig. 3.1); if event A is expressed as the shaded part, then the probability

$$P(A) = \frac{\text{Area of } A}{\text{Area of } X} \tag{3.4}$$

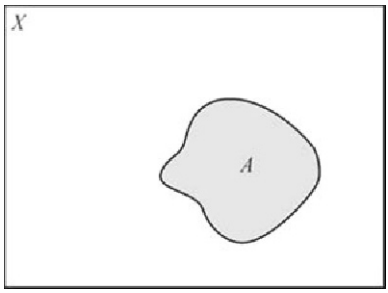


Fig. 3.1. Illustration of a geometrical definition of probability

Example

Assume that a single frequency signal has random phase ϕ, and the sample space is $[0,\ 2\pi]$, then the probability of ϕ falling in segment $[\pi/6,\ \pi/3]$ is equal to

$$\frac{\pi/3-\pi/6}{2\pi}=\frac{1}{12}$$

Statistical Definition of Probability

Suppose there are N independent experiments under the same conditions, and the number of occurrences of some event A is $n(A)$, we define the occurrence probability of event A in one experiment as it is easy to verify that the previous three definitions about probability satisfy the three axioms of probability.

$$P(A)=\lim_{N\to\infty}\frac{n(A)}{N} \tag{3.5}$$

Based on the three axioms of probability, some useful operational rules can be derived:

(1) The probability of any event A is always no greater than 1, i.e., $P(A)\le 1$.

(2) If $A\subset B$, then $P(A)\le P(B)$.

(3) For any N mutual non-intersect sets $A_1,\ ...,\ A_N$

$$P(A_1+A_2+...+A_N)=P(A_1)+P(A_2)+...+P(A_N)$$

(4) For any sets A and B

$$P(A+B)=P(A)+P(B)-P(AB) \tag{3.6}$$

Note that sometimes we use "event A" and other times we prefer to use "set A", depending on the situation. In fact, these expressions are equivalent, but the understanding is different. For example, AB can be understood to be the intersection set of A and B, and can also be understood to be the events A and B happening simultaneously.

3.1.2 Conditional Probability

Definition 3.2 *Suppose B is a set of sample space X and $P(B) > 0$ in the condition that B has already happened, then the conditional probability of event A occurring is defined as*

$$P(A \mid B) = P(AB) / P(B) \tag{3.7}$$

Theorem 3.1 (Product Theorem) *Suppose $P(B) > 0$, then*

$$P(AB) = P(B)P(A \mid B) \tag{3.8}$$

Theorem 3.2 (Total Probability Formula) *If $B_1, B_2, ..., B_n$ is n mutual non-intersect sets and $B_1 + B_2 + ... + B_n = X$, then for any $A \subset X$*

$$P(A) = P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + ... + P(B_n)P(A \mid B_n) \tag{3.9}$$

Theorem 3.3 (Bayes Formula) *If $B_1, B_2, ..., B_n$ is n non-mutual intersect sets and $B_1 + B_2 + ... + B_n = X$, then for any $A \subset X$, $P(A) > 0$, we have*

$$P(B_i A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{k=1}^{n} P(A \mid B_k)P(B_k)} \qquad (i = 1, ..., n) \tag{3.10}$$

Example

Assume workers A and B in turn operate a machine, and that the failure probability is 0.1 when A operates it and 0.2 when B operates it. Now if the machine has failed to work, what is the probability that it has failed as a result of A operating the machine?

Let B_1 = {A operates this machine}, B_2 = {B operates this machine} and A = {machine failure}. According to the above assumption,

$$P(B_1) = P(B_2) = 1/2, \quad P(A \mid B_1) = 0.1, \quad P(A \mid B_2) = 0.2$$

the probability to be found is

$$P(B_1 \mid A) = P(B_1)P(A \mid B_1) / P(A) = 1/3$$

Definition 3.3 (The Independence of Random Events) *If the occurrence of random event B does not affect the probability of occurrence of random event A, then the random events A and B are called independent. The mathematical expression of this definition is*

$$P(A) = P(A \mid B) \quad \text{or} \quad P(AB) = P(A)P(B) \tag{3.11}$$

3.1.3 Random Variable and Distribution Function

Definition 3.4 (Random Variable) *The real or complex value function* $\xi(\omega)$ *in sample space X is called the random variable, where* $\omega \in X$.

If there exists a measure $P(\cdot)$ in sample space X, in which random variable $\xi(\omega)$ is defined, then any random events corresponding to $\xi(\omega)$ can be given a probability.

Sometimes we omit the argument ω of $\xi(\omega)$ as ξ, if there is no difficulty in understanding random variable $\xi(\omega)$.

Definition 3.5 (Distribution Function) *Suppose* $\xi(\omega)$ *is the random variable in sample space X and* $P(\cdot)$ *is the probability measure in x, then*

$$F(x) = P\{\xi(\omega) \le x \mid x \in R\} \tag{3.12}$$

is called the distribution function of $\xi(\omega)$, where R denotes the real value set.

The distribution functions $F(x)$ for any random variable have the following three basic specifications:

(1) Monotonic non-decreasing: $F(b) \ge F(a)$, if $b > a$.

(2) Right continuous: $\lim_{x \to a^+} F(x) = F(a)$.

(3) $F(-\infty) = 0,\ F(\infty) = 1$.

Definition 3.6 (Discrete and Continuous Random Variable) *If the value set of random variable* $\xi(\omega)$ *is finite or countable, then* $\xi(\omega)$ *is called a discrete random variable. The value set is denoted as* $\{a_1, a_2, ..., a_n\}$, *and the probability* $p_i = P\{\xi(\omega) = a_i\}$ $(i = 1, 2, ..., n)$ *is called the distribution of* $\xi(\omega)$. *It is obvious that*

$$\sum p_i = 1$$

where the symbol of summation is taken as the set of all $\{p_i\}$, *and the number can be finite or countable infinite.*

If the distribution function of random variable $\xi(\omega)$ can be expressed by the integration of a non-negative function $p(x)$

$$F(x) = \int_{-\infty}^{x} p(t)\mathrm{d}t \tag{3.13}$$

and

$$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$$

then the random variable $\xi(\omega)$ is called a continuous random variable and the function $p(x)$ is called a distribution density function or a probability density function of random variable $\xi(\omega)$.

Examples

(1) Bernoulli binomial distribution. Assume that *E* represents a random experiment, and that the probability of outcome *A* of each experiment is *p*, i.e., $P(A) = p$. Repeating *n* times the experiment *E*, the probability that event *A* appears exactly *k* times ($k \le n$) is

$$P\{\xi(\omega) = k\} = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0,\ 1,\ ...,\ n) \tag{3.14}$$

where $\xi(\omega)$ represents the number of outcomes of event *A* in *n* experiments. Clearly, $\xi(\omega)$ is a discrete random variable, and its value set is {0, 1, ..., *n*}.

(2) Uniform distribution. Assume $-\infty < a < b < \infty$

$$p(x) = \begin{cases} 1/(b-a) & x \in [a,\ b] \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

is called a uniform distribution density and the distribution function is

$$F(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \le x < b \\ 1 & x \ge b \end{cases}$$

(3) Normal or Gaussian distribution. The distribution density function of a normal

(or Gaussian) random variable is

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{3.16}$$

Its distribution function is

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)\mathrm{d}t$$

Fig. 3.2 illustrates the distribution density function and the distribution function of uniform and Gaussian random variables.

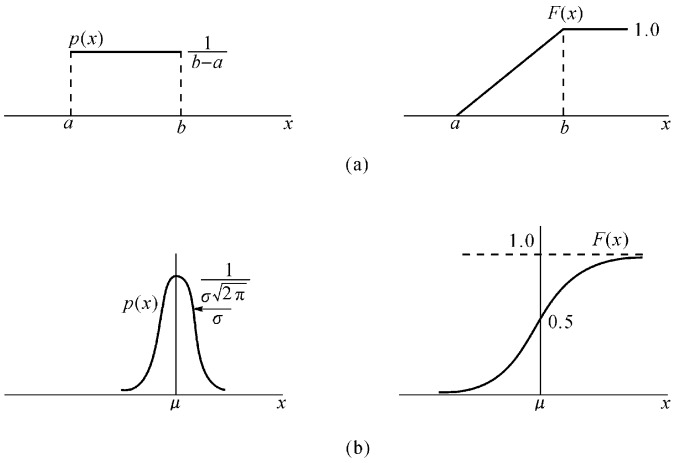


Fig. 3.2. Probability density function of a continuous random variable. (a) Uniform distribution; (b) Gaussian distribution

The function of random variables and multi-dimensional random variables

If $f(x)$ is an ordinary function and $\boldsymbol{\xi}(\omega)$ is a random variable, then $f(\boldsymbol{\xi}(\omega))$ is also a random variable.

In terms of several random variables we can construct a random vector, i.e., a multi-dimensional random variable. For example, based on the random variables $\boldsymbol{\xi}_1(\omega)$, $\boldsymbol{\xi}_2(\omega)$, ..., $\boldsymbol{\xi}_n(\omega)$, we can construct an n dimensional random vector where the symbol T denotes the transpose operation of the vector or matrix.

$$\boldsymbol{\xi}(\omega)=[\boldsymbol{\xi}_1(\omega),\ \boldsymbol{\xi}_2(\omega),\ ...,\ \boldsymbol{\xi}_n(\omega)]^{\mathrm{T}}$$

The distribution function of an n dimensional random variable is an n variable function

$$F(x_1, x_2, ..., x_n) = P\{\xi_1(\omega) \le x_1, \xi_2(\omega) \le x_2, ..., \xi_n(\omega) \le x_n\} \tag{3.17}$$

Definition 3.7 (Independence of Random Variables) *Two random variables* $\xi_1(\omega)$, $\xi_2(\omega)$ *in sample space X are called independent, if for any* x_1, x_2 *the events* $\{\xi_1(\omega) \le x_1\}$ *and* $\{\xi_2(\omega) \le x_2\}$ *are independent. For the independent random variable,* $p(x_1, x_2) = p_1(x_1)p_2(x_2)$ *, where* $p_1(x)$ *is the distribution density function of variable* $\xi_1(\omega)$*, and* $p_2(x)$ *is the distribution density function of variable* $\xi_2(\omega)$*.*

3.1.4 Digital Characteristics of Random Variables

There are several digital characteristics of random variables, which together describe important aspects of their behavior. Sometimes in comparing different random variables one examines only these digital characteristics and neglects the probability distribution function or density function.

Definition 3.8 (Mean Value) *Suppose* $\xi(\omega)$ *is a discrete type of variable, the probability distribution is*

$$a_1, a_2, ...$$
$$p_1, p_2, ...$$

The mean value (or mathematical expectation) of $\xi(\omega)$ is defined as

$$E(\xi) = \sum p_i a_i \tag{3.18}$$

where the symbol of summation is taken as the set of all $\{p_i a_i\}$. The number can be finite or countable infinite.

If $\xi(\omega)$ is a continuous random variable and the distribution density function is $p(x)$, then its mean value is defined as

$$E(\xi) = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x \tag{3.19}$$

Definition 3.9 (Variance of Random Variable) *The variance of random variable* ξ *is defined as the expectation value of random variable* $[\xi - E(\xi)]^2$*, i.e.*

$$\mathrm{Var}(\boldsymbol{\xi}) = E\,[\boldsymbol{\xi} - E(\boldsymbol{\xi})]^2 \tag{3.20}$$

For the discrete random variable

$$\mathrm{Var}(\boldsymbol{\xi}) = \sum p_i (a_i - E(\boldsymbol{\xi}))^2 \tag{3.21}$$

For the continuous random variable,

$$\mathrm{Var}(\boldsymbol{\xi}) = \int_{-\infty}^{\infty} (x - E(\boldsymbol{\xi}))^2 p(x)\mathrm{d}x \tag{3.22}$$

There is another notation for variance of a random variable, that is $D(\xi)$, i.e.

$$D(\boldsymbol{\xi}) = \mathrm{Var}(\boldsymbol{\xi})$$

Some Useful Formulas for the Operation of Mean Value and Variance

(1) A constant is a special random variable. The mean value of a constant is equal to itself, and the variance of a constant is equal to zero, i.e.

$$E(c) = c, \quad \mathrm{Var}(c) = 0$$

(2) The mean value of summation of two random variables is equal to the summation of the mean values of two random variables, i.e.

$$E(\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2) = E(\boldsymbol{\xi}_1) + E(\boldsymbol{\xi}_2) \tag{3.23}$$

where ξ_1 and ξ_2 are any two random variables.

(3) The variance of a random variable is equal to the difference between the mean value of the square of a random variable and the square of the mean value of this random variable, i.e.

$$\mathrm{Var}(\boldsymbol{\xi}) = E(\boldsymbol{\xi}^2) - (E(\boldsymbol{\xi}))^2 \tag{3.24}$$

(4) The mean value of the product of two independent random variables is equal to the product of the mean values of two random variables, i.e.

$$E(\boldsymbol{\xi}_1\boldsymbol{\xi}_2) = E(\boldsymbol{\xi}_1)E(\boldsymbol{\xi}_2) \tag{3.25}$$

If ξ_1 and ξ_2 are independent.

(5) The variance of summation of two independent random variables is equal to the summation of the variance of two random variables, i.e.

$$\mathrm{Var}(\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2) = \mathrm{Var}(\boldsymbol{\xi}_1) + \mathrm{Var}(\boldsymbol{\xi}_2) \tag{3.26}$$

If ξ_1 and ξ_2 are independent.

(6) If $f(x)$ is a real function, and ξ is a continuous random variable with probability density function $p(x)$, then

$$E(f(\boldsymbol{\xi})) = \int_{-\infty}^{\infty} f(x)p(x)\mathrm{d}x \tag{3.27}$$

(7) The square of the mean value of a random variable is equal to or less than the mean value of the square of the random variable, i.e.

$$(E(\boldsymbol{\xi}))^2 \le E(\boldsymbol{\xi}^2) \tag{3.28}$$

(8) Schwartz inequality: if ξ_1 and ξ_2 are two random variables, then

$$|E(\boldsymbol{\xi}_1\boldsymbol{\xi}_2)| \le [E(\boldsymbol{\xi}_1^2)E(\boldsymbol{\xi}_2^2)]^{1/2} \tag{3.29}$$

The proof of formulas above can almost be derived directly from the basic definitions and calculus.

Example

Suppose the random variable ξ obeys the distribution given by Eq. (3.15), then

$$E(\boldsymbol{\xi}) = (a+b)/2, \quad \mathrm{Var}(\boldsymbol{\xi}) = (b-a)^2/12 .$$

If ξ obeys the Gaussian distribution as shown in Eq.(3.16), then

$$E(\boldsymbol{\xi}) = \mu, \quad \mathrm{Var}(\boldsymbol{\xi}) = \sigma^2$$

Definition 3.10 (Covariance and Correlation Coefficient) *The covariance of two random variables* ξ and η *is defined as*

$$\mathrm{cov}(\boldsymbol{\xi},\eta) = E[(\boldsymbol{\xi} - E(\boldsymbol{\xi}))(\eta - E(\eta))] \tag{3.30}$$

The normalized covariance is called the correlation coefficient of random variables ξ and η

$$\rho(\boldsymbol{\xi},\eta) = \frac{\mathrm{cov}(\boldsymbol{\xi},\ \eta)}{\sqrt{\mathrm{Var}(\boldsymbol{\xi})\mathrm{Var}(\eta)}} \tag{3.31}$$

Based on Schwartz inequality

$$|\operatorname{cov}(\boldsymbol{\xi},\eta)\le[\operatorname{Var}(\boldsymbol{\xi})\operatorname{Var}(\eta)]^{1/2}$$

Clearly, $|\rho(\boldsymbol{\xi},\ \eta)|\le 1$.

Examples

(1) Two-dimensional Gaussian distribution.

Suppose random vector (ξ, η) obeys a two-dimensional Gaussian distribution

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-2\frac{1}{1-\rho^2}\left[\frac{(x-a)^2}{\sigma_1^2}-2\rho\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}$$

It can be shown that $\operatorname{cov}(\xi, \eta) = \sigma_1\sigma_2\rho$.

(2) Raleigh distribution.

Suppose ξ_1, ξ_2 are two independent random variables and both obey the Gaussian probability distribution with a mean value 0 and a variance σ^2 We are going to find the probability distribution function of random variable $\eta=(\boldsymbol{\xi}_1^2+\boldsymbol{\xi}_2^2)^{1/2}$. It is easy to show that

$$F(x)=P\{\eta\le x\}=\iint\limits_{x_1^2+x_2^2}\frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}(x_1^2+x_2^2)\right]\mathrm{d}x_1\mathrm{d}x_2\quad(x>0)$$

Let $x_1=r\cos\theta$, $x_2=r\sin\theta$, replace it in the above integration and take integral transform. It is then straightforward to calculate that

$$\begin{aligned}&F(x)=1-\exp(-x^2/(2\sigma^2))\quad(x>0)\\&p(x)=F'(x)=\begin{cases}0 & x<0\\ \dfrac{x}{\sigma^2}\exp(-x^2/2\sigma^2) & x\ge 0\end{cases}\end{aligned}\tag{3.32}$$

The probability density function in Eq. (3.32) is called the Rayleigh distribution. It is actually the distribution density of the envelope of the square sum of two Gaussian random variables. It is useful in many practical application problems.

3.1.5 *Large Number Law and Central Limit Theorem*

As we know, the probability rule is applied in the study of massive random phenomena. To explore the probability rule one should repeat an experiment many times under the same conditions, or under certain conditions observe or study

massive random phenomena. This work results in the so-called large number law and central limit theorem.

Theorem 3.4 (Chebychev Large Number Law) *Suppose* $\boldsymbol{\xi}_k$, $(k = 1, 2, ..., n)$ *is an independent random variable series with a constant c such that* $\mathrm{Var}(\boldsymbol{\xi}_k) \le c$, *for any k. Then for any given* $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{\xi}_k - E(\boldsymbol{\xi}_k))\right| \ge \varepsilon\right\} = 0 \qquad (3.33)$$

Theorem 3.5 (Bernoulli Large Number Law) *In the Bernoulli binomial distribution of Eq. (3.14), if the probability of occurrence of event A is p* ($0 \le p \le 1$), *and* ξ_n *denotes the number of occurrences of event* A_n *in the first n times of independent tests, then for any given* $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{\boldsymbol{\xi}_n}{n} - p\right| \mid \varepsilon\right\} = 0 \qquad (3.34)$$

Theorem 3.6 (Khinchine Large Number Law) *Suppose* ξ_n $(k = 1, 2, ..., n)$ *is an independent random variable with the same probability distribution and* $E(\xi_k) = \mu$ $(k = 1, 2, ..., n)$, *then for any given* $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\xi}_k - \mu\right| \ge \varepsilon\right\} = 0 \qquad (3.35)$$

Although the format of the above three large number laws is different, they essentially describe a basic fact, that when one repeats an experiment independently under the same conditions, then the probability that the mean value will be biased is very small when the number of experiments becomes large.

Theorem 3.7 (Central Limit Theorem) *Suppose* $\boldsymbol{\xi}_k$, $(k=1, 2, ..., n)$ *is an independent random variable with the same probability distribution and* $E(\xi_k) = \mu$, $\mathrm{Var}(\xi_k) = \sigma^2$, $(k = 1, 2, ..., n)$ *then*

$$\lim_{n\to\infty} P\left\{\frac{1}{\sqrt{n}\sigma}\sum_{k=1}^{n}(\xi_k - \mu) \le x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-z^2/2)\mathrm{d}z \qquad (3.36)$$

Central limit theorem has a very important meaning in theory and application: the average value of an independent random variable series with the same probability distribution function will eventually converge to the Gaussian probability distribution, regardless of what were the original probability functions.

3.1.6 Random Process (Stochastic Process)

Definition 3.11 *A set of random variables in sample space X is called random process. Strictly speaking, the random variable should expressed as* $x(\omega, t)$ $(\omega \in X,\ t \in R)$, *where R is the real value set. For any fixed* $\omega_0 \in X$, $x(\omega_0, t)$ *is an ordinary function, real or complex; and for any fixed* $t_0 \in R$, $x(\omega, t_0)$ *is a random variable defined in abstract sample space. In signal processing theory, we usually omit the argument* ω, *and indicate the random process as* $x(t)$, *but we need always to remember that it is a random variable for a fixed t.*

Sometimes the function of $x(\omega_0, t)$ is called the realization of random process. That is really an ordinary function in time domain. The signal we observe in the natural world is the typical realization of some random process, but it is impossible to exhaust all realizations for a random process. The main task for us is to study the behavior and performance of a random process from its limited realization.

Definition 3.12 (Mean Value Function and Variance Function) *If the distribution density function of* $x(t)$ *is* $p(x, t)$, *then*

$$\begin{aligned} \mu(t) &= \int xp(x,t)\mathrm{d}x \\ \sigma^2(t) &= \int (x-\mu(t))^2 p(x,t)\mathrm{d}x \end{aligned} \tag{3.37}$$

are called the mean value function and variance function, respectively.

Definition 3.13 (Correlation Function)

$$R(t_1, t_2) = E(x(t_1)x(t_2)) \tag{3.38}$$

is called the correlation function of $x(t)$, *and*

$$C(t_1, t_2) = E[(x(t_1)-\mu(t_1))((x(t_2)-\mu(t_2))] = R(t_1, t_2) - \mu(t_1)\mu(t_2) \tag{3.39}$$

is called the auto-covariance function of $x(t)$.

Definition 3.14 (Stationary Random Process) *Suppose* $x(t)$ *is a random process and if for any time delay* τ, $x(t)$ *and* $x(t+\tau)$ *have the same statistical performance, then the random process* $x(t)$ *is called stationary. In particular, if* $x(t)$ *has a probability density function, then*

$$p(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = p(x_1, x_2, \ldots, x_n; t_1+\tau, t_2+\tau, \ldots, t_n+\tau) \tag{3.40}$$

Stationary random process is one of the most important types of random process in signal processing theory. For a stationary random process the mean

value function does not depend on time and the correlation function depends only on the time difference.

Example

Assume a random signal $x(t) = A\cos(2\pi ft + \varphi)$, where A is a constant and φ is a random phase with uniform distribution in $(0,\ 2\pi)$. It is proved that $x(t)$ is a stationary random process and

$$\begin{aligned} &E[x(t)] = 0 \\ &R_x(\tau) = E[x(t)x(t+\tau)] = A^2\cos(2\pi f\tau)/2 \end{aligned} \tag{3.41}$$

There is one kind of stationary process for which the statistical performance based on the sample space can be derived in time space based on realization. This kind of random process is called an ergodic stationary random process.

Theorem 3.8 (Ergodic Theorem of Mean Value) *Suppose* $x(t)$ *is a stationary random process and* $E[x(t)] = \mu$, $E[x(t)x(t-\tau)] = R(\tau)$, *then the ensemble average is equal to the time average of* $x(t)$, *i.e.*

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)\mathrm{d}t = \mu$$

if and only if

$$\lim_{T\to\infty}\frac{1}{2T}\int_{0}^{2T}\left(1-\frac{\tau}{2T}\right)(R(\tau)-\mu^2)\mathrm{d}\tau = 0$$

Theorem 3.9 (Ergodic Theorem of Correlation Function) *Suppose* $x(t)$ *is a stationary random process and* $E[x(t)x(t-\tau)] = R(\tau)$, *set* $y(t,\ \lambda) = x(t+\lambda)x(t)$, $R_y(\tau) = E[y(t+\tau,\ \lambda)y(t,\ \lambda)$. *For the given* λ, *if*

$$\lim_{T\to\infty}\frac{1}{2T}\int_{0}^{2T}\left(1-\frac{\tau}{2T}\right)(R_y(\tau)-R^2(\lambda))\mathrm{d}\tau = 0$$

then

$$R(\lambda) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t+\lambda)x(t)\mathrm{d}t$$

Theorem 3.10 (Ergodic Theorem of Probability Distribution) *Suppose* $x(t)$ *is a stationary random process, for any given real value* x, *let*

$$y(x,\ t)=\begin{cases}1 & x(t)\le x\\ 0 & x(t)>x\end{cases}$$

Denote the probability distribution function of $(x(t+\tau),\ x(t))$ *as* $G(x_1,\ x_2;\ \tau)$, *then the necessary and sufficient condition of the time average of* $y(x,\ t)$ *converges to the distribution function of* $x(t)$, *i.e.*

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}y(x,\ t)\mathrm{d}t=F(x)$$

is

$$\lim_{T\to\infty}\frac{1}{2T}\int_{0}^{2T}(1-\frac{\tau}{2T})(G(x_1,\ x_2;\ \tau)-F^2(x))\mathrm{d}\tau=0$$

It is very difficult to prove that a stationary random process is ergodic, since we need to prove that any kind of time average is equal to the ensemble average. It is also difficult in practice to verify if the ergodic condition is satisfied. Any record of random signal we observe in an application situation is just some of many realizations of this random signal. In discussing detection and estimation problems in signal processing, with a few exceptions, it always inherently assumed that the signal we are discussing is ergodic stationary random process, or at least, in some sense it is locally ergodic.

3.2 Introduction to the Basic Concepts of Information Theory

Information theory and cybernetic theory were among the most important achievements of the 1940s[1-2]. They have supported the development of science and technology up to the present day. The information theory proposed by Shannon was the first to give a measure of information. His coding theory and transmission theory are the foundation of modern communication theory. Shannon put communication theory on a serious mathematical basis. The cybernetic theory proposed by Wiener was an innovation of technological science, the achievement of many scientists and technicians working in different fields. The signal feedback theory was introduced in Wiener's preliminary works and yielded the new concept of cybernetics. There is no doubt that the combination of modern information theory and cybernetic theory was one of the most important achievements of the 20th century and will continue to influence trends in the 21st century.

In this section the basic concepts of information theory would be introduced. These concepts and results from information theory are essential for sonar designers, especially for researchers in the field of underwater acoustic communication.

3.2.1 Information and Entropy

Intuitively, when we receive a message we are told that an event, whose occurrence is uncertain, has happened. That is to say a message will bring us some kind of information. It is reasonable to think that if an event with very small probability has occurred then it will bring us a large amount of information. For those events that will definitely happen, when they happen they will bring only very limited information. Therefore, the quantitative measure of information should be related to the probability of an event.

If the possible outcome of an experiment is $x_1, x_2, ..., x_n$, the probability of each outcome is $p_1, p_2, ..., p_n$. The result of each experiment will bring us some information. From another point of view, the experiment has some uncertainty. If we find a measure to describe the quantity of information, then we also find a way to describe the uncertainty of an experiment.

In communication theory, the possible outcomes $x_1, x_2, ..., x_n$, of an experiment are called discrete information source. Accordingly, there is a source called continuous information source. This supposes that there are infinite possible outcomes or infinite symbols to be transmitted. As an example, let us first introduce how to calculate the information contained in a finite source.

Suppose there is a finite source X expressed as

$$\begin{matrix} x_1, & x_2, & ..., & x_n \\ p_1, & p_2, & ..., & p_n \end{matrix}$$

The information contained by the event x_i is expressed as $I(x_i) = f(x_i)$, which, as we noted before, is a function of p_i. It must satisfy the following three axioms:

(1) Monotonic decreasing: if $p_1 > p_2$, then $I(x_2) < I(x_1)$;

(2) If p_i=0, then $I(x_i) \to \infty$; if p_i=1, then $I(x_i) = 0$.

(3) The information of summation of two independent events should equal the summation of the information of the two events.

As we can see, the three axioms about information are intuitive. For example, consider the second axiom: if an impossible event does happen, it will bring us infinite information whereas, in contrast, no information is gained if an event with 100% probability occurs.

In mathematics, it can be shown that a function that satisfies the three axioms should have the form

$$I(x_i) = -\log_2 p_i \tag{3.42}$$

except a constant coefficient.

Traditionally, the function in Eq. (3.42) takes the logarithmic base 2 form, because for a binary source with an equal probability distribution $p_1 = p_2 = 0.5$, it follows that $I(x_1) = I(x_2) = 1$.

Definition 3.15 (The Entropy of Information Source) *The average information for the source X*

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{3.43}$$

is called the entropy of this source. Sometimes, H is called the self-information of X

$$I(X, X) = H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{3.44}$$

Theorem 3.11 *The maximum value of the entropy of an n symbol source is* $\log_2 n$, *i.e.*

$$H(X) \le \log_2 n \tag{3.45}$$

Definition 3.16 (Mutual Contain Information) *Suppose X is a random variable with n possible results* $x_1, x_2, \ldots, x_n$, *and Y is a random variable with m possible results* $y_1, y_2, \ldots, y_m$, *then*

$$I(x_i, y_j) = \log_2 \left[\frac{P(x_i \mid y_j)}{P(x_i)} \right] \tag{3.46}$$

is called the conditional information of x_i *in the case that* y_j *happens, where*

$$P(x_i) = P\{X = x_i\} \quad \text{and} \quad P(x_i \mid y_j) = P\{X = x_i \mid Y = y_j\}$$

Because

$$\frac{P(x_i \mid y_j)}{P(x_i)} = \frac{P(y_j \mid x_i)}{P(y_j)}$$

therefore

$$I(x_i, y_j) = I(y_j, x_i) \tag{3.47}$$

This means that the mutual information content is symmetrical.

Definition 3.17 (Average Mutual Information Content) *The mutual*

information content of source X and Y is defined as

$$\begin{aligned} I(X,\ Y) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,\ y_j) I(x_i,\ y_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,\ y_j) \log_2 \frac{P(x_i,\ y_j)}{P(x_i)P(y_j)} \end{aligned} \tag{3.48}$$

It is proved that $I(X,\ Y) \geq 0$.

Theorem 3.12 (Independent Information Source) *If X and Y are independent, then* $I(x,\ y) = 0$.

Definition 3.18 (Conditional Entropy, Multiple Entropy)

$$H(X \mid Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,\ y_j) \log_2 \frac{1}{P(x_i \mid y_j)} \tag{3.49}$$

is called the conditional entropy of X in Y.

$$H(X,\ Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,\ y_j) \log_2 P(x_i,\ y_j) \tag{3.50}$$

is called the two-dimensional entropy of X and Y.

Theorem 3.13 (The Relationship between Entropy and Information)

$$I(X,\ Y) = H(X) - H(X \mid Y) = H(X) + H(Y) - H(X,\ Y) \tag{3.51}$$

$$H(X) \geq H(X \mid Y) \tag{3.52}$$

$$H(X,\ Y) \leq H(X) + H(Y) \tag{3.53}$$

$$H(X,\ Y) = H(Y) + H(X \mid Y) \tag{3.54}$$

Most of the above formulas can be derived directly from the definition of entropy and information. Note that, for a discrete information source (or random variable), the entropy is always a non-negative quantity. The self-information or mutual information content is also a non-negative quantity.

The meaning of Eqs. (3.51) and (3.52) is quite understandable. For example, Eq. (3.52) says that in the case of one event having already happened, the entropy of another event will be equal to or less than the entropy when nothing happened. Eq. (3.53) says that the entropy of the combination of two events will be less than the summation of the entropy of two events. There is a very important inequality in mathematics, which is necessary in the proof of Eq. (3.53).

Theorem 3.14 *Assume* p_i, q_i $(i = 1, 2, ..., n)$ *are two sets of non-negative values*

$$\sum p_i = 1, \ \sum q_i = 1 .$$

Then

$$-\sum_{i=1}^{n} q_i \log_2 q_i \le -\sum_{i=1}^{n} q_i \log_2 p_i \tag{3.55}$$

The necessary and sufficient condition for the equality is $p_i = q_i$, for any i.

Now we are going to discuss the entropy and information of a continuous source or random variable. Usually, the continuous situation can be derived from the discrete situation by using a similarity method, but sometimes it is invalid. For example, in the self-information given in Eq. (3.44), the quantity p_i should be replaced by $p(x)\mathrm{d}x$, but $\log_2 p(x)\mathrm{d}x$ becomes infinite. The problem with this kind of similarity is that the contents of a continuous source have infinite information, therefore it is necessary to have infinite bits to describe it. To overcome this problem, a quantity called differential entropy has been introduced, i.e.

$$H(X) = -\int p(x) \log_2 p(x)\mathrm{d}x \tag{3.56}$$

where $p(x)$ is the probability density function of X.

Note that this has no meaning like self-information in a discrete source.

The mutual information content of two random variables X and Y is defined as

$$I(X, Y) = \iint p(x)p(x \mid y) \log_2 \frac{p(y \mid x)p(x)}{p(x)p(y)} \mathrm{d}x\mathrm{d}y \tag{3.57}$$

Similarly, the conditional entropy is defined as

$$H(X \mid Y) = -\iint p(x \mid y) \log_2 p(x \mid y)\mathrm{d}x\mathrm{d}y \tag{3.58}$$

It is proved that

$$I(X,Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \tag{3.59}$$

Sometimes we are interested in the case where one source is discrete and another is continuous, i.e., where X consists of x_i $(i = 1, 2, ..., n)$ with probability distribution $P(x_i)$ and the probability density function of Y is $p(y)$. When X and Y are statistically correlated, then we define

$$p(y) = \sum_{i=1}^{n} p(y \mid x_i) P(x_i), X \text{ and } Y$$

The mutual information contained

$$I(x_i, y) = \log_2 \frac{p(y \mid x_i) P(x_i)}{p(y) P(x_i)} = \log_2 \frac{p(y \mid x_i)}{p(y)}$$

The average information between X and Y is defined as

$$I(X, Y) = \sum_{i=1}^{n} \int p(y \mid x_i) P(x_i) \log_2 \frac{p(y \mid x_i)}{p(y)} \mathrm{d}y \tag{3.60}$$

Examples

(1) Suppose a system transmits a 0 – 1 series, and the probability for transmitting 0 is q, and for transmitting 1 is $1-q$. Therefore, the entropy of this source is

$$H(X) = -q \log_2 q - (1-q) \log_2 (1-q) \tag{3.61}$$

Fig. 3.3 plots the figure of $H(X)$ as a function of q, showing that, the maximum value of entropy $H(X)$ is reached in $q = 1/2$.

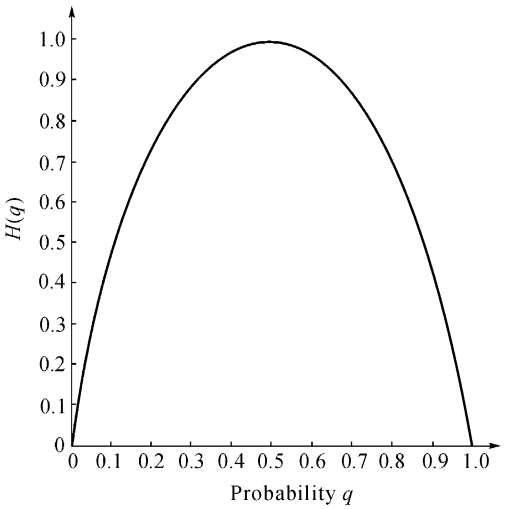


Fig. 3.3. Entropy of a binary information source

The binary transmitter / receiver system is the basic communication scheme. Suppose X and Y consist of 0, 1 symbols, the conditional probabilities are as

follows (Fig. 3.4): $P(X=0)=q$; $P(X=1)=1-q$; $P(Y=0\mid X=0)=1-p$; $P(Y=1\mid X=0)=p$; $P(Y=1\mid X=1)=1-p$; $P(Y=0\mid X=1)=p$.

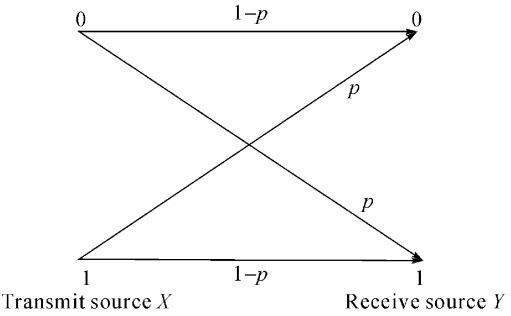


Fig. 3.4. Transmission model of binary

Based on Eqs. (3.49) and (3.50) we can calculate the conditional entropy and mutual information content (Fig. 3.5).

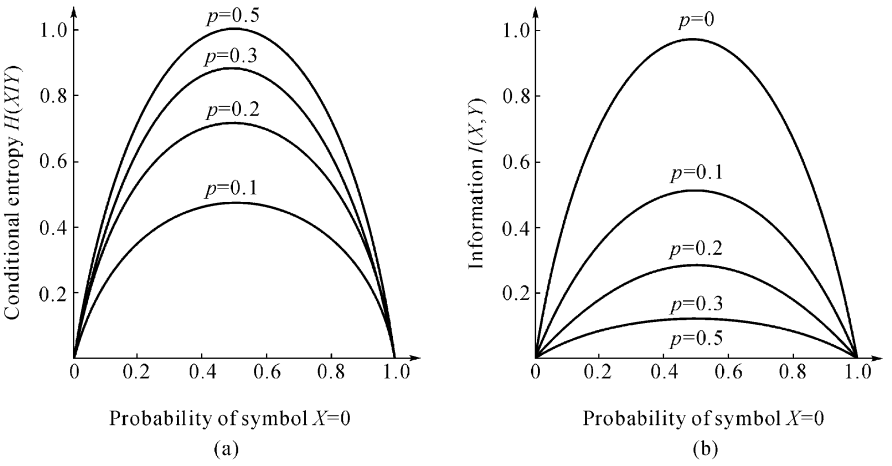


Fig. 3.5. Entropy and information of binary source. (a) Conditional entropy $H(X|Y)$; (b) Information $I(X,Y)$

(2) Suppose X is a discrete random variable, taking the values of $x_1 = A$, $x_2 = -A$ with equal probability. The conditional probability density functions of Y given A and $-A$ are

$$p(y \mid A) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-(y-A)^2 / (2\sigma^2)]$$

$$p(y \mid -A) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-(y + A)^2 / (2\sigma^2)]$$

From Eq. (3.60), it follows that

$$I(X, Y) = \frac{1}{2} \int \left[p(y \mid A) \log_2 \frac{p(y \mid A)}{p(y)} + p(y \mid -A) \log_2 \frac{p(y \mid -A)}{p(y)} \right] \mathrm{d}y$$

where

$$p(y) = [p(y \mid A) + p(y \mid -A)] / 2$$

3.2.2 The Coding Theorem of a Discrete Information Source

Suppose a discrete source has L elements x_i $(i = 1, ..., L)$. The entropy of this source is

$$H(X) = -\sum_{i=1}^{L} p_i \log_2 p_i \le \log_2 L$$

where p_i is the probability of transmitting x_i.

When discussing coding problems, transmission problems are often considered simultaneously. This is because the main aim of coding is more effective transmission and reception. The number of codes is obviously an important factor for good transmission.

Suppose a 0 – 1 series is used for coding the L elements (symbols) in source X, where L is in the form of the power of 2, i.e., $L = 2^R$. Therefore, to express L as a different symbol, it is necessary to have a 0 – 1 series with length R. Since $R = \log_2 L$, $H(X) \le R$, the effectiveness of coding can be expressed by $R / H(X)$. In the case where the probabilities of each symbol occurrence are equal, $R = H(X)$. This means that each symbol of length R is 100% effective. In fact, to reduce the transmission rate, it is necessary to determine how to reduce the length R. One method is to relax the rule of the one-to-one correspondence relation in coding. That means it is necessary for only those symbols which have high probability to have a unique code. The other symbols with low probability correspond to the same code. Thus, the necessary number of codes for transmitting L symbols in X will be reduced.

Divide L symbols of X into J groups and make a one-to-one correspondence for each group with one of $2^N - 1$ codes in length N. Each code needs N/J bits. Of

coures, this will result in some mistakes in decoding. Shannon's coding theorem answers the relationship between N and J.

Theorem 3.15 (Coding Theorem) *Suppose X is a finite information source with entropy $H(X)$. Divide X into J groups and make binary coding. The length of code is N. For any given $\varepsilon > 0$, if*

$$R = \frac{N}{J} \geq H(X) + \varepsilon$$

then we can find a coding method such that the decoding error p_e is arbitrary and small. In contrast, if $R \leq H(X) - \varepsilon$, then for a sufficiently large J, error p_e can be close to 1 and arbitrary.

Shannon's coding theorem points out the coding direction of finite information source in fixed length code. There is also a coding theorem for variable length codes, but as it is much more complicated, we will not describe it here.

For a continuous information source, there is a similar coding theorem, which is related to differential entropy.

Theorem 3.16 (Maximum Entropy of a Continuous Source) *Assume X is a continuous information source with probability density function $p(x)$, $\int p(x)\mathrm{d}x = 1$.*

(1) If the mean square value of X is limited, i.e., $\int x^2 p(x)\mathrm{d}x = \sigma^2$, then the maximum value of differential entropy

$$H(X) = -\int p(x)\log_2 p(x)\mathrm{d}x$$

is

$$\max\{H(X)\} = \log_2 \sigma\sqrt{2\pi e} \quad \text{(b)} \tag{3.62}$$

(2) If the peak value of X is limited, i.e.

$$p(x) = \begin{cases} 1/(2A) & -A \leq x \leq A \\ 0 & \text{otherwise} \end{cases}$$

then the maximum value of differential entropy is

$$\max\{H(X)\} = \log_2(2A) \quad \text{(b)} \tag{3.63}$$

Suppose X is band limited and the upper cut off frequency is W. After transmitting signal x, the received signal is $y = x + n$, where n is independent background noise which obeys Gaussian distribution $p_0(n)$, i.e., $p(y \mid x) = p_0(n)$. Based on the definition of conditional entropy (Eq. (3.58)), we have

$$\begin{aligned} H(Y \mid X) &= -\int p(x)\mathrm{d}x \int p(y \mid x)\log_2 p(y \mid x)\mathrm{d}y \\ &= -\int p(x)\mathrm{d}x \int p_0(n)\log_2 p_0(n)\mathrm{d}n \\ &= H(N) \end{aligned}$$

$$I(X,Y) = H(Y) - H(Y \mid X) = H(Y) - H(N)$$

The capacity per second of the transmission channel is $C = \max[H(Y) - H(Y \mid X)]2W$, where $2W$ is the sample number per second. Suppose the signal power is σ_s^2, and the noise power is σ_n^2, then

$$\max[H(Y)] = \log_2 \sqrt{2\pi e(\sigma_s^2 + \sigma_n^2)}$$

Because

$$H(Y \mid X) = H(N) = \log_2 \sqrt{2\pi e\sigma_n^2}$$

is a fixed value

$$\begin{aligned} C &= \max[H(Y) - H(Y \mid X)]2W \\ &= W\log_2\left(1 + \frac{\sigma_s^2}{\sigma_n^2}\right) \text{ (bps)} \end{aligned} \tag{3.64}$$

This is Shannon's famous capacity theorem for information transmission. As we can see, the channel capacity depends on the signal bandwidth and the SNR. To increase the ability of channel transmission, it is necessary to increase the signal bandwidth or the SNR. Fig. 3.6 plots the normalized value of C/W against the SNR.

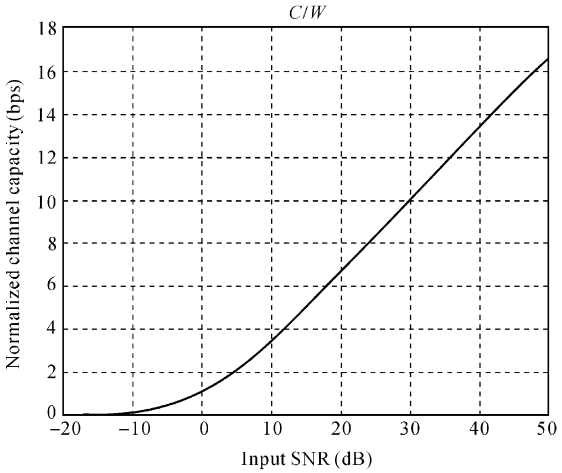


Fig. 3.6. Normalized channel capacity

3.3 The Optimum Receiving Theory of Weak Signal in Background Noise

One of the most important problems in many application areas is to extract the signal of interest from background noise. Background noise is random, the occurrence of signal and the behavior of signal are also random. Therefore, it is reasonable to deal with the signal extraction problem using methods based on probability theory and statistical estimation. That is to say that signal detection and parameter estimation problems are statistical hypothesis testing problems in mathematical statistics.

3.3.1 Basic Concepts of Statistical Hypothesis Tests

An "hypothesis" is a possible explanation for the outcome of a statistical problem. For example, in simple sonar signal detection there are two kinds of hypothesis: signal present and signal absent (noise only). For each hypothesis there is a probability expression of possible results. Based on this probability expression and the observation results, one can make a decision. We will start our discussion from the simplest situation.

A binary test is the simplest form of hypothesis test. There are only two possible hypotheses:

(1) H_0: noise only or signal absent, $x = n$.

(2) H_1: signal plus noise or signal present, $x = s + n$.

The detection problem is based on an observation about some random variables. The task for statistical testing is to make a decision, based on the observation result x, as to which hypothesis is correct. All possible results of x comprise an observation space. The binary test actually divides the observation space into two regions of X_0 and X_1. When the observation result falls into X_0, then we believe H_0 is correct, otherwise H_1 is correct. X_0 and X_1 are called the acceptance region and rejection region, respectively. Since the value of x is a random variable, our decision also has a random property. The problem is to choose the decision region as well as possible, such that the error probability is as small as possible.

First of all, if the result x already happens, it is natural to ask which is larger: $P(H_0 \mid x)$ or $P(H_1 \mid x)$? If $P(H_0 \mid x) > P(H_1 \mid x)$, choose H_0; If $P(H_1 \mid x) > P(H_0 \mid x)$, choose H_1. Denote this kind of decision by

$$\frac{P(H_1 \mid x)}{P(H_0 \mid x)} \begin{cases} > 1 & H_1 \text{ true} \\ < 1 & H_0 \text{ true} \end{cases} \tag{3.65}$$

This decision rule is called the maximum à posterior probability (MAP) test criteria.

By using the Bayes formula (Eq. (3.9)) in probability theory, Eq. (3.65) can be rewritten in a more useful format.

Since

$$P(H_i \mid x) = \frac{P(x \mid H_i)P(H_i)}{P(x)} \quad (i = 0,\ 1)$$

Therefore

$$\frac{P(H_1 \mid x)}{P(H_0 \mid x)} = \frac{P(x \mid H_1)}{P(x \mid H_0)} \frac{P(H_1)}{P(H_0)}$$

The decision rule in Eq. (3.65) becomes

$$\frac{P(x \mid H_1)}{P(x \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)} \tag{3.66}$$

The ratio $\Lambda(x) = P(x \mid H_1) / P(x \mid H_0)$ in Eq. (3.66) is called the likelihood ratio, and has a very important meaning in statistical test theory.

From Eqs. (3.65) and (3.66), we can see that in using the MAP rule it is necessary to calculate $P(H_1 \mid x)$ and $P(H_0 \mid x)$, or to have prior information about $P(H_0)$ and $P(H_1)$. In general, this is impossible because when we face target detection problems, it is difficult to say whether the target should be present, and even more difficult to estimate the probability. Therefore, it is necessary to find another kind of decision rule. Before doing that, let us consider four kinds of conditional probability in binary tests:

(1) $P(D_0 \mid H_0)$: H_0 true (no signal), decision D_0;

(2) $P(D_1 \mid H_1)$: H_1 true (signal present), decision D_1;

(3) $P(D_1 \mid H_0)$: H_0 true (no signal), decision D_1;

(4) $P(D_0 \mid H_1)$: H_1 true (signal present), decision D_0;

The probability in (3) is called "false alarm"; the probability in (4) is called "missing detection". False and missing are two kinds of error detection. Denote $p_0(x) = p(x \mid H_0)$, $p_1(x) = p(x \mid H_1)$. Suppose the decision spaces of X_0 and X_1 are defined as

$$X_0 = \left\{ x_0 \;\middle|\; \frac{p_1(x)}{p_0(x)} < Z \right\},\quad X_1 = \left\{ x \;\middle|\; \frac{p_1(x)}{p_0(x)} \geq Z_0 \right\}$$

When $x \in X_0$, D_0 is correct; when $x \in X_1$, D_1 is correct.

The false alarm probability

$$\alpha = P(D_1 \mid H_0) = \int_{X_1} p_0(x)\mathrm{d}x \tag{3.67}$$

The missing detection probability

$$\beta = P(D_0 \mid H_1) = \int_{X_0} p_1(x)\mathrm{d}x \tag{3.68}$$

The probability $p_d = 1 - \beta$ is called the detection probability, i.e., the probability of the correct detection.

In application problems, the seriousness of different errors is not same. To express the cost of various types of error, it is necessary to have some kinds of measure for each error. One can try to adopt a strategy such that the average cost is minimized. If it is difficult to quantify the cost of errors for a particular application problem, one can choose another criterion.

3.3.2 Optimum Detection Criterion

Some optimum detection criteria are introduced here. Their requirements for prior information are different.

Criterion of Minimum Total Error Probability

This criterion aims to minimize the total error probability. The total error probability is

$$\begin{aligned} p_e &= P(H_1 D_0) + P(H_0 D_1) \\ &= P(H_1)P(D_0 \mid H_1) + P(H_0)P(D_1 \mid H_0) \end{aligned}$$

Denoting $P(H_1) = \pi_1$, we have

$$p_e = \pi_1 \beta + (1 - \pi_1)\alpha \tag{3.69}$$

The criterion that minimizes p_e is called the criterion of minimum total error probability. Clearly, from Eq. (3.69), it is necessary to have the information of $P(H_1)$ to find the minimum value of p_e.

Criterion of Maximum SNR

This criterion requires the SNR in the system output to be maximized. The SNR is defined as

$$d = \frac{\text{output when signal present} - \text{output when signal absent}}{\text{output fluctuation without signal}} \tag{3.70}$$

The calculation of SNR is of course dependent on the signal processing schema.

Minimum Mean Square Error (MSE)

If some of the prior information of the signal of interest is known, it is possible to find a method such that the mean square error between the system output and the desired signal is minimized. Suppose $d(t)$ is the desired signal and $x(t)$ is the output of the detection system

$$\varepsilon^2 = E[(x(t) - d(t))^2] \tag{3.71}$$

The criterion for minimizing ε^2 is called MMSE.

Neyman-Pearson Criterion

The Neyman-Pearson criterion (N-P) is essentially the basic criterion in mathematical statistical test theory. In the condition that the false alarm probability is small or equal to a positive value α, the aim is to find a method such that the detection probability p_d is maximized.

3.3.3 N-P Test

The following theorem explores the relationship between the N-P criterion and the likelihood ratio function.

Theorem 3.17 (N-P Theorem) *Given a likelihood test*

$$\varphi(x)=\begin{cases}1 & \Lambda(x)\geq K\\ 0 & \Lambda(x)<K\end{cases}$$

Define

$$E_0[\varphi(x)]=\int_X \varphi(x)p_0(x)\mathrm{d}x=\int_{X_1} p_0(x)\mathrm{d}x=\alpha$$

and

$$E_1[\varphi(x)]=\int_X \varphi(x)p_1(x)\mathrm{d}x=1-\beta$$

For any other test $\psi(x)$, *if*

$$E_0[\psi(x)]\leq\alpha$$

then

$$E_1[\psi(x)]\leq E_1[\varphi(x)]$$

Therefore φ is the test method which has maximum detection probability and a false alarm probability equal to or less than α.

Now, we are going to prove this theorem.

It is easy to verify that

$$1-\beta=E_1[\varphi(x)]=\int_X \varphi(x)p_1(x)\mathrm{d}x=\int_{X_1} p_1(x)\mathrm{d}x$$

$$1-\beta'=E_1[\psi(x)]=\int_X \psi(x)p_1(x)\mathrm{d}x$$

It is necessary to prove, if $E_0[\psi(x)]=\int_X \psi(x)p_0(x)\mathrm{d}x\leq\alpha$, then $1-\beta\geq 1-\beta'$.

Denote $X^+=\{x\mid\varphi(x)-\psi(x)>0\}$ and $X^-=\{x\mid\varphi(x)-\psi(x)<0\}$, i.e., X^+ consists of those elements for which $\varphi(x)=1$ and $\psi(x)<1$; X^- consists of those elements for which $\varphi(x)=0$ and $\psi(x)>0$. Therefore, $\Lambda(x)>K$ and $x\in X^+$, and $\Lambda(x)<K$ and $x\in X^-$. For the elements neither belonging to X^+ nor X^-, we have $\varphi(x)=\psi(x)$. It follows that

$$\int_X[\varphi(x)-\psi(x)][p_1(x)-Kp_0(x)]\mathrm{d}x=\int_{X^+\cup X^-}[\varphi(x)-\psi(x)][p_1(x)-Kp_0(x)]\mathrm{d}x$$

Because $p_1(x)-Kp_0(x)>0$, when $\varphi(x)-\psi(x)>0$, and $p_1(x)-Kp_0(x)<0$,

when $\varphi(x)-\psi(x)<0$, therefore the integrated function is non-negative, i.e.

$$\int_X [\varphi(x)-\psi(x)][p_1(x)-Kp_0(x)]\mathrm{d}x \geq 0$$

or

$$\int_X [\varphi(x)-\psi(x)]p_1(x)\mathrm{d}x \geq K\int_X [\varphi(x)-\psi(x)]p_0(x)\mathrm{d}x$$

From the assumption

$$\int_X \varphi(x)p_0(x)\mathrm{d}x=\alpha, \quad \int_X \varphi(x)p_0(x)\mathrm{d}x \leq \alpha$$

therefore, the left side of the above equation is non-negative, so we have

$$\int_X [\varphi(x)-\psi(x)]p_1(x) \geq 0$$

That proves the theorem.

From the previous discussion we know that the N-P criterion concludes the maximum likelihood ratio test. Fig. 3.7 shows the meaning of α and β. Because of the importance of the maximum likelihood ratio test, sometimes one needs to know the detailed behavior of α and $p_d = 1-\beta$ as a function of the SNR. This kind of figure is called the receiver operating curve (ROC), and the output SNR given in Eq. (3.70) is the parameter (Fig. 3.8).

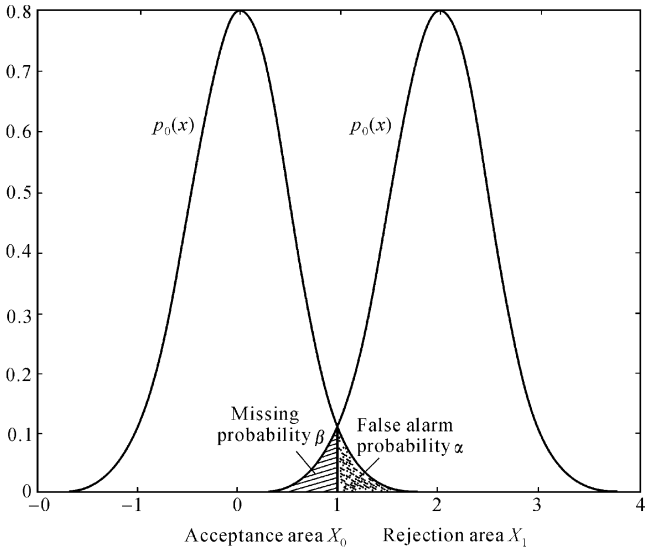


Fig. 3.7. False alarm and missing probability

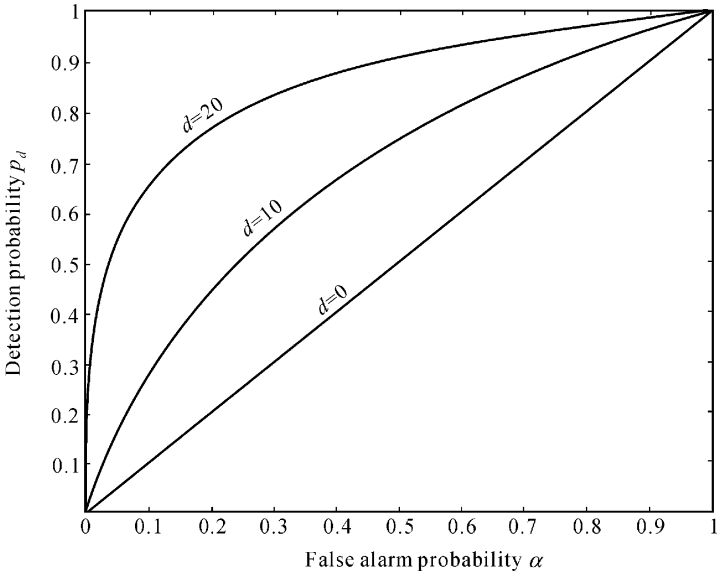


Fig. 3.8. Receiver operating curve

Example (*square detector or energy detector*)

Assume both signal $s(t)$ and noise $n(t)$ obey Gaussian distribution and $E[s(t)] = E[n(t)] = 0$, $E[s^2(t)] = \sigma_s^2$, $E[n^2(t)] = \sigma_n^2$.

(1) H_0: $x(t) = n(t)$, noise only;

(2) H_1: $x(t) = s(t) + n(t)$, signal plus noise.

It is easy to show that

$$p_1(x) = \frac{1}{\sqrt{2\pi(\sigma_s^2 + \sigma_n^2)}} \exp\left[-\frac{x^2}{2(\sigma_s^2 + \sigma_n^2)}\right]$$

$$p_0(x) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{x^2}{2\sigma_n^2}\right]$$

The rejection region is

$$X_1 = \left\{ x \mid \frac{p_1(x)}{p_0(x)} \geq Z_0 \right\} = \left\{ x \mid x^2 \geq \frac{\sigma_n^2(\sigma_n^2 + \sigma_s^2)}{\sigma_s^2} \ln \frac{Z_0\sqrt{\sigma_n^2 + \sigma_s^2}}{\sigma_n} \right\} \tag{3.72}$$

We can see from Eq. (3.72) that the optimum receiver is actually a square detector or energy detector. This example introduces a very important fact that the

optimum detector of Gaussian signal in Gaussian background noise is the square detector. Peterson was first to obtain this result in 1954[26].

3.3.4 Multiple Observations

The discussion in the previous section related to one observation in a binary test. It is not difficult to generalize the one observation results to multiple observations, i.e., based on the multiple observations of a random variable to make a decision between H_0 and H_1.

Suppose there are N data observations: $x_1, ..., x_N$, such as a series of observation results for the same physical parameter. The behavior of these observation data can be described by the conditional probability density function

$$p(x_1, x_2, ..., x_N \mid H_0) = p_0(x_1, x_2, ..., x_N)$$
$$p(x_1, x_2, ..., x_N \mid H_1) = p_1(x_1, x_2, ..., x_N)$$

To introduce the vector notation, let

$$\boldsymbol{x} = (x_1, x_2, ..., x_N)^{\mathrm{T}}$$

Consider $\boldsymbol{x}$ as one point in N-dimension space X. The process of making a decision is to divide space X into two parts: X_0 and X_1. When $\boldsymbol{x} \in X_0$, decide H_0 is true, when $\boldsymbol{x} \in X_1$, decide H_1 is true. Based on the N-P criterion, with a given false alarm probability α, the optimum detection method, which maximizes the detection probability, is as follows:

$$\Lambda(\boldsymbol{x}) \begin{cases} H_1 \\ > \\ < \\ H_0 \end{cases} \eta \tag{3.73}$$

where η is defined by the α

$$\int_{\eta}^{\infty} p_0(x)\mathrm{d}x = \alpha \tag{3.74}$$

The proof of the decision rule for the N-dimension case is similar to that for

one dimension: simply replace the scalar of one dimension by the vector of N dimensions, and understand that the integration of one dimension is an abstract measure of the integration of N-dimensions.

Example (*average detection*)

Suppose there are N independent observation values $x_1, x_2, ..., x_N$, and each x_i obeys the same Gaussian distribution. Under hypothesis H_0, with mean value 0 and variance σ^2, and under hypothesis H_1, with mean value m and variance σ^2; $P(H_0) = P(H_1) = 0.5$. We are going to find the rule of minimum total probability error.

Firstly, look at Eq. (3.69): $P(H_1) = \pi_1 = 0.5$, therefore

$$p_e = (\alpha + \beta)/2$$

Thus, the test that minimizes p_e is the likelihood ratio test, and

$$\Lambda(x) = \frac{p_1(x)}{p_0(x)} \begin{cases} H_1 \\ > \\ < \\ H_0 \end{cases} \frac{P(H_0)}{P(H_1)} = 1$$

In terms of the assumption

$$p(x_i \mid H_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - m)^2}{2\sigma^2} \right]$$

$$p(x_i \mid H_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{x_i^2}{2\sigma^2} \right]$$

Denoting $\boldsymbol{x} = (x_1, x_2, ..., x_N)^{\mathrm{T}}$, and $\boldsymbol{m} = (m_1, m_2, ..., m_N)^{\mathrm{T}}$, the combination probability function can be written as

$$p(\boldsymbol{x} \mid H_1) = \frac{1}{(2\pi\sigma^2)^{2/N}} \exp\left[-\frac{1}{2\sigma^2} (\boldsymbol{x} - \boldsymbol{m})^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{m}) \right]$$

$$p(\boldsymbol{x} \mid H_0) = \frac{1}{(2\pi\sigma^2)^{2/N}} \exp\left[-\frac{1}{2\sigma^2} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} \right]$$

Therefore, the likelihood ratio is

$$\Lambda(\boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid H_1)}{p(\boldsymbol{x} \mid H_0)} = \exp\left[-\frac{1}{\sigma^2}(\boldsymbol{m}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}/2) \right]$$

The rule for decision making is

$$\exp\left[-\frac{1}{\sigma^2}(\boldsymbol{m}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}/2) \right] \underset{H_0}{\overset{H_1}{\gtrless}} \frac{N}{2}m$$

Taking the logarithm of both sides, it follows that

$$\boldsymbol{m}^{\mathrm{T}}\boldsymbol{x} \underset{H_0}{\overset{H_1}{\gtrless}} \boldsymbol{m}^{\mathrm{T}}\boldsymbol{m}/2, \quad \text{or} \quad \sum_{i=1}^{N} x_i \underset{H_0}{\overset{H_1}{\gtrless}} \frac{N}{2}m \tag{3.75}$$

Thus the test statistic of the observation data $x_1, x_2, \ldots, x_N$ is the average value.

3.3.5 Wald Sequential Test

Clearly, in the test rule described above, one must choose between H_0 and H_1 regardless of the number of observations. But it is natural to expect that this kind of decision-making rule could be improved. That is, if there are insufficient observations to make a decision, one can continue to collect new data until we are able to make a choice between H_0 and H_1. This kind of test is called the Wald sequential test[10].

In the N-P criterion, the decision process compares the likelihood ratio function $\Lambda(x)$ with a threshold value K, which is defined by the false alarm probability. In the Wald test, two thresholds K_0, K_1 ($K_0 < K_1$) are defined, which are calculated from the false alarm probability α and the missing probability β.

If $\Lambda(x) > K_1$, choose H_1; if $\Lambda(x) < K_0$, choose H_0; and if $K_0 \le \Lambda(x) \le K_1$, it means that there are insufficient data to make a decision and one needs to collect more data. Wald proved that for given α, β, this kind of test should finish (i.e., the test procedure is convergent) and the number of observations needed, on

average, is less than that required for a fixed length test. The theoretical presentation of Wald theory is beyond the scope of this book. We list only the main conclusion of his theory.

Theorem 3.18 (Wald Sequential Test) *Suppose* $L(x) = \ln \Lambda(x)$. *In the case that* H_1 *is true, the average observation number is*

$$E(N \mid H_1) = \frac{(1-\beta)\ln K_1 + \beta \ln K_0}{E(L(x) \mid H_1)} \tag{3.76(a)}$$

In the case that H_0 *is true, the average observation number is*

$$E(N \mid H_0) = \frac{\alpha \ln K_1 + (1-\alpha)\ln K_0}{E(L(x) \mid H_0)} \tag{3.76(b)}$$

The probability of a test finish is 1, *i.e.*

$$\lim_{n\to\infty} P\{N \geq n\} = 0$$

For any given α and β, the average observation number calculated from Eq. (3.76) is minimal for any sequential tests.

3.4 Wiener Filtering, Matched Filtering and Adaptive Filtering for Stationary Random Signal

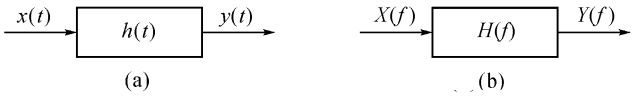
In signal detection theory, the stationary random signal is a popular and simple signal model. Suppose the input signal of a receiver is $x(t) = s(t) + n(t)$, where $s(t)$ and $n(t)$ are stationary random signals. $s(t)$ represents signal and $n(t)$ represents background noise. In most cases $s(t)$ and $n(t)$ can be considered as independent random signals. Of course, it depends on the practical situation and we will show the applicable condition, if necessary.

3.4.1 Basic Relation of Input / Output of a Linear System for Stationary Random Signal

Consider a linear system with unit impulse response $h(t)$ and Fourier transform $H(f)$ (Fig. 3.9). If the input and output signals of this system are $x(t)$ and $y(t)$, respectively, then according to the results of Chapter 2, we have the following

useful equation

$$\begin{aligned} y(t) &= \int x(\tau)h(t-\tau)\mathrm{d}\tau \\ X(f) &= \int x(t)\exp(-2\pi \mathrm{j}ft)\mathrm{d}t \\ Y(f) &= \int y(t)\exp(-2\pi \mathrm{j}ft)\mathrm{d}t \\ H(f) &= \int h(t)\exp(-2\pi \mathrm{j}ft)\mathrm{d}t \\ Y(f) &= H(f)X(f) \end{aligned} \tag{3.77}$$



(a) (b)

Fig. 3.9. Block diagram linear system in time and frequency domain. (a) Time domain; (b) Frequency domain

In general, $X(f)$, $Y(f)$ and $H(f)$ are complex functions. Because the detection problem in signal processing is usually related to the second moment of a random variable, it is necessary to list the main concepts and formulas of the correlation function and power spectrum, most of which can be derived directly from Eq. (2.65).

Assume $x(t)$ and $y(t)$ are two stationary random signals, and that the two-dimensional vectors $(x(t), y(t))$ are also stationary. Define the cross-correlation function and cross-power spectrum as follows

$$R_{xy}(\tau) = E[x(t)y(t-\tau)] \tag{3.78}$$

$$K_{xy}(f) = \int_{-\infty}^{\infty} R_{xy}(\tau)\exp(-2\pi \mathrm{j}f\tau)\mathrm{d}\tau \tag{3.79}$$

Clearly

$$R_{yx}(\tau) = R_{xy}(-\tau) \tag{3.80}$$

Therefore

$$K_{yx}(f) = K_{xy}^{*}(f) \tag{3.81}$$

If $y(t) = x(t)$, then

$$R_x(\tau) = R_{xx}(\tau) = E[x(t)x(t-\tau)]$$

Clearly

$$R_x(\tau) = R_x(-\tau) \tag{3.82}$$

i.e., the auto-correlation function is symmetry about 0.

Theorem 3.19 *Suppose the input of a linear system H(f) is x(t) and the output is y(t), then*

$$K_{yy}(f) = |H(f)|^2 K_{xx}(f) \tag{3.83}$$

$$K_{yx}(f) = H(f)K_{xx}(f) \tag{3.84}$$

As an example, we are going to prove Eq. (3.83).

$$\begin{aligned} K_{yy}(f) &= \int R_{yy}(\tau)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau \\ &= \int \mathrm{d}\tau \int \mathrm{d}u \int \mathrm{d}v h(u)h(v)R_{xx}(\tau+v-u)\exp(-2\pi \mathrm{j} f\tau) \\ &= \int h(u)\exp(-2\pi \mathrm{j} fu)\mathrm{d}u \int h(v)\exp(2\pi \mathrm{j} fv)\mathrm{d}v \int R_{xx}(\tau+v-u)\exp(-2\pi \mathrm{j} f(\tau+v-u))\mathrm{d}\tau \\ &= |H(f)|^2 K_{xx}(f) \end{aligned}$$

Theorem 3.20 *If the stationary random signal $x(t)$ passes through two linear systems $H_1(f)$, $H_2(f)$ (Fig. 3.10) with output signals $y_1(t)$ and $y_2(t)$, respectively, then the cross power spectrum of $y_1(t)$ and $y_2(t)$ is*

$$K_{y_1y_2}(f) = H_1(f)H_2^*(f)K_{xx}(f) \tag{3.85}$$

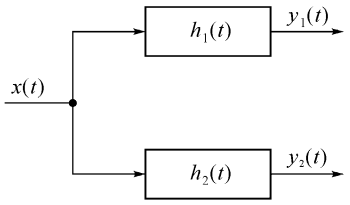


Fig. 3.10. Single input / multi-output linear system

The proof of theorem 3.20 is similar to that of theorem 3.19. In fact, Eq. (3.83) is a special case of Eq. (3.85).

The main topic of this section is the optimum detection theory of stationary random signal, which strictly speaking is the problem of detecting stationary random signal or deterministic signal in the background of stationary random

noise. The Wiener filtering theory deals with optimum stationary random signal detection in the background of stationary random noise. Therefore, this is a classical topic in signal processing, and Wiener's research work provides a perfect solution to this problem[44].

The matched filtering theory deals with the problem of detecting deterministic signal in the background of stationary random signal. Turin's paper from the early 1960s provides a good introduction[30].

The adaptive signal processing technique is an advanced optimum signal processing algorithm with a self-learning function. It adjusts its own parameters to match a slowly-changed environment such that the system behavior reaches an optimum state under certain criteria. The fundamentals of this technique are described by Widrow and Griffths[15-17, 19-20].

3.4.2 Wiener Filtering

Wiener filtering aims to optimize the detection of stationary signal in a background of stationary noise (Fig. 3.11). Suppose the input signal is $x(t) = s(t) + n(t)$, where $s(t)$ is signal and $n(t)$ is noise. Assume $s(t)$ and $n(t)$ are stationary random process and are independent. The impulse response function of the linear system is $h(t)$. The output of the system is $y(t)$. According to the theory of linear systems, we have

$$y(t) = \int h(t-\tau)x(\tau)\mathrm{d}\tau$$

It is necessary to find an optimum linear system $h(t)$, such that the mean square is minimum.

$$I = E[(y(t) - s(t))^2] \tag{3.86}$$

Intuitively, we need to find an optimum linear system such that the mean square error between the output and the signal component of the input is minimized.

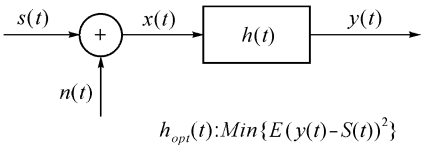


Fig. 3.11. Wiener filtering of stationary random signal

Denote the minimum value of Eq. (3.86) as $I_{\min}$ and the optimum value of $h(t)$ as $h_{\text{opt}}(t)$. A linear filter with impulse response $h_{\text{opt}}(t)$ is called a Wiener filter. Because of the importance of Wiener filters in the detection theory of

random signal, we will introduce two different ways to resolve the minimum value of Eq. (3.86).

Define the auto-correlation function of signal and noise as follows

$$R_{ss}(\tau) = E[s(t)s(t-\tau)]$$
$$R_{nn}(\tau) = E[n(t)n(t-\tau)]$$
$$R_{xx}(\tau) = E[x(t)x(t-\tau)] = R_{ss}(\tau) + R_{nn}(\tau)$$

Now the auto-power spectrum functions are

$$K_{ss}(f) = \int R_{ss}(\tau)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau$$
$$K_{nn}(f) = \int R_{nn}(\tau)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau$$
$$K_{xx}(f) = \int R_{xx}(\tau)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau$$

Suppose the transfer function of a linear system, i.e., the Fourier transform of $h(t)$, is $H(f)$, then

$$K_{yy}(f) = |H(f)|^2 K_{xx}(f)$$

Expanding Eq. (3.86), it follows that

$$\begin{aligned} I &= E[(y(t)-s(t))^2] \\ &= E[s^2(t)] + E[y^2(t)] - 2E[s(t)y(t)] \\ &= E[s^2(t)] + \int |H(f)|^2 K_{xx}(f)\mathrm{d}f - 2\int H(f)K_{xs}(f)\mathrm{d}f \end{aligned} \tag{3.87}$$

definition of $K_{xs}(f)$ can be derived from Eq. (3.79). Consider I as a function of $H(f)$, and take its differential. Based on the functional theory in mathematics, it is easy to find h_{opt} (t) or $H_{\mathrm{opt}}(f)$. We have

$$\frac{\partial I}{\partial H} = H(f)K_{xx}(f) - K_{xs}(f) = 0$$

This obtains

$$H_{\mathrm{opt}}(f) = K_{xx}^{-1}(f)K_{xs}(f) \tag{3.88}$$

or

$$H_{\text{opt}}(f) = K_{xx}^{-1}(f)K_{sx}^{*}(f) \tag{3.89}$$

Substituting Eq. (3.89) into Eq. (3.87) to obtain

$$I_{\min} = E[s^2(t)] - \int \frac{|K_{xs}(f)|^2}{K_{xx}(f)} \mathrm{d}f \tag{3.90}$$

$$h_{\text{opt}}(t) = \int H_{\text{opt}}(f) \exp(2\pi \mathrm{j} f t) \mathrm{d}f \tag{3.91}$$

Eqs. (3.88), (3.89) and (3.91) are the expressions of the resolution of the Wiener filter.

Now we turn to resolve the Wiener problem from a different point of view: this will yield the famous Wiener-Hopf Equation. As we know that

$$y(t) = \int h(t-\tau)x(\tau)\mathrm{d}\tau$$

$$\begin{aligned} I &= E[(y(t)-s(t))^2] \\ &= E[s^2(t)] - 2E[s(t)y(t)] + E[y^2(t)] \\ &= E[s^2(t)] - 2\int h(\tau)R_{sx}(\tau)\mathrm{d}\tau + \iint h(\tau_1)h(\tau_2)R_{xx}(\tau_2-\tau_1)\mathrm{d}\tau_1\mathrm{d}\tau_2 \end{aligned} \tag{3.92}$$

To find the minimum value $I_{\min}$, we are going to adopt so-called "micro-disturb" method in mathematics. Assume

$$h(t) = h_{\text{opt}}(t) + \alpha\delta h_{\text{opt}}(t) \tag{3.93}$$

where α is a scalar factor and δ represents a symbol of variation, i.e., $\delta h_{\text{opt}}(t)$ is a micro disturb of $h_{\text{opt}}(t)$. Substituting Eq. (3.93) into Eq. (3.92) obtains

$$\begin{aligned} I = {} & R_{ss}(0) - 2\int [h_{\text{opt}}(\tau) + \alpha\delta h_{\text{opt}}(\tau)]R_{sx}(\tau)\mathrm{d}\tau \\ & + \iint [h_{\text{opt}}(\tau_1) + \alpha\delta h_{\text{opt}}(\tau_1)][h_{\text{opt}}(\tau_2) + \alpha\delta h_{\text{opt}}(\tau_2)]R_{xx}(\tau_2-\tau_1)\mathrm{d}\tau_1\mathrm{d}\tau_2 \end{aligned} \tag{3.94}$$

This is the function of $\alpha[0,\ 1]$, denoted as $I(\alpha)$. The minimum value should be reached at $\alpha = 0$. Therefore, it follows that

$$\frac{\partial I(\alpha)}{\partial \alpha}\Big|_{\alpha=0} = 0$$

However, it is straightforward to calculate that

$$\frac{\partial I(\alpha)}{\partial \alpha} = -2\int \delta h_{\text{opt}}(\tau) R_{sx}(\tau)\mathrm{d}\tau + \iint h_{\text{opt}}(\tau_2)\delta h_{\text{opt}}(\tau_1) R_{xx}(\tau_2-\tau_1)\mathrm{d}\tau_2\mathrm{d}\tau_1$$
$$+ \iint h_{\text{opt}}(\tau_1)\delta h_{\text{opt}}(\tau_2) R_{xx}(\tau_2-\tau_1)\mathrm{d}\tau_1\mathrm{d}\tau_2$$
$$+ 2\iint \alpha\delta h_{\text{opt}}(\tau_1)\delta h_{\text{opt}}(\tau_2) R_{xx}(\tau_2-\tau_1)\mathrm{d}\tau_1\mathrm{d}\tau_2$$

When α equals zero, this equation should solve to 0, that is

$$2\int \delta h_{\text{opt}}(\tau_2)[-R_{xs}(\tau_2) + \int R_{xx}(\tau_2-\tau_1)h_{\text{opt}}(\tau_1)\mathrm{d}\tau_1]\mathrm{d}\tau_2 = 0 \tag{3.95}$$

Since Eq. (3.95) should be valid for every $\mathrm{d}h_{\text{opt}}(t)$, the integrated function in Eq. (3.95) must be equal to 0, i.e.

$$-R_{xs}(\tau) + \int R_{xx}(\tau-\tau_1)h_{\text{opt}}(\tau_1)\mathrm{d}\tau_1 = 0 \tag{3.96}$$

or

$$R_{xs}(\tau) = \int R_{xx}(\tau-\tau_1)h_{\text{opt}}(\tau_1)\mathrm{d}\tau_1 \tag{3.97}$$

This is the integral equation in which $h_{\text{opt}}(t)$ should be satisfied.

Taking the Fourier transform on both sides of Eq. (3.97), it follows that

$$K_{xs}(f) = K_{xx}(f)H_{\text{opt}}(f)$$

This is another expression of Eq. (3.88).

Note that in the resolution of the optimum transfer function, it is supposed that $s(t)$ and $n(t)$ are independent. Independence is not a necessary condition, but without it the proof is much more complicated.

Similar to eq. (3.90), $I_{\min}$ can be also expressed in time domain

$$I_{\min} = R_{ss}(0) - \int R_{xs}(\tau)h_{\text{opt}}(\tau)\mathrm{d}\tau \tag{3.98}$$

3.4.3 Matched Filter

Now consider the problem of the optimum detection of deterministic signal in the background of Gaussian random noise. Here the optimum criterion is not the minimum mean square error but the maximum SNR. Suppose the input signal of a

linear system with impulse response $h(t)$ is $s(t)$, and the background noise is $n(t)$. At first, for simplification, assume $n(t)$ is white noise, i.e., the power spectrum is constant at $N_0/2$ over the frequency band [– W, W]

$$K_n(f)=\begin{cases} N_0/2 & |f|\le W \\ 0 & \text{otherwise} \end{cases} \tag{3.99}$$

Suppose the amplitude spectrum of the signal is $S(f)$, the requirement of a matched filter is that the SNR in the output of the linear system has a maximum value in a certain observation time (Fig. 3.12).

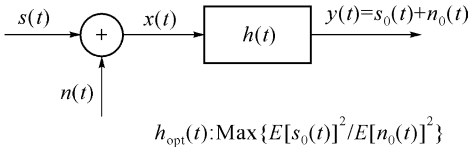


Fig. 3.12. Matched filtering of stationary signal in noise background

Denote the output signal of a linear system as $y_0(t)$, based on the characteristics of linear systems, it follows that

$$y_0(t)=x(t)*h(t)=\int h(\tau)[s(t-\tau)+n(t-\tau)]\mathrm{d}\tau$$

The signal component of the output in the moment t_0 is

$$d_s=\int h(\tau)s(t_0-\tau)\mathrm{d}\tau=\int H(f)S(f)\exp(2\pi\mathrm{j}ft_0)\mathrm{d}f \tag{3.100}$$

The power of the noise component of the output is

$$\sigma_n^2=\int|H(f)|^2K_n(f)\mathrm{d}f=\frac{N_0}{2}\int|H(f)|^2\mathrm{d}f$$

Therefore, the SNR in the output is

$$(SNR)_{\text{out}}=\frac{d_s^2}{\sigma_n^2}=\frac{\left[\int S(f)H(f)\exp(2\pi\mathrm{j}ft_0)\mathrm{d}f\right]^2}{N_0\int|H(f)|^2\,\mathrm{d}f/2}$$

In terms of Schwartz inequality Eq. (3.29), i.e., for real functions $f(x)$ and $g(x)$

$$|\int f(x)g(x)\mathrm{d}x|\le[\int f^2(x)\mathrm{d}x]^{1/2}[\int g^2(x)\mathrm{d}x]^{1/2}$$

Therefore,

$$|\int S(f)H(f)\exp(2\pi \mathrm{j}ft_0)\mathrm{d}f|^2 \le \int|H(f)|^2\,\mathrm{d}f\int|S(f)\exp(2\pi \mathrm{j}ft_0)|^2\,\mathrm{d}f$$

That means that

$$(SNR)_{\text{out}} \le \frac{\int|S(f)|^2\mathrm{d}f}{N_0/2} = \frac{2\sigma_s^2}{N_0} \tag{3.101}$$

where $d_s^2 = \int|S(f)|^2\,\mathrm{d}f$ is the energy of the signal component in the input. In Eq. (3.101), the equality is valid if and only if

$$H(f) = cS^*(f)\exp(-2\pi \mathrm{j}ft_0) \tag{3.102}$$

where, c is a constant. Afilter with the transfer function in Eq. (3.102) is called a matched filter for signal $s(t)$.

Denote

$$H_{\text{opt}}(f) = cS^*(f)\exp(-2\pi \mathrm{j}ft_0)$$

and

$$h_{\text{opt}}(t) = F^{-1}[H_{\text{opt}}(f)] = cs(t_0 - t) \tag{3.103}$$

Eq. (3.103) shows that the impulse response function of the matched filter is just a copy of the signal component in the input, but it reverses the time order. This is why the filter is called a "matched" filter.

It is easy to calculate the system gain from a matched filter. Suppose the signal duration is T, then the mean power value per unit time is σ_s^2/T, and the input power of noise is $\sigma_n^2 = N_0W$. Therefore, the input SNR is

$$(SNR)_{\text{in}} = \frac{\sigma_s^2/T}{N_0W}$$

From Eq. (3.101), it follows that

$$G_{\text{opt}} = \frac{(SNR)_{\text{out}}}{(SNR)_{\text{in}}} = \frac{2\sigma_s^2/N_0}{\sigma_s^2/(TN_0W)} = 2TW \tag{3.104}$$

This is to say that the system gain from the matched filter is twice the product of time and bandwidth. This is a very important conclusion in signal detection.

If the background noise is not a white Gaussian noise, the expression of the optimum transfer function of a matched filter is a little different. Suppose the power spectrum density function is $K_n(f)$. The signal component of the output in the moment t_0 is

$$d_s = \int h(\tau)s(t_0 - \tau)\mathrm{d}\tau = \int H(f)S(f)\exp(2\pi \mathrm{j}ft_0)\mathrm{d}f$$

The power of noise component of the output is

$$\sigma_n^2 = \int |H(f)|^2 K_n(f)\mathrm{d}f$$

The output SNR is

$$(SNR)_{\text{out}} = \frac{d_s^2}{\sigma_n^2} = \frac{|\int S(f)H(f)\exp(2\pi \mathrm{j}ft_0)\mathrm{d}f|^2}{\int |H(f)|^2 K_n(f)\mathrm{d}f}$$

In terms of Schwartz inequality, it follows that

$$(SNR)_{\text{out}} \le \int \frac{|S(f)\exp(2\pi \mathrm{j}ft_0)|^2}{K_n(f)}\mathrm{d}f \tag{3.105}$$

The equal sign is valid if, and only if,

$$H(f) = \frac{c}{K_n(f)}S^*(f)\exp(-2\pi \mathrm{j}ft_0) \tag{3.106}$$

Based on the convolution theorem, the system expressed in Eq. (3.106) can be considered as the cascade of two linear systems. The first system has the transfer function $H_1(f) = 1/\sqrt{K_n(f)}$, when the signal plus noise $x(t) = s(t) + n(t)$ passes through this filter. The noise component is whitened, giving it a flat spectrum over the frequency band. The second linear system is simply the matched filter for white background noise. This time, as the signal component already has the spectrum $S(f)/\sqrt{K_n(f)}$, the optimum transfer function of the matched filter should be

$$H_2(f) = \frac{c}{\sqrt{K_n(f)}}S^*(f)\exp(-2\pi \mathrm{j}ft_0) \tag{3.107}$$

Finally

$$H_{\text{opt}}(f) = H_1(f)H_2(f) = \frac{c}{K_n(f)} S^*(f)\exp(-2\pi \mathrm{j} f t_0) \tag{3.108}$$

The matched filter is used extensively in many areas, particularly in active signal detection.

3.4.4 Adaptive Filtering

Adaptive theory was proposed by Widrow et al. in the 1960s. This theory has been applied extensively in many areas of signal processing and has become an important branch of signal processing theory.

The adaptive filter is an optimum linear system. The parameters of this filter are not preset and fixed in the operation, but are adaptively adjusted in the operation procedure. The system can identify changes in the environment and learn from the input / output data, feedback the error function to the system, and finally complete the system adjustment and reach a steady state.

Fig. 3.13 is a block diagram of a single input, single output adaptive filter. Assume the system input is

$$x(t) = s(t) + n(t)$$

where, $s(t)$ is a signal and $n(t)$ is a noise, and $s(t)$ and $n(t)$ are independent random signals. The system output is $y(t)$. Suppose $d(t)$ is the desired signal. The main task of an adaptive filter is to find an optimum transfer $H_{\text{opt}}(f)$ function such that

$$I = E[\varepsilon^2(t)] = E[(y(t) - d(t))^2] \tag{3.109}$$

is minimized (Fig. 3.13).

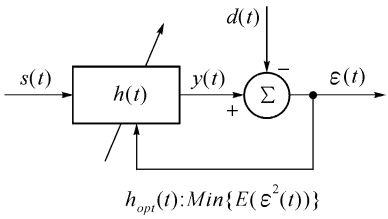


Fig. 3.13. Block diagram of adaptive filtering

Using the notation for the correlation function and the power spectrum density functions of $x(t)$, $y(t)$ and $d(t)$, and expanding Eq. (3.109), it follows that

$$\begin{aligned} I &= E[(y(t)-d(t))^2] \\ &= E[d^2(t)] - 2E[y(t)d(t)] + E[y^2(t)] \\ &= E[d^2(t)] - 2\int K_{yd}(t)\mathrm{d}f + \int K_{yy}(f)\mathrm{d}f \\ &= E[d^2(t)] - 2\int H(f)K_{xd}(f)\mathrm{d}f + \int |H(f)|^2 K_{xx}(f)\mathrm{d}f \end{aligned} \tag{3.110}$$

Similar to the Wiener solution, the optimum solution $H_{\text{opt}}(t)$ of Eq. (3.110) such that $I = I_{\min}$ is

$$H_{\text{opt}}(f) = \frac{K_{xd}^*(f)}{K_{xx}(f)} \tag{3.111}$$

and

$$I_{\min} = E[d^2(t)] - \int H_{\text{opt}}(f)K_{xd}(f)\mathrm{d}f \tag{3.112}$$

If the practical filter is different from $H_{\text{opt}}(t)$, then the mean square error I should have some bias to $I_{\min}$. Here is an important formula to estimate the bias:

$$\begin{aligned} I &= E[d^2(t)] - 2\int H(f)K_{xd}(f)\mathrm{d}f + \int |H(f)|^2 K_{xx}(f)\mathrm{d}f \\ &= E[d^2(t)] - 2\int H_{\text{opt}}(f)K_{xd}(f)\mathrm{d}f + \int |H_{\text{opt}}(f) - H(f)|^2 K_{xx}(f)\mathrm{d}f \\ &= I_{\min} + \int |H_{\text{opt}}(f) - H(f)|^2 K_{xx}(f)\mathrm{d}f \end{aligned} \tag{3.113}$$

Widrow describes an algorithm to find $H_{\text{opt}}(t)$ based on the input data series. A transversal filter (tapped delay line) is used to approximate $H_{\text{opt}}(t)$. Fig. 3.14 shows a block diagram of a tapped delay line. The impulse response function is

$$h(t) = \sum_{k=0}^{N-1} w_k \delta(t - k\Delta) \tag{3.114}$$

and the transfer function is

$$H(f) = \sum_{k=0}^{N-1} w_k \exp(-2\pi \mathrm{j} f k\Delta) \tag{3.115}$$

In Eq. (3.110), the optimum transfer function is to be found over all linear systems, but if we restrict our linear system family to a tapped delay line with $N-1$ taps, then the resolution should be different. To distinguish these two kinds of linear systems, denote

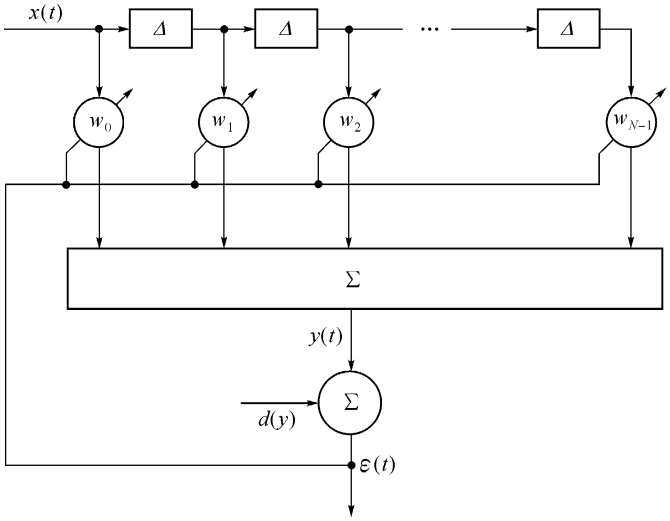


Fig. 3.14. Implementation of an adaptive filter using a tapped delay line

$$J = E[(y(t) - d(t))^2] \tag{3.116}$$

The filter is restricted to a tapped delay line with $N-1$ taps. If the solution of Eq. (3.116) is such that J reaches $J_{\min}$, then we have

$$J_{\min} \geq I_{\min}$$

Denote

$$\boldsymbol{w} = [w_1,\ w_2,\ ...,\ w_{N-1}]^{\mathrm{T}}$$
$$\boldsymbol{x} = [x_1,\ x_2,\ ...,\ x_{N-1}]^{\mathrm{T}}$$

where $x_i = x(t - i\Delta)$ $(i = 0,\ ...,\ N-1)$ (Fig. 3.14).
Now we have

$$\begin{aligned} J &= E[\varepsilon^2(t)] = E[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} - d(t)]^2 \\ &= E[d^2(t)] - 2\boldsymbol{R}_{xd}^{\mathrm{T}}\boldsymbol{w} + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_{xx}\boldsymbol{w} \end{aligned} \tag{3.117}$$

where $\boldsymbol{R}_{xx}$ represents the auto-correlation matrix of $\boldsymbol{x}$, and $\boldsymbol{R}_{xd}$ represents the cross-correlation vector of $\boldsymbol{x}$ and $d(t)$. Eq. (3.117) shows that J is a quadratic form of vector $\boldsymbol{w}$. To find the minimum value of J, it is necessary to find the gradient of J for $\boldsymbol{w}$ and force the gradient to equal zero.

There is a basic formula for a vector gradient

$$\nabla_{w}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{w}) = \boldsymbol{A}\boldsymbol{w} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$$

where $\boldsymbol{A}$ is a matrix and $\boldsymbol{w}$ is a vector. Based on this formula, we have

$$\nabla_{w}(J) = -2\boldsymbol{R}_{xd} + \boldsymbol{R}_{xx}\boldsymbol{w} + \boldsymbol{R}_{xx}^{\mathrm{T}}\boldsymbol{w} = -2\boldsymbol{R}_{xd} + 2\boldsymbol{R}_{xx}\boldsymbol{w}$$

Therefore

$$\boldsymbol{w}_{\mathrm{opt}} = \boldsymbol{R}_{xx}^{-1}\boldsymbol{R}_{xd} \tag{3.118}$$

Substituting Eq. (3.117) into Eq. (3.118), it follows that

$$J_{\min} = E[d^{2}(t)] - \boldsymbol{w}_{\mathrm{opt}}^{\mathrm{T}}\boldsymbol{R}_{xd} \tag{3.119}$$

Clearly, Eq. (3.118) is similar to Eq. (3.111), and Eq. (3.119) is similar to Eq. (3.112).

If we denote the optimum transfer function of $\boldsymbol{w}_{\mathrm{opt}}$ as $H_{0}(f)$, then Eq. (3.113) becomes

$$J_{\min} = I_{\min} + \int |H_{\mathrm{opt}}(f) - H_{0}(f)|^{2} K_{xx}(f)\mathrm{d}f \tag{3.120}$$

From Eq. (3.118) it is easy to see that to resolve $\boldsymbol{w}_{\mathrm{opt}}$, it is necessary to find the inverse matrix of the correlation matrix. The direct calculation is difficult and complicated. Many different algorithms have been devised that can calculate $\boldsymbol{w}_{\mathrm{opt}}$ in real time based on the input data series. One method, called the steepest descent algorithm, proposed by Widrow et al.[16], is introduced here. Firstly, choose an arbitrary preset value for $\boldsymbol{w}$, and then start from this preset value to find the optimum value along the path of the negative gradient of $\varepsilon(t)$ in each step by using an iterative algorithm.

Suppose the value of $\boldsymbol{w}$ in l-th step is $\boldsymbol{w}(l)$, let

$$\boldsymbol{w}(l+1) = \boldsymbol{w}(l) + \mu\nabla(l) \tag{3.121}$$

where $\nabla(l)$ represents the gradient of $E[\varepsilon^{2}(l)]$ and μ is a coefficient to manage the convergence of the iterative procedure. It is easy to show that

$$\nabla(l) = -2\boldsymbol{R}_{xd} + 2\boldsymbol{R}_{\mathbf{xx}}\boldsymbol{w}(l) \tag{3.122}$$

In Eq. (3.122) there are some terms such as $\boldsymbol{R}_{xd}$, $\boldsymbol{R}_{xx}$, which are the mean values of random processes; these will be replaced by transient quantities, i.e.

$$\nabla(l) = \nabla[\varepsilon^2(l)] = 2\varepsilon(l)\boldsymbol{x}(l)$$

therefore

$$\boldsymbol{w}(l+1) = \boldsymbol{w}(l) + 2\mu\varepsilon(l)\boldsymbol{x}(l) \tag{3.123}$$

This is the kernel formula of Widrow's algorithm in adaptive filtering. Note that $\varepsilon(l) = y(l) - d(l) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}$. The terms appearing in Eq. (3.123) are all real data which can be measured or observed in real time. Therefore, it is easy to realize in hardware. It is also proved that

$$\lim_{l\to\infty} E[\boldsymbol{w}(l)] = \boldsymbol{w}_{\mathrm{opt}} \tag{3.124}$$

provided μ is small enough. This is to say that the mean value of the weight coefficient vector $\boldsymbol{w}(l)$ will converge to the optimum value $\boldsymbol{w}_{\mathrm{opt}}$ when l becomes infinite, regardless of the preset value of $\boldsymbol{w}(l)$.

Since $\boldsymbol{w}(l)$ approximates to the optimum value, $\boldsymbol{w}_{\mathrm{opt}}$ is in a sense a mean value of convergence. Therefore, the procedure of $J = E[\varepsilon^2(l)]$ when close to $J_{\min}$ may show fluctuation, as a function of l. The curve of $J(l) = E[\varepsilon^2(l)]$ is called the learning curve of adaptive filtering and reflects the behavior of the adaptive process in iteration. It is not difficult to find the relationship between $J(l) = E[\varepsilon^2(l)]$ and $J_{\min}$

$$\begin{aligned} J(l) = E[\varepsilon^2(l)] &= E[d^2(l)] - 2\boldsymbol{R}_{xd}^{\mathrm{T}}\boldsymbol{w}(l) + \boldsymbol{w}^{\mathrm{T}}(l)\boldsymbol{R}_{xx}\boldsymbol{w}(l) \\ &= J_{\min} + [\boldsymbol{w}(l) - \boldsymbol{w}_{\mathrm{opt}}]^{\mathrm{T}}\boldsymbol{R}_{xx}[\boldsymbol{w}(l) - \boldsymbol{w}_{\mathrm{opt}}] \end{aligned} \tag{3.125}$$

This is an important formula in analyzing the behavior of an adaptive filtering system.

Adaptive filtering theory is used extensively in radar, sonar and speech signal processing. Fig. 3.15 illustrates four examples of applications.

Fig. 3.15 (a) is a block diagram of an adaptive noise canceling filter. Suppose $m(t)$ and $n(t)$ are correlated and independent of $s(t)$. The main task of this filter is to adjust $m(t)$ such that

$$J = E[\varepsilon^2(t)] = E[s(t) + n(t) - x(t)]^2$$

is minimized, where $x(t)$ is the output of the adaptive filter. According to the assumption, it obtains

$$J = E[s^2(t)] = E[n(t) - x(t)]^2$$

J will reach its minimum value when

$$(SNR)_{\text{out}} = \frac{E[s^2(t)]}{E[n(t)-x(t)]^2}$$

reaches its maximum value. In other words, the minimum mean square error criterion and maximum SNR criterion are identical in this situation.

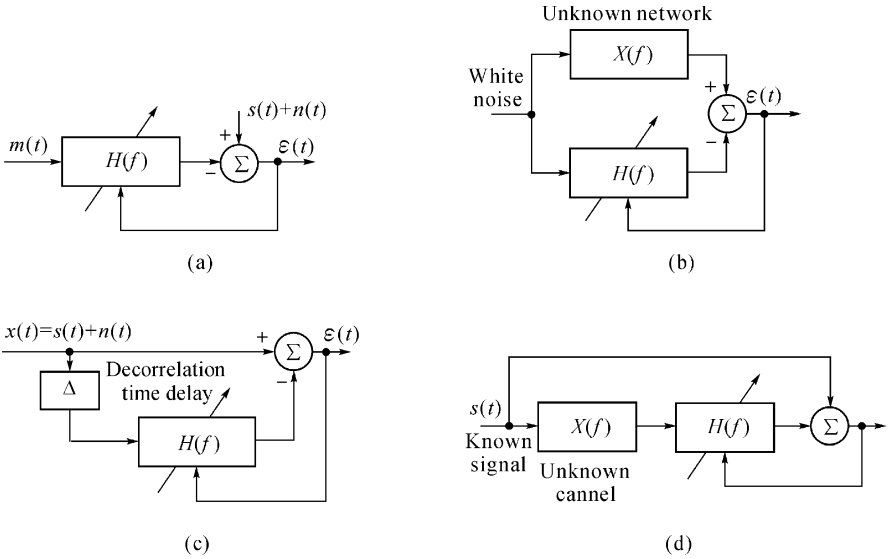


Fig. 3.15. Examples of adaptive filter applications. (a) Adaptive noise canceling; (b) Adaptive test of network; (c) Adaptive line enhancer; (d) Adaptve equalizer

Fig. 3.15(b) is a block diagram of an adaptive network characteristic tester. Suppose $X(f)$ is an unknown network, and that it is necessary to measure the transfer function of this network. By using an adaptive filter $H(f)$, the parameter can be calculated in real time. The white noise is excited in both filters of $H(f)$ and $X(f)$. The output of these two filters is expressed as $x_1(t)$, $x_2(t)$. $H(f)$ is adjusted such that

$$J = E[\varepsilon^2(t)] = E[x_1(t) - x_2(t)]^2$$

is minimized. Clearly, the minimum value will be reached when $H(f) = X(f)$. Thus, the transfer function of the unknown network will be solved.

Fig. 3.15(c) shows an adaptive filter called an "adaptive line enhancer". It is used to extract the single frequency component from a signal with a broadband continuous frequency spectrum. Suppose the system input signal is $x(t) = s(t) + n(t)$,

where $s(t)$ is a random signal with some single frequency components and $n(t)$ is white noise, independent of $s(t)$. Delay $x(t)$ to $x(t-\Delta)$ such that $n(t)$ and $n(t-\Delta)$ are uncorrelated. The signals $x(t-\Delta)=s(t-\Delta)+n(t-\Delta)$ pass through the filter $H(f)$ to become

$$\tilde{x}(t-\Delta)=\tilde{s}(t-\Delta)+\tilde{n}(t-\Delta) \tag{3.126}$$

to find the minimum value of

$$J=E[s(t)+n(t)-\tilde{s}(t-\Delta)-\tilde{n}(t-\Delta)]^2$$

$$=E[s(t)-\tilde{s}(t-\Delta)]^2+E[n^2(t)]+E[\tilde{n}^2(t-\Delta)]$$

This actually minimizes $E[s(t)-\tilde{s}(t-\Delta)]^2$ and $E[\tilde{n}^2(t-\Delta)]$. In other words, in Eq. (3.126), it is necessary to make $E[\tilde{n}^2(t-\Delta)]$ as small as possible and at the same time to make the difference between $s(t)$ and $\tilde{s}(t-\Delta)$ as small as possible. Thus, the single frequency components in $\tilde{x}(t-\Delta)$ are considerably enhanced.

Fig. 3.15(d) illustrates an adaptive equalizer of a transmission channel. This is an equalizing technique that is used extensively in transmission systems. The only assumption is the input signal $s(t)$ must be known and is sent to the receiving end. An adaptive filter is placed in the receiving end, too. It is necessary to adjust the adaptive filter such that the mean square error $E[\varepsilon^2(t)]$ is minimized. It is easy to prove that

$$H_{\text{opt}}(f)=1/X(f) \tag{3.127}$$

In this ideal case any signal transmitted from the input end should pass through the channel without distortion, i.e., the original signal is received at the output end.

Note that if the transfer function $X(f)$ of the transmission channel has some zero points, then $H_{\text{opt}}(f)$ will have some pole points. This will result in a very long tapped delay line which can sometimes be hard to realize.

3.5 Kalman Filtering for Non-stationary Digital Signal

The optimum filtering problem of a non-stationary random signal will be discussed in this section. We are not going to discuss the filtering problems of general non-stationary random signals because that is actually an unsolved

theoretical problem. There are some theoretical models that are already applied in radar and sonar. Most of these models were established for target tracking and prediction. Application topics in radar are quite different from those in sonar. As movement of a target in space is subject to the laws of universal gravitation and kinematics, the prediction of target movement is easier in radar than in sonar. The movement of underwater vessels is less restricted, making modeling more complicated.

3.5.1 Kalman Filtering of a One-Dimensional Observation Model

Kalman filtering, named after its inventor, is a kind of linear filtering that appears in an iteration format[45].

To introduce the technique of Kalman filtering and the necessity of an iteration algorithm, a simple example is derived as follows.

Suppose the observation value is

$$x(n) = s(n) + u(n) \tag{3.128}$$

where $s(n)$ is the signal sample and $u(n)$ is the observation noise. We aim to use the data of $x(1)$, $x(2)$, ..., $x(n)$ to estimate or predict the signal value in the $(n+1)$-th moment, i.e., $s(n+1)$ in Chapter 2, it was mentioned that the mean value is a good estimation

$$\hat{s}(n) = [x(1) + x(2) + ... + x(n)] / n \tag{3.129}$$

In this estimation, it is necessary to have a very large memory to store the data $x(1)$, $x(2)$, ..., $x(n)$, when n is sufficiently large. But if we rewrite Eq. (3.128) in this form, we obtain

$$\hat{s}(n+1) = \frac{x(1) + ... x(n-1)}{n} + \frac{x(n)}{n} = \frac{n-1}{n}\hat{s}(n-1) + \frac{1}{n}x(n) \tag{3.130}$$

This means that to obtain $\hat{s}(n+1)$, it needs to store only $\hat{s}(n)$ and new data $x(n)$. This is the advantage of an iteration algorithm. The Kalman algorithm is a specific form of iteration algorithm.

The most important aspect of introducing a new algorithm is to establish the right model. A one-dimensional Kalman filtering algorithm is presented as follows.

Suppose $s(n)$ is a Gaussian sequence with zero-mean, it is defined by the Markov model as

$$s(n) = a(n)s(n-1) + w(n) \tag{3.131}$$

where $a(n)$ is a given sequence, $w(n)$ is white noise and $E[w(n)] = 0$. In general, the sequence expressed in Eq. (3.130) is non-stationary. If $a(n)$ is not dependent on n, then $s(n)$ is stationary sequence. The observation data is shown in Eq. (3.128). It is supposed that the observation noise $u(n)$ is independent of $w(n)$ and $s(n)$. Denote the variance of $w(n)$ and $u(n)$ by σ_w^2 and σ_u^2, respectively. Our task is to find the linear estimate $\hat{s}(n+1)$ for $s(n+1)$ from data $x(1)$, $x(2)$, ..., $x(n)$ that has the minimum mean square error. By using the vector notation in section 3.4, it follows that

$$\hat{s}(n+1) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$$

$$I = E[s(n+1) - \hat{s}(n+1)]^2 = E[s^2(n+1)] - 2E[s(n+1)\hat{s}(n+1)] + E[\hat{s}^2(n+1)]$$

Differential I for vector $\boldsymbol{w}$ to obtain the gradient

$$\nabla_w I = 2E[(s(n+1) - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x})\boldsymbol{x}] = 0 \tag{3.132}$$

Eq. (3.132) is called the orthogonal principle in parameter estimation, i.e., the error vector $s(n+1) - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$ is orthogonal to $\boldsymbol{x}$. Since $s(n)$ is not a stationary random signal, it is impossible to directly resolve Eq. (3.132). In fact, there is no second order moment for $s(n)$. We now turn to the derivation of the optimum solution of the iteration procedure.

Suppose

$$\varepsilon(n) = s(n) - \hat{s}(n), \quad p(n) = E[\varepsilon^2(n)] \tag{3.133}$$

$$v(n) = x(n) - \hat{s}(n) = s(n) + u(n) - \hat{s}(n) \tag{3.134}$$

The quantity $v(n)$ in Eq. (3.134) is called innovation. It is the difference between the n-th observation and the estimation $\hat{s}(n)$ for signal. Since $\hat{s}(n)$ is the linear combination of $x(i)$ $(i = 1, 2, ..., n-1)$, $v(n)$ is the linear combination of $x(i)$ $(i = 1, 2, ..., n)$.

Based on the orthogonal principle, clearly $\varepsilon(n)$ is orthogonal to $\hat{s}(n)$, i.e., $E[\varepsilon(n)\hat{s}(n)] = 0$. Therefore,

$$\begin{aligned} E[s(n)v(n)] &= E[s(n)(\varepsilon(n)+u(n))] \\ &= E[s(n)\varepsilon(n)] \\ &= E[(\varepsilon(n)+\hat{s}(n))\varepsilon(n)] \\ &= E[\varepsilon^2(n)] = p(n) \end{aligned}$$

Thus

$$E[v^2(n)] = E[s(n)-\hat{s}(n)+u(n)]^2 = p(n)+\sigma_u^2(n) \tag{3.135}$$

Note that

$$\begin{aligned} s(n) &= a(n)s(n-1)+w(n) \\ \hat{s}(n+1) &= a(n+1)\hat{s}(n)+b(n)v(n) \end{aligned} \tag{3.136}$$

where

$$b(n) = a(n+1)\frac{E(s(n)v(n)}{\sigma_v^2(n)}$$

Therefore

$$\hat{s}(n+1) = a(n+1)+a(n-1)k(n)v(n) \tag{3.137}$$

and

$$k(n) = b(n)/a(n+1)$$

It follows that

$$\hat{s}(n+1) = a(n+1)[(1-k(n))\hat{s}(n)+k(n)x(n)] \tag{3.138}$$

$$p(n) = a^2(n+1)(1-k(n))p(n)+\sigma_w^2(n+1) \tag{3.139}$$

$$k(n) = \frac{p(n)}{p(n)+\sigma_u^2(n)} \tag{3.140}$$

Now the Kalman filtering procedure concludes as follows $n=1$, $p(1)=\sigma^2$ (usually greater than σ_u^2, σ_w^2), $\hat{s}(1)=s$ (usually is set to be 0).

Obtain data σ_u^2, σ_w^2, $a(n+1)$, $x(n)$, to find

$$k(n) = \frac{p(n)}{p(n) + \sigma_u^2(n)}$$

$$\hat{s}(n+1) = a(n+1)[\hat{s}(n) + k(n)(x(n) - \hat{s}(n))]$$

$$p(n) = a^2(n+1)(1 - k(n))p(n) + \sigma_w^2(n+1)$$

$$n \to n+1$$

The block diagram of this algorithm is shown in Fig. 3.16.

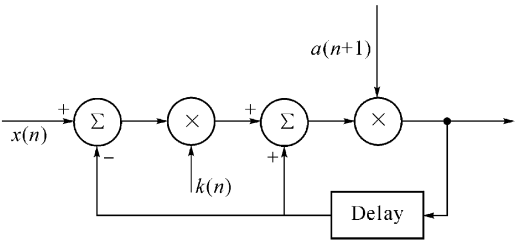


Fig. 3.16. Kalman filtering of a single channel

Example

Assume $s(n) = 0.6s(n-1) + w(n)$, $x(n) = s(n) + u(n)$ and $\sigma_w^2 = 0.25$, $\sigma_u^2 = 0.5$. Let $\hat{s}(1) = 0$, $p(1) = 1$. It follows that $k(1) = 2/3$, $\hat{s}(2) = 0.4x(1)$, $p(2) = 0.37$, $k(2) = 0.425$, $\hat{s}(3) = 0.138x(1) + 0.255x(2)$, $p(3) = 0.326$.

In the Kalman filtering model, if $a(n)$ is not dependent on n, and $w(n)$, $u(n)$ are stationary random series, then $p(n)$ has a limit value, denoted by p

$$p = a^2(1-k)p + \sigma_w^2, \quad k = p/(p + \sigma_u^2)$$

and

$$p = \lim_{n\to\infty} \frac{\sigma_w^2 + \sigma_u^2(a^2 - 1) + \{[\sigma_w^2 + \sigma_u^2(a^2 - 1)]^2 + 4\sigma_u^2\sigma_w^2\}^{1/2}}{2} \tag{3.141}$$

This is the error of Kalman filtering in steady state.

3.5.2 Kalman Filtering of Multiple Channels

In general, the detection and tracking of underwater targets are carried out using multiple sensors, because an array of multiple sensors has more high system gain, and therefore has long ranging distance and locating accuracy. Kalman filtering of

multiple channels is the generalized case of the one channel situation. We list only the final results and neglect the theoretical details

$$\boldsymbol{s}(n) = \boldsymbol{A}(n)\boldsymbol{s}(n-1) + \boldsymbol{w}(n) \tag{3.142}$$

$$\boldsymbol{x}(n) = \boldsymbol{H}(n)\boldsymbol{s}(n) + \boldsymbol{u}(n) \tag{3.143}$$

where $\boldsymbol{s}(n) : m \times 1$, signal vector; $\boldsymbol{A}(n) : m \times m$, state matrix; $\boldsymbol{w}(n) : m \times 1$, zero mean Gaussian white noise $\boldsymbol{x}(n) : q \times 1$, observation vector $\boldsymbol{H}(n) : q \times m$, matrix; $\boldsymbol{u}(n) : q \times 1$, measurement error vector.

Assume

$$E[\boldsymbol{w}(k)\boldsymbol{w}^{\mathrm{T}}(k)] = \begin{cases} Q(k) & k = i \\ 0 & k \neq i \end{cases}$$

$$E[\boldsymbol{u}(k)\boldsymbol{u}^{\mathrm{T}}(k)] = \begin{cases} R(k) & k = i \\ 0 & k \neq i \end{cases}$$

$$\boldsymbol{\varepsilon}(n) = \boldsymbol{s}(n) - \hat{\boldsymbol{s}}(n)$$

$$\boldsymbol{p}(n) = E[\boldsymbol{\varepsilon}(n)\boldsymbol{\varepsilon}^{\mathrm{T}}(n)]$$

The algorithm of Kalman filtering is as follows:

$$n = 1$$

$$\boldsymbol{p}(1) = \sigma^2 \boldsymbol{I} \ (\boldsymbol{I} \text{ is the unit matrix})$$

$$\hat{\boldsymbol{s}}(1) = \boldsymbol{s}$$

The obtained data are

$$\boldsymbol{R}(n),\ \boldsymbol{Q}(n),\ \boldsymbol{A}(n+1),\ \boldsymbol{x}(n)$$

$$\boldsymbol{K}(n) = \boldsymbol{p}(n)\boldsymbol{H}^{\mathrm{T}}(n)[\boldsymbol{H}(n)\boldsymbol{p}(n)\boldsymbol{H}^{\mathrm{T}}(n) + \boldsymbol{R}(n)]^{-1} \tag{3.144}$$

$$\hat{\boldsymbol{s}}(n+1) = \boldsymbol{A}(n+1)[\hat{\boldsymbol{s}}(n) + \boldsymbol{K}(n)(\boldsymbol{x}(n) - \boldsymbol{H}(n)\hat{\boldsymbol{s}}(n))] \tag{3.145}$$

$$\boldsymbol{p}(n+1) = \boldsymbol{A}(n+1)[(\boldsymbol{I} - \boldsymbol{K}(n)\boldsymbol{H}(n))\boldsymbol{p}(n)]\boldsymbol{A}^{\mathrm{T}}(n+1) + \boldsymbol{Q}(n+1) \tag{3.146}$$

$$n \to n+1 .$$

3.6 Parameter Estimation of Random Signal

In this section, we are going to discuss the problem of parameter estimation. As noted above, signal detection and parameter estimation are two basic topics in modern signal processing theory. To estimate parameters of random signal, it is necessary to set some assumptions for the statistical behavior of observational data. The term chosen to estimate a signal parameter is called a statistic. Different models result in different statistics.

In discussing signal detection problems, several models have been developed. Some models need the information of prior probability distribution and others need knowledge of the cost when an erroneous decision is made. From a theoretical point of view, the problem of random signal parameter estimation can be considered as a procedure in which the necessary information of the statistical characteristics of the estimated parameter is released step by step.

A Bayes estimate needs knowledge of the cost of error probability and the complete information of the prior probability distribution. The maximum estimate avoids the requirement of information of prior probability. Furthermore, the linear minimum mean square error estimate removes all assumptions about probability distribution, and needs only the information of the mean value and variance. As the last step, the minimum quadratic estimate considers the estimation problem as a popular function in mathematical analysis.

Clearly, the more prior information we have about the parameter to be estimated, the greater accuracy we can achieve. Therefore, in practice, the main principle is that we should use all the information that we can gather.

3.6.1 Test of Stationariness and Ergodicness of a Random Signal

To estimate the parameter of a random signal, it is possible to use only one realization sample of this random signal, which means that an assumption of ergodicness is inherent in the estimation procedure. Stationariness is certainly also a preset assumption.

In practice, the stationariness of random signal must be tested, but the ergodicness is hard to verify, because sometimes a large amount of samples are difficult to manage. However, we should remember that the random signal to be processed is an ergodic random signal, otherwise it is not reasonable to use the time average statistics to replace ensemble average statistics.

Note that for a Gaussian random signal, if the correlation function $R_{xx}(\tau)$ satisfies $\int |R_{xx}(\tau)| \, \mathrm{d}\tau < \infty$, then the signal is ergodic, which means that for a Gaussian random signal the stationariness test is equivalent to the ergodicness test.

In mathematical statistics theory, there are several methods that can be used to test if a random process is stationary. A useful method is to divide a random series

$x(n)$ ($n = 1, 2, ...$) into many groups, and then to calculate the mean value, variance and probability density function. The degree of dispersion of the results for each group is used to test if $x(n)$ is a stationary random series. The theoretical derivation is beyond the scope of this book.

3.6.2 Basic Requirements for a Statistic

In the problem of parameter estimation, there are some basic requirements for the statistic. Denote the parameter to be estimated as θ, and the statistic (or estimate) as $\hat{\theta}$.

Unbiased

The estimation value depends on the observation data, therefore the estimation value will change when we take a new set of data. Since the estimate itself is random, it also has a mean value, variance, etc. It is reasonable that the estimate value should be close to the true value, i.e.

$$E(\hat{\theta} - \theta) = 0 \tag{3.148}$$

The estimate which satisfies Eq. (3.147) is called an unbiased estimate, otherwise it is called biased and the difference between the estimated value and the true value is called the bias.

Consistency

When the amount of observation data increases, it is desirable that the density function of the estimate should concentrate in the neighborhood of the true value, i.e., the variance of the estimate should be close to zero, i.e., for any $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|\hat{\theta}_n - \theta| > \varepsilon\} = 0 \tag{3.148}$$

where θ_n represents the estimation value from n samples. Any estimate which satisfies Eq. (3.148) is called a consistent estimate.

Effectiveness

It is reasonable to hope that the estimate $\hat{\theta}$ we used has minimum variance

among all estimates, and therefore that the average error is minimized, i.e., for any other estimate $\tilde{\theta}$, we have

$$E[(\hat{\theta}-\theta)^2] \le E[(\tilde{\theta}-\theta)^2] \tag{3.149}$$

An estimate satisfying Eq. (3.149) is called an effective estimate.

In signal processing of digital sonar, incidental angle estimation from beam patterns, range estimation from echo time difference, and target speed estimation from Doppler frequency shift of carrier frequency all belong to the parameter estimation problem, although their physical meanings are quite different.

3.6.3 Some Estimates Used Frequently in Sonar Design

The estimates used frequently in sonar design include bayes estimate, minimum mean square estimate, median value estimation, maximum posterior estimate, maximum likelihood estimation.

3.6.3.1 Bayes Estimate

Suppose θ is the parameter to be estimated and X is the observation space. Estimate $\hat{\theta}(x)$ is the function on X. Consider the cost (or risk) function C as the function of $\varepsilon(x)=\theta-\hat{\theta}$. There are several different standards for the function $C(\varepsilon(x))$, three of which are considered here

$$C(\varepsilon(x)) = \varepsilon^2(x) = [\theta-\hat{\theta}(x)]^2 \tag{3.150a}$$

$$C(\varepsilon(x)) = |\varepsilon(x)| = |\theta-\hat{\theta}(x)| \tag{3.150b}$$

$$C(\varepsilon(x)) = \begin{cases} 1 & |\varepsilon(x)| \ge \Delta/2 \\ 0 & |\varepsilon(x)| < \Delta/2 \end{cases} \tag{3.150c}$$

If the probability density function $p(x,\ \theta)$ is known, then the average cost is

$$\overline{C} = \iint C(\varepsilon(x))p(x,\ \theta)\mathrm{d}x\mathrm{d}\theta = \int\left(\int C(\theta-\hat{\theta}(x))p(\theta \mid x)\mathrm{d}\theta\right)p(x)\mathrm{d}x \tag{3.151}$$

Since the integrand function of the first fold integral is non-negative, minimizing it is equivalent to minimizing $\overline{C}$, i.e., to find the minimum value of

$$\overline{C}(\hat{\theta} \mid x) = \int C(\theta - \hat{\theta}(x)) p(\theta \mid x) \mathrm{d}\theta \tag{3.152}$$

The estimate $\hat{\theta}(x)$ which minimizes Eq. (3.152) is called a Bayes estimate. It is actually the conditional expectation of the cost function.

3.6.3.2 Minimum Mean Square Estimate

Substituting $\overline{C}(\theta - \hat{\theta}) = (\theta - \hat{\theta})^2$ to Eq. (3.152) obtains

$$\overline{C}_{\mathrm{ms}}(\hat{\theta} \mid x) = \int (\theta - \hat{\theta})^2 p(\theta \mid x) \mathrm{d}\theta$$

$$\frac{\partial \overline{C}_{\mathrm{ms}}}{\partial \theta} = -2 \int \theta p(\theta \mid x) \mathrm{d}\theta + 2\hat{\theta} \int p(\theta \mid x) \mathrm{d}\theta$$

It is proved that

$$\hat{\theta}_{\mathrm{ms}} = \int \theta p(\theta \mid x) \mathrm{d}\theta \tag{3.153}$$

This is the minimum mean square estimate, also called the conditional mean square estimate.

3.6.3.3 Median Value Estimation

Substituting $\overline{C}(\theta - \hat{\theta}) = \mid \theta - \hat{\theta} \mid$ to Eq. (3.153) obtains

$$\begin{aligned} \overline{C}_{\mathrm{abs}}(\hat{\theta} \mid x) &= \int \mid \theta - \hat{\theta} \mid p(\theta \mid x) \mathrm{d}\theta \\ &= \int_{-\infty}^{\hat{\theta}} (-\theta + \hat{\theta}(x)) p(\theta \mid x) \mathrm{d}\theta + \int_{\hat{\theta}}^{\infty} (-\hat{\theta}(x) + \theta) p(\theta \mid x) \mathrm{d}\theta \end{aligned}$$

Taking the derivative on both sides, and letting the result equal zero, obtains

$$\int_{-\infty}^{\hat{\theta}_{\mathrm{abs}}} p(\theta \mid x) \mathrm{d}\theta = \int_{\hat{\theta}_{\mathrm{abs}}}^{\infty} p(\theta \mid x) \mathrm{d}\theta \tag{3.154}$$

where $\hat{\theta}_{\mathrm{abs}}$ is called the median value estimate, because it divides the integral into two equal parts, as we see in Eq. (3.154).

3.6.3.4 Maximum Posterior Estimate

Substituting Eq. (3.150c) to Eq. (3.152), it follows that

$$\overline{C}_{\text{unf}}(\hat{\theta} \mid x) = 1 - \int_{\hat{\theta}-\Delta/2}^{\hat{\theta}+\Delta/2} p(\theta \mid x)\mathrm{d}\theta$$

To find the minimum value of the left side in this expression, it is necessary to find the maximum value of the integral on the right side. If Δ is small, it is equivalent to maximizing $p(\theta \mid x)$ in the segment $[\hat{\theta}_{\text{unf}} - \Delta/2,\ \hat{\theta}_{\text{unf}} + \Delta/2]$. In other words, we have to find the maximum value of $p(\theta \mid x)$ for θ. So this estimate is called the maximum posterior estimate $\hat{\theta}_{\text{map}}$. Since function ln y is monotonic for variable y, $\hat{\theta}_{\text{map}}$ can be resolved using Eq. (3.155)

$$\frac{\partial \ln p(\theta \mid x)}{\partial \theta}\Big|_{\theta=\hat{\theta}_{\text{map}}} = 0 \tag{3.155}$$

3.6.3.5 Maximum Likelihood Estimation

The maximum posterior estimation must have the information of $p(\theta)$. If $p(\theta)$ is unknown, we can find another way to choose a new estimate instead of $\hat{\theta}_{map}$.

Since

$$p(\theta \mid x) = \frac{p(x \mid \theta)p(\theta)}{p(x)}$$

we can find the maximum value of $p(x \mid \theta)$. This yields the maximum likelihood estimate $\hat{\theta}_{\text{ml}}$. It is the solution of Eq. (3.156)

$$\frac{\partial \ln p(x \mid \theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}_{\text{ml}}} = 0 \tag{3.156}$$

3.6.4 Cramer-Rao Low Bound

In statistical mathematics theory, the Cramer-Rao low bound is one of the most important theorems as it gives the best possible result that an estimate can attain.

The basis of this theorem is as follows: assume we want to estimate some parameter of a random signal and have already chosen a suitable estimate. If the estimation error is equal to the Cramer-Rao low bound, then this theorem tells us that it is not necessary to find other estimates.

Theorem 3.21 (Cramer-Rao Theorem) *Assuming the likelihood function of parameter* θ *is* $L(x \mid \theta)$ *and the expectation value of estimate* $\hat{\theta}$ *is* $E(\theta) = \psi(\theta)$, *then the variance* $\mathrm{Var}(\hat{\theta})$ *of* $\hat{\theta}$ *satisfies the following inequality*

$$\mathrm{Var}(\hat{\theta}) \geq \frac{(\partial \psi(\theta) / \partial \theta)^2}{E[\partial \ln L(x \mid \theta) / \partial \theta]^2} \tag{3.157}$$

In particular, if the estimate $\hat{\theta}$ is unbiased, i.e., $\psi(\theta) = \theta$, Eq. (3.157) can be simplified as

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{E[\partial \ln L(x \mid \theta) / \partial \theta]^2} \tag{3.158}$$

Proof: Since the likelihood function is also the probability density function, it follows that $\int L(x \mid \theta)\mathrm{d}x = 1$. Deriving this expression for θ, we obtain

$$\int \frac{\partial L(x \mid \theta)}{\partial \theta}\mathrm{d}x = \int \frac{\partial \ln L(x \mid \theta)}{\partial \theta} L(x \mid \theta)\mathrm{d}x = 0$$

i.e.

$$E\left[\frac{\partial \ln L(x \mid \theta)}{\partial \theta}\right] = 0$$

Since the expectation operation is for variable x, for any function of $\psi(\theta)$, it should be

$$E\left[\psi(\theta)\frac{\partial \ln L(x \mid \theta)}{\partial \theta}\right] = 0$$

Based on the assumption, we have $E(\hat{\theta}) = \int \hat{\theta} L(x \mid \theta)\mathrm{d}x = \psi(\theta)$.

Taking the derivate for θ, we obtain

$$\frac{\partial \psi(\theta)}{\partial \theta} = \int \hat{\theta} \frac{\partial \ln L(x \mid \theta)}{\partial \theta} L(x \mid \theta)\mathrm{d}x$$

The expectation of

$$\psi(\theta)\left(\frac{\partial \ln L(x \mid \theta)}{\partial \theta}\right)$$

is zero, therefore

$$\int[\hat{\theta}-\psi(\theta)]\frac{\partial \ln L(x \mid \theta)}{\partial \theta}L(x \mid \theta)\mathrm{d}x=\frac{\partial \psi(\theta)}{\partial \theta}$$

By using the Schwartz inequality, it follows that

$$\int[\hat{\theta}-\psi(\theta)]^2 L(x \mid \theta)\mathrm{d}x\int\left[\frac{\partial \ln L(x \mid \theta)}{\partial \theta}\right]^2 L(x \mid \theta)\mathrm{d}x \geq \left[\frac{\partial \psi(\theta)}{\partial \theta}\right]^2$$

Thus, we obtain

$$\mathrm{Var}(\hat{\theta}) \geq \frac{\left[\frac{\partial \psi(\theta)}{\partial \theta}\right]^2}{E\left[\frac{\partial \ln L(x \mid \theta)}{\partial \theta}\right]^2}$$

The right sides of Eqs. (3.157) and (3.158) are called the Cramer-Rao low bound. This is a limit value which an estimate can reach.

If we use the expression

$$E\left[\frac{\partial \ln L(x \mid \theta)}{\partial \theta}\right]^2=-E\left[\frac{\partial^2 \ln L(x \mid \theta)}{\partial \theta^2}\right] \tag{3.159}$$

then we can obtain another form of expression of the Cramer-Rao low bound.

It is necessary to show that an estimate with an estimation variance which is equal to the Cramer-Rao low bound does not always exist. If it does exist, it must be the maximum likelihood estimate and also be unique. However, because the proof is complicated and beyond the scope of this book, we will not discuss this here.

3.6.5 *Example (Mean Value Estimate)*

Assume we have N independent data observations x_i $(i = 1, ..., N)$ for one parameter θ and that the observation error $x_i - \theta$ obeys a Gaussian probability

distribution. Then the maximum likelihood estimate for θ is the mean value estimate, i.e.

$$\theta_{\text{ml}} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Based on the assumptions, we have

$$p(x_i \mid \theta) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x_i - \theta)^2}{2\sigma^2}\right] \tag{3.160}$$

Since x_i $(i = 1, \ ..., \ N)$ are independent, we have

$$p(\boldsymbol{x} \mid \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x_i - \theta)^2}{2\sigma^2}\right] \tag{3.161}$$

Taking the logarithm and deriving it, we have

$$\frac{\partial \ln p(\boldsymbol{x} \mid \theta)}{\partial \theta} = \frac{1}{\sigma^2}\left[\sum_{i=1}^{N} -(\theta - x_i)\right] \tag{3.162}$$

Letting Eq. (3.162) equal zero, we obtain

$$\hat{\theta}_{\text{ml}} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The Cramer-Rao low bound can be found in the following expression

$$\frac{\partial^2 \ln p(\boldsymbol{x} \mid \theta)}{\partial \boldsymbol{x}^2} = -\frac{N}{\sigma^2}$$

$$\text{Var}(\hat{\theta}_{\text{ml}} - \theta) = \sigma^2 / N \tag{3.163}$$

3.6.6 *Model-Free Estimates*

The estimates discussed above depend more or less on the model of the observation data. Model-free estimates can be understood that do not depend on models. These kinds of estimates are especially useful for estimating the mean value, variance, correlation function and power density spectrum of random signals.

Suppose the sample sequences are $x(0),\ ...,\ x(N-1)$.

Mean Value Estimation

$$\overline{x} = \frac{1}{N}\sum_{i=0}^{N-1} x(i) \tag{3.164}$$

It is easy to prove that $\overline{x}$ is unbiased. The variance of the mean value estimation is

$$\mathrm{Var}(\overline{x}) = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sigma_{ij}$$

where $\sigma_{ij} = E[(x(i)-\overline{x})(x(j)-\overline{x})]$ is the covariance of $x(i)$ and $x(j)$. For the stationary white noise, $\sigma_{ij} = 0,$ for $i \neq j$. Therefore

$$\mathrm{Var}(\overline{x}) = \frac{1}{N}C_{xx}(0)$$

Estimation of Auto-Correlation Function

$$\hat{R}_{xx}(k) = \frac{1}{N-k}\sum_{i=0}^{N-k-1} x(i)x(i+k)\ (k = 0,\ ...,\ N-1) \tag{3.165}$$

Note $\hat{R}_{xx}(-k) = \hat{R}_{xx}(k)$.

Since the number of samples is limited, for different *k*, the number of samples in the average expression differs. This is actually the truncation effect in parameter estimation.

It is easy to prove that

$$E[\hat{R}_{xx}(k)] = R_{xx}(k)\ \ (k \le N-1)$$

$$\mathrm{Var}[\hat{R}_{xx}(0)] = \frac{2\sigma_x^4}{N}$$

$$\mathrm{Var}[\hat{R}_{xx}(k)] = \frac{\sigma_x^4}{N-|k|}\ \ (1 \le |k| \le N-1)$$

Estimation of Power Spectrum Density Function

As we know

$$K_{xx}(f) = \int R_{xx}(\tau)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau$$

Therefore, it is reasonable to adopt the following expression for estimating the power spectrum of a random signal

$$\hat{K}_{xx}(f) = \int \hat{R}_{xx}(\tau)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau \tag{3.166}$$

$$\hat{K}_{xx}(f) = \sum_{k=-(N-1)}^{N-1} \hat{R}_{xx}(k)\exp(-2\pi \mathrm{j} kf) \left(|f| < \frac{1}{2} \right) \tag{3.167}$$

where $\hat{R}_{xx}(k)$ is defined by Eq. (3.165).

References

[1] Wiener, N.: Cybernetics, or Control and Communication in the Animal and the Machine, 2nd Edn. MIT Press, London (1962)

[2] Shannon, C.: A mathematical theory of communication. Bell System Tech. J. 27-3, pp. 379-423 (1948)

[3] Hancock, J. C., Wintz, P. A.: Signal Detection Theory. McGraw-Hill, New York (1966)

[4] Van Tress, H. L.: Detection, Estimation, and Modulation Theory. Wiley, New York (1968)

[5] Chung, K. L.: Elementary Probability Theory with Stochastic Process, 2nd Edn. Springer-Verlag, New York (1975)

[6] Feller, E.: An Introduction to Probability Theory and its Applications, 2nd Edn. Wiley, New York (1957)

[7] McDonough, R. N., Whalen, A. D.: Detection of Signals in Noise, 2nd Edn. Academic Press, Maryland Heights (1995)

[8] Moon, T. K.: Similarity methods in signal processing. IEEE Trans. SP-44-4, pp. 827-833 (1996)

[9] Papoulis, A.: Probability, Random Variable and Stochastic Process. McGraw-Hill, New York (1965)

[10] Wald, A.: Sequential Analysis. Wiley, New York (1959)

[11] Winder, A., Loda, C. J.: Space-time Information Processing, 1st Edn. Peninsula Pub., Wisconsin (1988)

[12] Swets, J. A. (Eds.): Signal Detection and Recognition by Human Observers. Penisula Pub., Wisconsin (1988)

[13] Schultheiss, P. M., Weinstein, E.: Detection and estimation: a summary of results. In: Proc. of NATO ASI on Underwater Acoustics and Signal Processing, pp. 379-410 (1980)
[14] Burdic, W. S.: Radar Signal Processing. Prentice Hall, Englewood (1968)
[15] Widrow, B., Steams, S. D.: Adaptive Signal Processing. Prentice Hall, Englewood (1985)
[16] Widrow, B., Mantey, P. E., Griffiths, L. J., Goode, B. B.: Adaptive antenna systems. In: Proc. of IEEE 57, pp. 2143-2159 (1967)
[17] Griffiths, L.: A simple adaptive algorithm for real time processing in antenna arrays. In: Proc. of IEEE 57, pp. 1696-1704 (1969)
[18] Anderson, V. C.: Side lobe interference suppression with an adaptive null processor. J. Acoust. Soc. Amer. 69, pp. 185-190 (1981)
[19] Monzino, R. A., Miller, T. W.: Introduction to Adaptive Arrays. Wiley, New York (1980)
[20] Poor, H. V.: On robust Wiener filtering. IEEE Trans. AC-25, pp. 531-536 (1980)
[21] Gilsson, T. H., Black, C. I., Sage, A. P., Incorporated, T. I.: On sonar signal analysis. IEEE Trans. AES-6, pp. 37-50 (1970)
[22] Owsley, N. L.: Sonar Array Processing. In: Array Signal Processing, Haykin, S. (Eds), Prentice Hall, Englewood (1985)
[23] Hinich, M. J., Bloom, M. C.: Statistical approach to passive target tracking. J. Acoust. Soc. Amer. 69, pp. 167-173 (1981)
[24] Deprettere, E. (Eds): SVD and Signal Processing. North Holland, New York (1988)
[25] Thompson, J. S., Grant, P. M., Mulgrew, B.: Performance of spatial smoothing algorithm for correlated sources. IEEE Trans. SP-44-4, pp. 1041-1045 (1996)
[26] Peterson, W. W., Birdsall, T., Fox, W.: The theory of signal detectability. IRE Trans. IT-4, pp. 171-212 (1954)
[27] Kay, S. M.: Fundamentals of Statistical Signal Processing: Detection Theory. Prentice Hall, Englewood (1993)
[28] Kay, S. M.: Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice Hall, Englewood (1993)
[29] Knight, W. C., Pridham, R. G., Kay, S. M.: Digital Signal Processing for Sonar, In: Proc. of IEEE 69-11, pp. 1451-1506 (1981)
[30] Turin, G. L.: An introduction to matched filters. IRE Trans. IT-6, pp. 311-330 (1960)
[31] Gibson, J. D., Melsas, J. L.: Introduction to Nonparametric Detection with Applications. IEEE Press, New Jersey (1996)
[32] Chen, T. et al., (Eds.): The Past, Present and Future of Multimedia Signal processing. IEEE Signal Processing Magazine, pp. 28-51 (1997)
[33] Mohanty, N.: Random Signals Estimation and Identification. Van Norstrand, New York (1986)
[34] Krishnaiah, P. R., Kandal, L. N.: Handbook of Statistics, Vol.2. North Holland Pub., New York (1982)

[35] Oppenhaim, A. V., Willsky, A. S. Hamid, S.: Signals and Systems. Prentice Hall, Englewood (1985)
[36] Edelblute, D. J., Fisk, J. M., Kinnison, G. L.: Criteria for optimum signal detection theory for arrays. J. Acoust. Soc. Amer. 41, pp. 199-205 (1967)
[37] Schtoeder, M. R.: Linear prediction, entropy and signal analysis. IEEE ASSP Magazine, pp. 3-11 (1984)
[38] Feinstain, A.: Foundation of Information Theory. Mc-Graw Hill, New York (1958)
[39] Middleton, D., Esposito, R.: Simultaneous optimum detection and estimation of signal in noise. IEEE IT-14-3, pp. 434-444 (1968)
[40] Jaffer, A.G., Gupta, S. C.: Coupled detection-estimation of Gaussian processes in Gaussian noise. IEEE IT-18-1, pp. 106-110 (1972)
[41] Olmo. G., Magli, E., Presti, L. L.: Joint statistical signal detection and estimation Part I : theoretical aspects of the problems. Signal Processing 80, pp. 57-73 (2000)
[42] Shin, F. B., Kil, D. H.: Full spectrum signal processing using a classify-before-detection paradigm. J. Acoust. Soc. Amer. 99-4 pp.2188-2197 (1996)
[43] http://en.wikipedia.org. Accessed 1 Jan 2011
[44] Wiener, N.: Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications. Wiley, New York (1949)
[45] Kalman, R. E.: A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Engr. 82D, pp. 34-45 (1960)

4

General Principles of Sonar Design

The dry end of modern digital sonar is actually a large, multi-functional digital computer. The design of this computer depends on the configuration of the wet end and on our knowledge and understanding of the ocean environment and the behavior of underwater sound propagation. Therefore, sonar design is a combination of underwater acoustic physics, underwater acoustic engineering, radio electronics and transducer manufacture.

The sonar equation, as a popular tool, is used extensively in sonar theoretical studies and in discussions of technical feasibility. The sonar equation combines the most important concepts in transducers, underwater sound channels, ocean environments (ambient noise, reverberation), target characteristics (radiated noise or signal reflect ability), optimum detection of signal, and signal wave form design.

In this chapter, starting from the sonar equation, the general design principles of sonar design are discussed. Our focus is on the general theory of sonar design rather than on digital or analogue systems. The method for using the sonar equation is introduced and the various parameters appearing in the sonar equation are explained in detail. Some important concepts in underwater acoustic physics are also introduced.

The basic theoretical principles of sonar design and related signal processing techniques can be found in many classical publications in the field of underwater acoustics[1-12].

4.1 Determination of Sonar System Specifications

Any sonar has its own purpose, and the mission requirements are the most important factor in sonar design. In designing a specific sonar system, it is first necessary to define the main task and tactical mission of the sonar system, and then to define its required performance and specifications. Therefore, the tactical and technical specifications should guide the design and direct the realization of a

sonar system. In this section we will discuss how to define the technical specifications based on the sonar task and tactical mission.

4.1.1 Category of Sonar

The general procedure of sonar design is shown in Fig. 4.1. It is divided into four steps: task / mission definition, tactical specifications, functional module and decomposition of technical specifications.

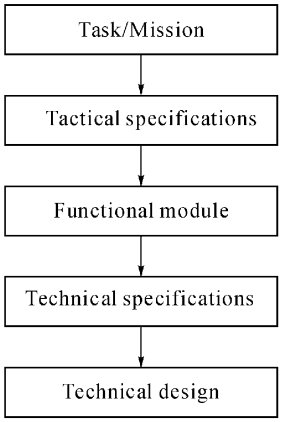


Fig. 4.1. Procedure of sonar design

The task or mission for a sonar system includes the installation objective, the main performance requirements, the operation environment and the relationship with other equipment already installed in vessels or on platforms. Tactical specifications include the range distance, the accuracy of ranging, the bearing accuracy, the resolution ability of multi-targets and the target recognition ability. For the civil use of sonar, the tactical specifications are often included in the technical specifications. Various functional modules of a sonar system are necessary for the realization of tactical specifications. The fourth step in sonar design is to decompose the technical specifications into a functional module, so that the hardware designer can implement the technical specifications. The technical specifications are the guarantee for the performance of the sonar tactical mission. For example, to reach the ranging accuracy, it is necessary to define the technical requirements for time delay estimation, to realize the reliability specification, and to propose the life test specifications of transducers and electronic components.

There are many different kinds of sonar systems. The technical specification strongly depends on the type of sonar. We will briefly introduce the methods for classifying sonar.

Based on the installation platform or vessel, sonar can be classified as surface ship sonar, submarine sonar, helicopter sonar, shore-based sonar, sonobuoy or torpedo sonar, etc. (Fig. 4.2).

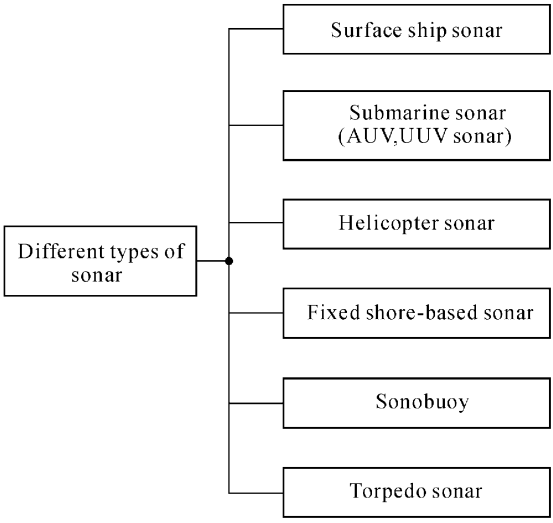


Fig. 4.2. Classification of sonar by platform

If we classify a sonar system based on its mission, it can be categorized as active sonar, passive sonar, underwater communication sonar, underwater counter-measure sonar, torpedo alarming sonar or sonar decoy, etc. (Fig. 4.3).

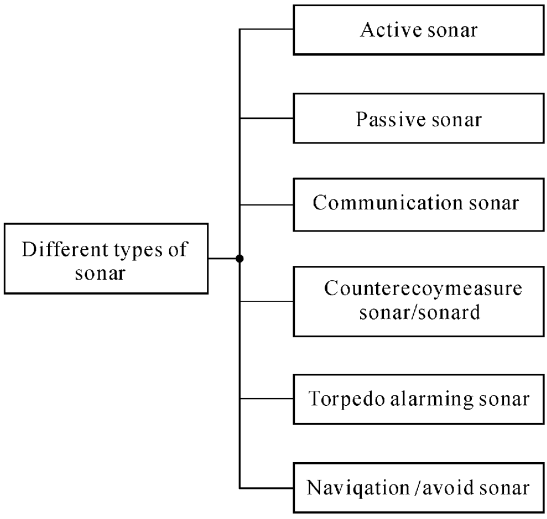


Fig. 4.3. Classification of sonar by function

Active sonar has the capability for echo ranging and Doppler speed measurement. Passive sonar has passive bearing or passive ranging performance.

A sonar system can be also categorized based on the installation position of the wet end: bow array sonar, conformal array sonar, flank array sonar, towed linear array sonar and variable depth sonar, sonobuoy, etc. (Fig. 4.4)

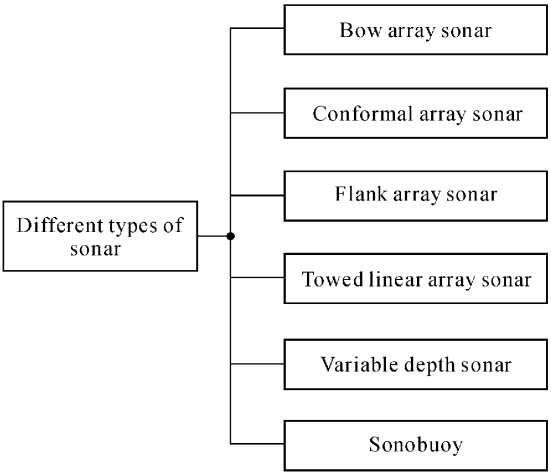


Fig. 4.4. Classification of sonar by the wet end or installation position

If we are concerned about the finer classes of sonar systems, each category can be divided into many sub-categories, such as side scan sonar, mine avoidance sonar, sea bottom image sonar, and mine hunting sonar. A modern submarine has up to 20 different types of sonar systems.

4.1.2 Relationship between Tactical and Technical Specifications

The tactical and technical specifications of sonar systems are two sides of one problem. The requirements of practical application can be viewed from different perspectives. Tactical specifications reflect military requirements, i.e., the kind of performance needed from a specific sonar system. Technical specifications reflect the concerns of the sonar designer, i.e., the kind of specifications needed, such that the sonar system can be designed to meet the military requirements. Some tactical and technical specifications are the same, for example, the reliability demand and maintainability.

To gain a better understanding of the tactical and technical specifications, we will briefly illustrate the basic working scenario of various sonar systems. This will show how the tactical and technical specifications are inherent in sonar design philosophy and how a sonar designer can trade off the military demands.

Fig. 4.5 shows a block diagram of a typical active sonar system, with the

transmitting and receiving function using the same array. Usually, the beamforming unction in the receiver is different from that in the transmitter because the narrowband transmitting beam can be controlled by a phased shift network, but the receiving beamforming is wideband.

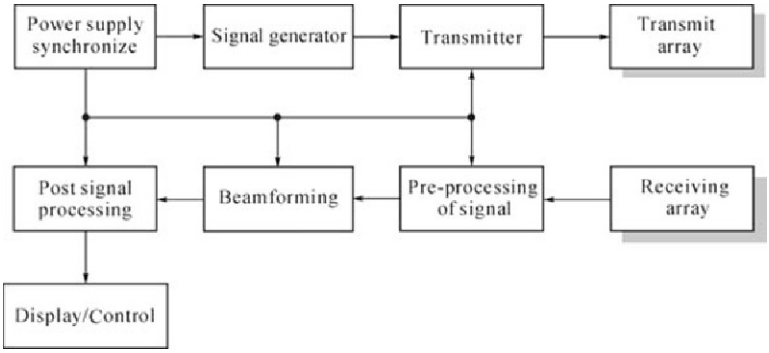


Fig. 4.5. Block diagram of the principle of active sonar

Fig. 4.6 is a block diagram of passive sonar. The signal processor in passive sonar is usually divided into two parts, narrowband processing and wideband processing, because spectrum analysis is necessary for parameter estimation and target recognition.

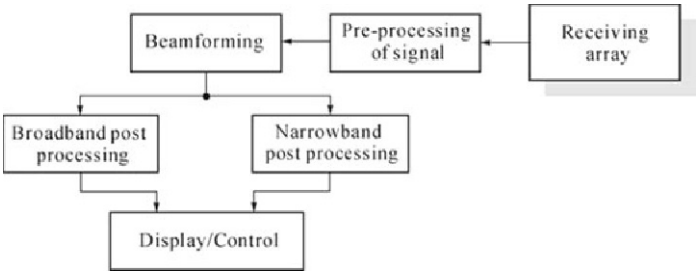


Fig. 4.6. Block diagram of signal processing of passive sonar

Fig. 4.7 shows a block diagram of a phased control transmitter. A phase shift network controls the steering direction of the transmitting beam. Therefore, the transmitting energy can be concentrated in a preset direction. The phased control network makes the beam scan in 360° space such that the active sonar realizes all-direction detection.

Fig. 4.8 illustrates the principle of a beamforming system of an arc receiving array. A preset time delay network can form multiple beams which steer in a 0 – 360° direction. Usually, the incoming underwater sound wave to the receiver is expected to be a plane wave, except in the case of passive ranging sonar. The assumption of a plane wave or spherical wave is important in the design of the time delay calculation of preset beamforming.

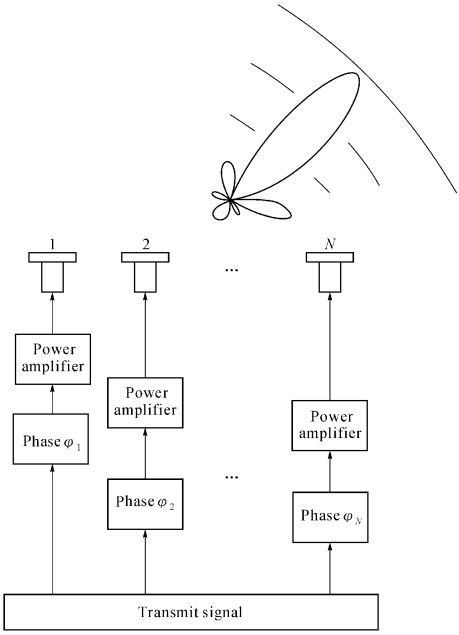


Fig. 4.7. Illustration of phased control transmitter

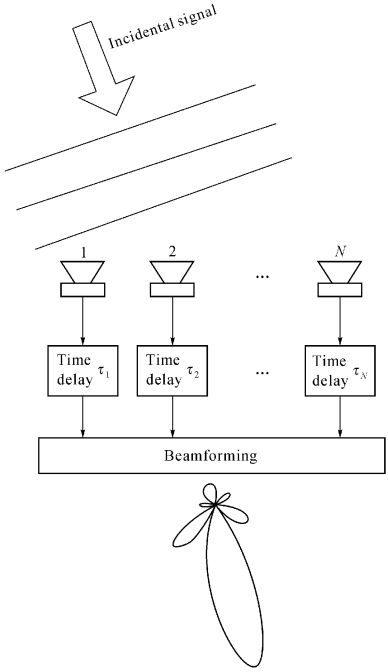


Fig. 4.8. Principle of pre-time delay beamforming

In Figs. 4.9 to 4.15, a series of block diagrams of sonar systems are shown and some typical tactical applications of modern sonar are presented[13].

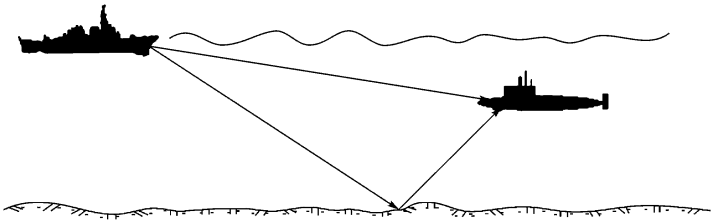


Fig. 4.9. Active sonar on a surface ship (for echo detection of submarines)

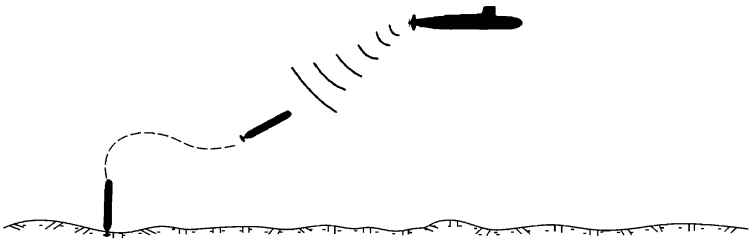


Fig. 4.10. Passive detection of submarines by sea bed mounted torpedoes

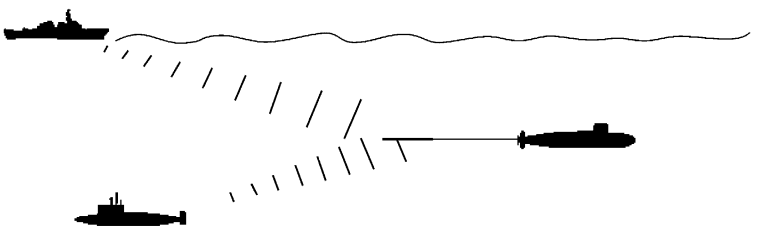


Fig. 4.11. Passive detection of underwater target noise

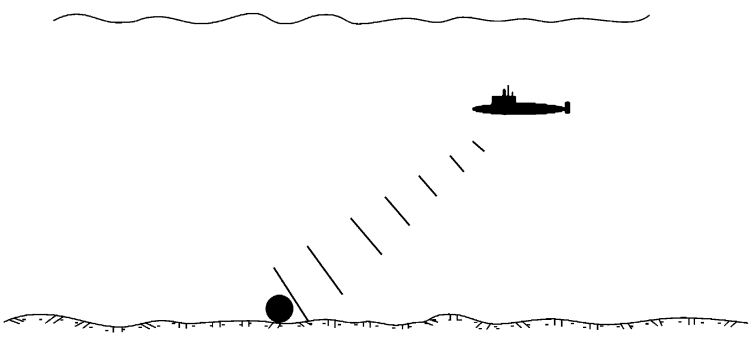


Fig. 4.12. Passive detection of underwater target noise by mines

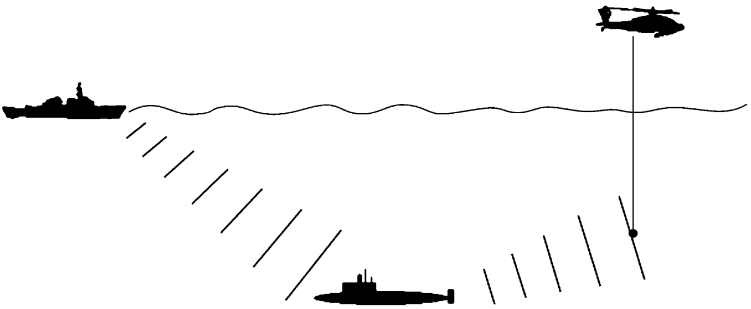


Fig. 4.13. Submarine detection by combined surface ship and helicopter

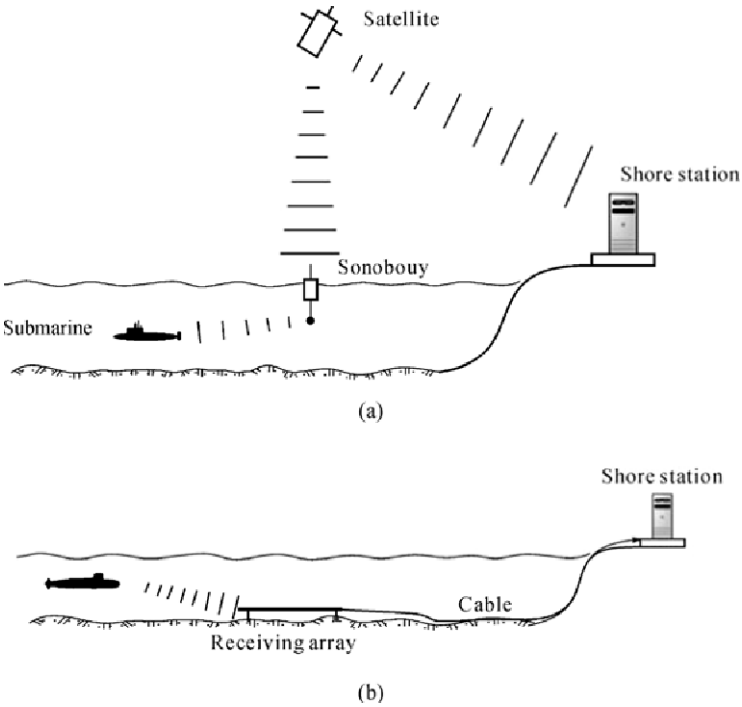


Fig. 4.14. (a) Communication between submarine and shore-based station; (b) Shore-based sonar system

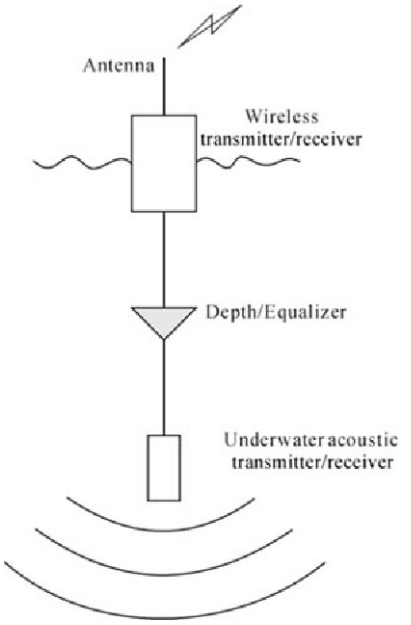


Fig. 4.15. Illustration of sonobuoy operation

4.1.3 Technical Specification Related Concepts in Sonar Design

Before we describe the technical specifications of a sonar system, it is necessary to introduce some concepts that are related to the quantitative description of sonar specifications. These quantitative descriptions are necessary for the discussion of sonar design and the verification of design results. In any situation, the technical specifications defined must be tested quantitatively, either in a laboratory, lake or sea trial. Specifications which can not be tested are meaningless in practice.

4.1.3.1 Sound Pressure and Sound Power

In the centimeter gram second (CGS) unit system, the unit of sound pressure is the bar

$$1\text{bar} = 1\text{dyn}/\text{cm}^2 \tag{4.1}$$

Sound power is equal to the square of sound pressure, i.e.

$$I = p^2 \tag{4.2}$$

When p is in bar, and I is in W/cm^2, it follows that

$$I = 0.66 \times 10^{-12} p^2 \tag{4.3}$$

or

$$p = 1.23 \times 10^6 \sqrt{I} \tag{4.4}$$

It is usual to use dB for expressing sound pressure and sound power. If a sound pressure p_0 is chosen to be a reference sound pressure, define it as 0 dB, and then any other sound pressure p has the dB value of $20 \lg p / p_0$ dB.

Similarly, if sound power I_0 is the 0 dB reference, then any other power I has the dB value of $10 \lg I / I_0$ dB.

Since 1971, the reference value of sound pressure has been expressed in μPa or kPa, 1μPa = 10^{-6} Pa = 10^{-6} N / m^2.
It is easy to calculate the relationship between bar and μPa, that is

$$1 \text{ Pa} = 1 \text{ N/m}^2 = \frac{1000 \text{ g} \times 100 \text{ cm}}{10^4 \text{cm}^2} = 10 \text{ dyn/cm}^2 = 10 \text{ bar}$$

4.1.3.2 Source Level and Spectrum Level

In describing the quantity of radiated sound, the concept of sound level is usually adopted. This indicates the sound pressure or sound intensity per unit area, received at a distance of 1 m from a sound source. The value is called the index sound pressure or index sound intensity. If we sketch the sound spectrum, then the sound level is actually the sound energy in a certain frequency band. Suppose the power spectrum of sound signal $x(t)$ is $K_x(f)$ and the unit is Pa^2 / Hz or bar^2 / Hz, then the sound energy in the band $[W_1, W_2]$ is

$$SL_0 = \int_{W_1}^{W_2} K_x(f) \mathrm{d}f \tag{4.5}$$

It is easy to translate SL_0 into a dB value, i.e.

$$SL = 10 \lg SL_0 \tag{4.6}$$

SL is called the dB value of a sound level and $\sqrt{K_x(f)}$ is called the spectrum level. The dB value is expressed as

$$SPL(f) = 20\lg\sqrt{K_x(f)} \tag{4.7}$$

A simplified, but not accurate method, to find SL from $SPL(f)$ is

$$SL = SPL(f_0) + 10\lg[W_2 - W_1] \tag{4.8}$$

where f_0 is some value between W_1 and W_2. Sometimes $f_0 = (W_1 + W_2)/2$ or $f_0 = \sqrt{W_1 W_2}$ is an acceptable choice.

A typical signal spectrum is shown in Fig. 4.16, where the horizontal axes are frequency and the vertical axes are spectrum sound level. This is a semi-log curve, and is a straight line. In most cases, in a semi-log figure, signal spectrum is a straight line. This phenomenon can be explained by the following examples.

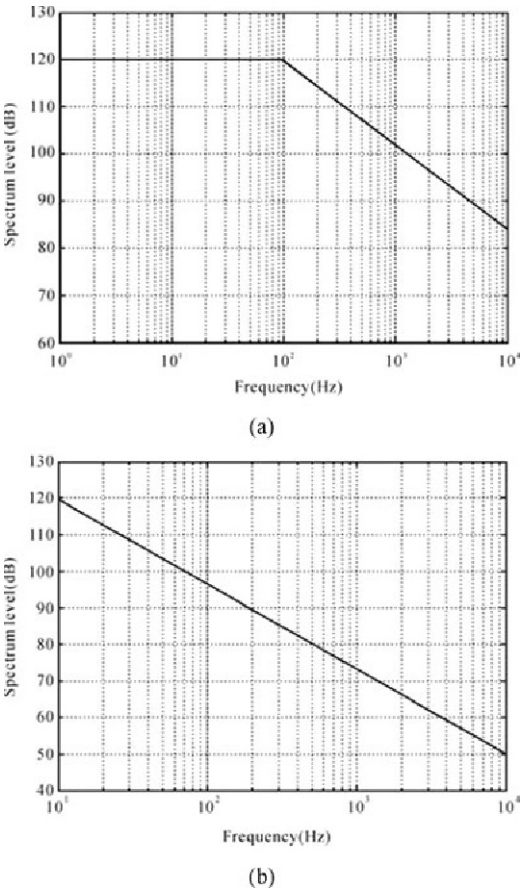


Fig. 4.16. Typical spectrum of target noise and ambient noise. (a) Typical spectrum of target noise; (b) Typical spectrum of ambient noise

Examples

(1) Spectrum of white noise. Suppose the signal is white noise with an auto-correlation function

$$R_x(\tau) = \frac{N}{2}\delta(\tau) \tag{4.9}$$

where $\delta(\cdot)$ is the Dirac function. Its power spectrum is

$$K_x(f) = \frac{N}{2}\int \delta(\tau)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau = N/2 \quad (-\infty < f < \infty) \tag{4.10}$$

This is a linear function of f and is, of course, a straight line in semi-log figure.

(2) Signal with an exponential auto-correlation function. In many situations the auto-correlation function of signal $x(t)$ has the form

$$R_x(\tau) = \sigma_x^2 \exp(-\alpha|\tau|) \quad (\alpha > 0) \tag{4.11}$$

Therefore, the power spectrum is

$$K_x(f) = \sigma_x^2 \int \exp(-\alpha|\tau|)\exp(-2\pi \mathrm{j} f\tau)\mathrm{d}\tau = 2\sigma_x^2 \frac{\alpha}{\alpha^2 + 4\pi^2 f^2} \tag{4.12}$$

When $f >> \alpha$, we have

$$K_x(f) \approx 2\sigma_x^2 \frac{\alpha}{4\pi^2 f^2}$$

It follows that

$$10\lg K_x(f) = 10\lg\left(\frac{\alpha\sigma_x^2}{2\pi^2}\right) - 20\lg f \tag{4.13}$$

This is also a straight line in semi-log figure.

We used to express the power spectrum in a way similar to that shown in Fig. 4.16 for radiated noise and ambient noise. The reference level of the sound level or spectrum level is 1 $\mu\mathrm{Pa}^2$/Hz or 1 $\mu\mathrm{Pa}/\sqrt{\mathrm{Hz}}$.

Except for the sound level of radiated noise, the sound level of a transmitting signal of active sonar can be calculated. Suppose the transmitting signal is spread in a spherical shape in a very small region. The unit of sound power P is W. The index of sound pressure p with unit bar at a distance of 1 m from a source is

$$P = 0.66\times10^{-12}\times p^2\times4\pi\times10^4$$

$$p^2 = \frac{P}{0.66\times10^{-12}\times4\pi\times10^4}$$

If P = 1 W, then p = 3464.1 bar. The index sound level (at 1 m) is

$$SL(1\text{ W}) = 20\lg p = 70.8\text{ dB}\quad(\text{ref. 1 bar} = 0\text{ dB}).$$

If we consider 1 μPa as 0 dB, then based on Eq. (4.5), 1 bar = 10^5 μPa, SL(1W) = 170.8 dB. Note that these values are a little different from those that appeared in traditional English publications where 1 yard was usually taken as the standard distance.

4.1.3.3 Detection Index

In detection procedures, regardless of whether active sonar detection in time / frequency bins or passive sonar detection in bearing bins is used, the "target present" decision is made using signal post-processing. If in some resolution bins, the result of post-processing is affected by accidental changes, then we can say only that the target has some probability of being present. Fig. 4.17 shows the procedure of target detection in the post-processing of a sonar system.

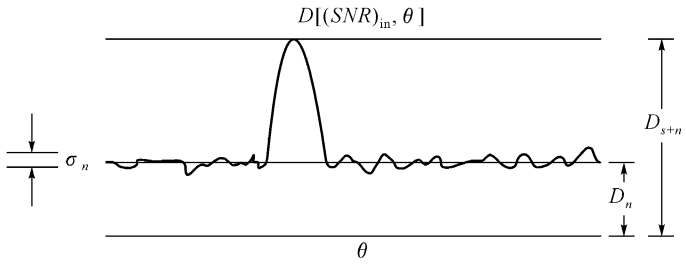


Fig. 4.17. Explanation of detection index

The horizontal axis x can be time (a-scan, as in active sonar detection), bearing angle (beam pattern, as in passive sonar detection) or frequency (as in spectrum analysis). The signal is embedded in the fluctuation of background noise. Suppose the average value of noise is

$$D_n = \overline{D}(-\infty, x)$$

where $-\infty$ represents the input SNR $(SNR)_{\text{in}} = -\infty$. If the signal is present, it results in accidental change

$$D_{s+n} - D_n = \overline{D}((SNR)_{\text{in}}, x).$$

The fluctuation of background noise is denoted by σ; the detection index is defined as

$$d = \frac{D_{s+n} - D_n}{\sigma} \tag{4.14}$$

The logarithmic value $DT = 20\lg d$ of detection index d is called the recognition coefficient. The relationship between DT and the detection probability p_d can be calculated based on the ROC curve as described in Chapter 3. For example

Table 4.1

Detection index d	1	2	3
Recognition coefficient DT	0 dB	6 dB	9.5 dB
Detection probability p_d	0.68	0.954	0.997

In sonar design a common value of DT is 6 dB; therefore the detection probability should be 0.954. If in one test the sonar system does not satisfy this index, we can say that the sonar system has failed to pass the test.

4.1.3.4 Accuracy of Parameter Estimation: Origin Moment and Center moment

The distance, incidental angle, and moving speed of the target to be detected are all random variables with some kind of probability distribution. Suppose the true value of the parameter to be estimated is θ, and the estimation value is $\hat{\theta}$, then the estimation error is $\hat{\theta} - \theta$, and the estimation accuracy can be considered as the variance of $\hat{\theta}$. If the estimate is unbiased, i.e., $E(\hat{\theta}) = \theta$, then the estimation accuracy is

$$\text{Var}[\hat{\theta}]^{\frac{1}{2}} = [E(\hat{\theta} - \theta)^2]^{\frac{1}{2}} \tag{4.15}$$

In Chapter 3, we proved that any variance of unbiased estimation will be no better than the Cramer-Rao low bound.

If for some reason, in practical measurement, the estimation error includes systematic error, which should be taken away from the random measurement error. In mathematical statistics, this kind of error concludes to two different types of moments of random variables: origin moment and center moment. We will briefly introduce the concepts of origin and center moments. In sonar signal processing, the first and second moments of a variable are used most often, and the first and

second origin moments are exchangeable with the first and second center moments.

The first center moment is $\mu_1 = E[\hat{\theta} - E(\theta)]$, and the second center moment is $\mu_2 = E[\hat{\theta} - E(\theta)]^2$. The first origin moment is $\nu_1 = E(\hat{\theta})$, and the second origin moment is $\nu_2 = E(\hat{\theta}^2)$.

It is easy to prove that

$$\nu_2 = E(\hat{\theta}^2) = \mu_2 + \nu_1^2 \tag{4.16}$$

Therefore, the first and second origin and center moments can be distinguished from each other. But, the third or higher moments do not have this relationship, except for very specific random variables, and so origin and center moments equal to or above the third order cannot be resolved.

4.1.3.5 Low Probability Events

In discussing the detection problem, we have mentioned that when the detection index is 6 dB, then the detection probability is more than 95%. If in one trial, the sonar system does not reach the technical specification, we can definitely confirm that the sonar system has failed to pass the examination, because the probability that a sonar system does not satisfy the specification is only 5%, and an event with small probability should not occur in one test, which is the rule of low probability events. This rule is extensively used in sonar specification verification.

In the case where the detection probability is not close to 1, one should run the test more than once. The number of tests depends on the probability.

Example

Suppose a sonar system has a target recognition function and under certain conditions, the correct probability for target recognition is 75%. Then we can calculate the probability of multiple tests as follows:

$P\{\text{The correct decision number } \geq 6 \text{ in a 10 times test}\} = 0.931$

$P\{\text{The correct decision number } \geq 7 \text{ in a 10 times test}\} = 0.782$

$P\{\text{The correct decision number } \geq 8 \text{ in a 10 times test}\} = 0.527$

It is reasonable to consider that in the 10 times test if the number of correct decisions is fewer than 6, then the verification is failure to pass.

4.1.3.6 Reliability Concept

The reliability and maintainability specifications are very important in the theory and application of sonar design. Some important concepts about reliability will be discussed here (Fig. 4.18).

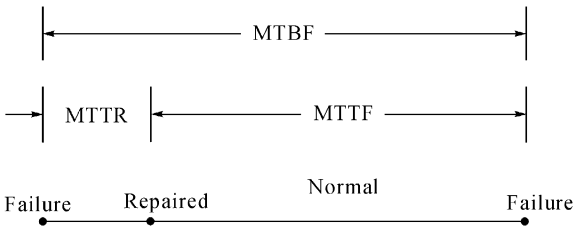


Fig. 4.18. Explanation of the system reliability concept

MTBF (mean time between failures) indicates the average time between two failures; MTTF (mean time to failure) indicates the average normal working time; MTTR (mean time to repair) indicates the average time to repair a system failure. Clearly,

$$MTBF = MTTR + MTTF \tag{4.17}$$

Generally speaking, $MTTR \ll MTTF$, therefore, we can consider that $MTTF \approx MTBF$.

Since a sonar system usually is very complicated, the failure of a part or some subsystems may not affect the normal operation of the whole system. Therefore, it is necessary to identify which kinds of failures will be fatal. There is another definition of reliability, called the mean time between critical failures (MTBCF).

The concept of "critical failure" refers to a serious failure that results in the system being completely stopped. Sometimes the MTBF is called the specification of basic reliability, and MTBCF is called the task reliability.

Corresponding to the concepts of MTBCF and MTBF, there are two quantities called task reliability A_0 and basic reliability A_i

$$A_0 = \frac{MTBCF}{MTBCF + MTTR}, \quad A_i = \frac{MTBF}{MTBF + MTTR} \tag{4.18}$$

Since $MTBCF \geq MTBF$, $A_0 \geq A_i$.

4.1.4 Basic Concepts of Sonar Specifications

It is well known that there are many different kinds of sonar systems, each with its own specification system. We list the most popular specifications in sonar design as following.

(1) Transmitting source level: Expressed in MPa, the transmitting source level of a transducer should be related to the directivity index and frequency.

(2) Source level or spectrum level of radiated noise: Expressed in MPa, the source level or spectrum level of radiated noise should be related to the movement

state of a vessel.

(3) Source level of a platform or source level of ambient noise: The platform noise and ambient noise are background noise. Expressed in MPa, the source level of a platform should be related to the movement state, and the ambient noise should be related to the sea state.

(4) Detecting distance and tracking distance of passive sonar: Tracking distance indicates the range over which sonar steadily tracks a target from nearer to farther. Detecting distance indicates the range over which sonar detects the target from farther to nearer. After detecting the target it may lose contact. Usually, the detecting distance is shorter than the tracking distance. Detecting distance and tracking distance depends on the frequency, ocean environment, radiated noise, etc.

(5) Ranging distance of active sonar: The ranging distance of active sonar is usually the result of a single transmit: when the recognition coefficient DT = 0 dB, the detection probability is 68%. If the number of transmitting signals is three, then the probability of at least one detecting a target is 95%. The specification of ranging distance is related to the source level of the transmitting signal, the frequency, and the strength of the target and ocean environment.

(6) Bearing accuracy of passive sonar: Clearly, the bearing accuracy depends on distance and this specification is often defined in relation to a certain distance, for example, 80% of maximum ranging distance. Since in a sea trial the target is moving, the bearing accuracy of passive sonar should be the result of a statistical average. If the mean value of an estimate is variable, it is inherent that the estimate must be unbiased.

Suppose θ is the incidental angle of a signal and the N measurement values are $x_1, ..., x_N$. The measurement accuracy is defined as

$$\sigma_0^2 = \frac{1}{N-1}\sum_{i=1}^{N}[x_i - \frac{x_1 + ... + x_N}{N}]^2 \tag{4.19}$$

where $1/(N-1)$ is necessary for obtaining unbiased estimation. In practice, the incidental angles to be measured are $\theta_1, \ldots, \theta_N$ and the measurement values are $x_1, \ldots, x_N$. Therefore, the estimation accuracy is defined as

$$\sigma_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \theta_i)^2 \tag{4.20}$$

Because the true values $\theta_1, ..., \theta_N$ come from a different ensemble, σ_1^2 may be different from σ_0^2 unless the estimates are unbiased. The specification of bearing accuracy may vary with bearing angle.

(7) Estimation of error of vessel radial speed in active sonar: The estimation of

radial speed is related to the frequency estimation and, therefore, can be considered as the estimation error of frequency measurement. It depends on the center frequency of the transmitting signal.

(8) Multiple target resolution ability: This is usually defined as the specification of multiple target resolution ability for two equal strength targets. The measurement should be taken at a distance of less than the maximum ranging distance. The specification may be related to the bearing angle.

(9) Identification and classification of target: The target identification and classification is a single choice from multiple possibilities. It is usually defined in DT = 6 dB.

(10) Reliability and maintainability: These specifications are given often in relation to different parts of a sonar system. For example, the dry end and wet end of a sonar system should have different reliability specifications.

The specifications described above are only the main specifications in sonar design and do not cover all the specifications of concern to a sonar designer. Based on the demands of a sonar system, it is necessary to modify some specifications and propose new concepts. For example, in synthetic aperture sonar, we have to consider the specifications of along track resolution and cross track resolution instead of the pure distance resolution.

4.2 Design Procedure of Digital Sonar: the Sonar Equation

The sonar equation is an important concept in sonar design. It reflects the substantial relationship between the media and sonar, sonar systems and detected targets, and the sonar operation state. Once the wet end and signal format are chosen, the main parameters of the sonar equation are confirmed, including the source level, location of the receiver, frequency, length of pulse, and the band width of modulation. The acoustic sound field, the ocean environment, will have a considerable affect on sonar performance. The main parameters of the sound field include wind speed, sound speed profile, temperature, pressure, depth, and the characteristics of the sea floor.

The sonar equation objectively reflects the main parameters in sonar design, and is therefore the key concept in sonar design.

4.2.1 Parameters of the Sonar Equation

There are three types of parameters in the sonar equation:

(1) Parameters defined by the sonar system

SL: Index sound pressure, i.e., receiving sound pressure at 1 m distance from the transmitter.

GS: Space gain of the sonar system. This is obtained from the sonar array. In

the case of the transmitter array, space gain is also denoted by the directional coefficient *DI*.

GT: Time gain of the sonar system. This is obtained from the time integration of signal processing.

DT: Recognition index, also called the threshold of detection. This indicates the minimum output SNR, from which a sonar operator can extract the target.

(2) Parameters defined by the target

SL: Index sound pressure of radiated noise, measured at a 1 m distance from source.

TS: Target strength, indicating the ability of a target to gather sound energy and re-transmit it.

(3) Parameters defined by the ocean environment

NL (or *NL+RL*): Ambient noise (or ambient noise plus reverberation) level. This is sometimes called background noise, and may consist of ambient noise, reverberation or self-noise of a sonar platform.

TL: Transmission loss, indicating the propagation attenuation from 1 m to a certain distance.

All parameters described above are expressed in values of dB, and usually are related to frequency.

4.2.2 Active Sonar Equation

Fig. 4.19 is used to derive the active sonar equation.

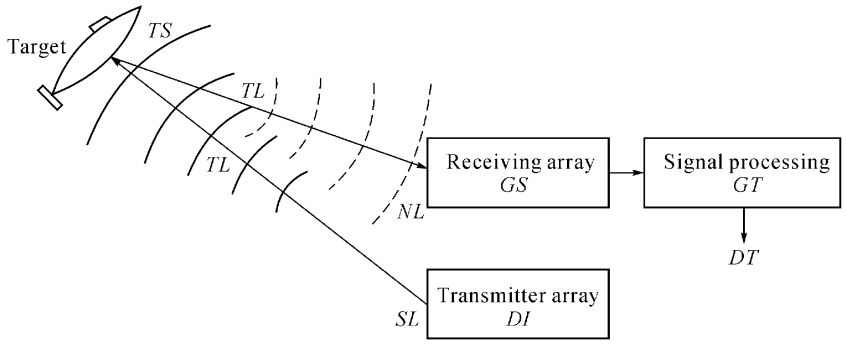


Fig. 4.19. Derivation of active sonar equation

The expression of the active sonar equation is

$$SL - 2TL + TS + GS + GT - NL = DT \tag{4.21}$$

Suppose the power of a transmitting signal is P, then the index sound pressure at 1 m distance from a sound source is

$$SL = 170.8 + \lg P + DI \quad \text{(dB)} \tag{4.22}$$

where DI is the transducer directional coefficient. If the transducer is omni-directional, then $DI = 0$ dB.

Now, let us explain the active sonar equation.

Suppose the index sound pressure is SL and it propagates to a target at distance r. The sound pressure becomes $SL–TL$. Since the target strength is TS, when the signal is re-transmitted (i.e., as the echo) from the target, the sound pressure is $SL–TL+TS$. After the same distance r of propagation, when the echo arrives at the receiving array, the sound pressure becomes $SL–2TL+TS$. Therefore, the signal excess in each hydrophone is $SL–2TL+TS–NL$, where NL can consist of ambient noise or ambient noise plus reverberation (RL). Clearly, the SNR for each hydrophone is

$$(SNR)_{\text{in}} = SL - 2TL + TS - NL \tag{4.23}$$

By adding the space gain and time gain to $(SNR)_{\text{in}}$, we obtain the output SNR of the sonar system

$$(SNR)_{\text{out}} = SL - 2TL + TS - NL + GS + GT \tag{4.24}$$

At the maximum ranging distance, the value expressed in Eq. (4.24) should be equal to the minimum detectable SNR, i.e., the recognition index

$$DT = SL - 2TL + TS - NL + GS + GT$$

This is Eq. (4.20), and is called the active sonar equation.

Generally speaking, the aim of using the sonar equation is to find the ranging distance of a sonar system, i.e., to resolve r from Eq. (4.21). We define a quantity called figure of merit (FOM) as follows

$$FOM = \frac{SL + TS - NL + GS + GT - DT}{2} \tag{4.25}$$

Set

$$FOM = TL(r) \tag{4.26}$$

From this we can find the ranging distance. We will illustrate the solution method later.

As we can see from Eq. (4.21), each term is related to frequency. Therefore, from a theoretical point of view, we can take the derivate for Eq. (4.21) and make it zero to find the optimum frequency. But clearly, this idea is meaningless, since the relationship between each parameter and the frequency is very complicated. No simple method in calculus can be used to find the maximum value of a sonar equation. The optimum frequency for a sonar system actually is the result of a trade-off among many factors.

4.2.3 Passive Sonar Qquation

The expression for a passive sonar equation (Fig. 4.20) is

$$SL - NL - TL + GS + GT = DT \tag{4.27}$$

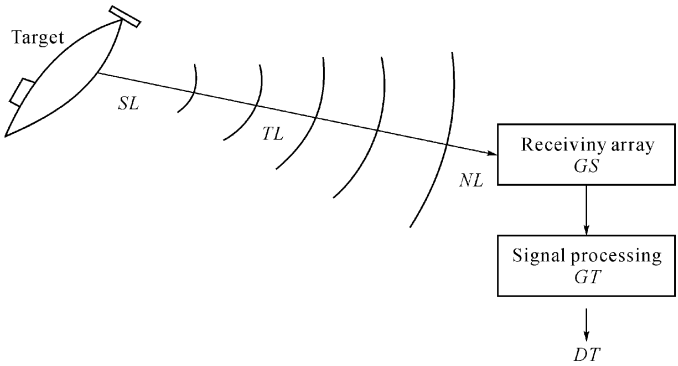


Fig. 4.20. Derivation of passive sonar equation

Suppose the index sound pressure of radiated noise of a target is SL. After propagation over the distance r, at the receiving array the signal becomes $SL - TL$. Therefore, the receiving signal of a single hydrophone is $SL - TL - NL$, where NL consists of ambient noise or ambient noise plus platform noise. The input SNR is

$$(SNR)_{\text{in}} = SL - TL - NL \tag{4.28}$$

Adding the space gain and time gain to $(SNR)_{\text{in}}$, the output SNR should be

$$(SNR)_{\text{out}} = SL - TL - NL + GS + GT \tag{4.29}$$

At the maximum distance of passive sonar, the value on the right side of Eq. (4.29) should be equal to the recognition coefficient:

$$SL - TL - NL + GS + GT = DT$$

This is the passive sonar Eq. (4.27). The FOM for passive sonar is

$$FOM = SL - NL + GS + GT - DT \tag{4.30}$$

In application, the DT in the sonar equation is a quantity that is difficult to define precisely. For vision detection, a $DT \geq 6$ dB is usually chosen; for auditory detection, the DT strongly depends on the operator. A well-trained sonar operator may have a much smaller DT value than a new operator. Therefore, the sonar operator plays a very important role in sonar detection.

4.2.4 Calculation of the Sonar Ranging Distance

The sonar equation can be conveniently used to calculate sonar ranging distance, However, some parameters are related to the ocean environment. Therefore, the same sonar system, i.e., with the same GS, GT, in different seasons or different sea areas may have quite different ranging distances.

In Fig. 4.21, a practical example is presented. A sonar system with $FOM = 90$ dB, in a different place and with a different sound speed profile has a totally different ranging distance.

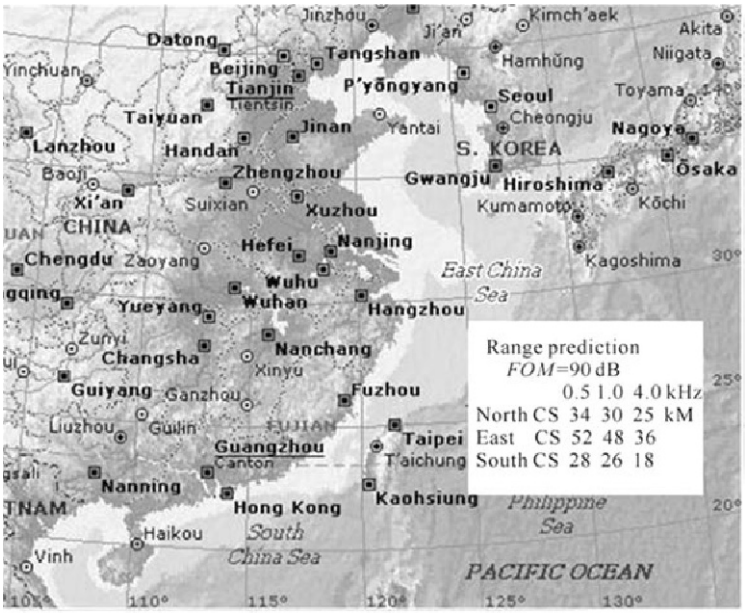


Fig. 4.21. Numerical example of sonar range prediction (FOM = 90 dB) (North China Sea: Isothermal layer; East China Sea: Weak positive gradient; South China: Isothermal layer)

In calculating sonar ranging distance, the source level SL is a parameter that should be precisely defined. For active sonar it is also important to know the target strength. Fig. 4.22 presents some examples of target radiated noise[1]. The noise radiated from a destroyer at a speed of 20 knots and a submarine at an electro-machine speed of 6 knots are illustrated. At 1,000 Hz, the spectrum level of the surface ship noise and the submarine noise is 138 dB and 114 dB, respectively. But at the same frequency, the spectrum level of ambient noise is 73 dB. Based on this figure, we can find the value of $SL-NL$. In calculating the ranging distance there are two different methods that will give the same results.

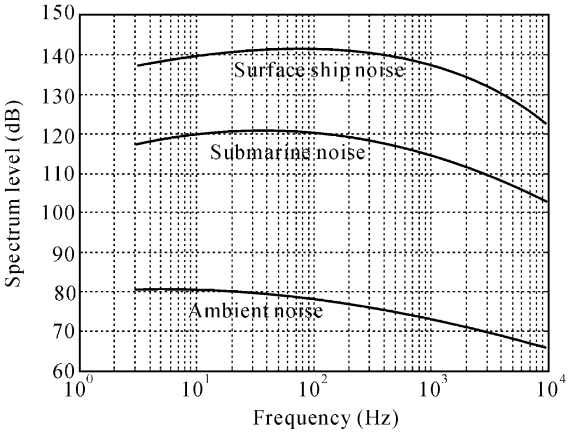


Fig. 4.22. Example of the spectrum of underwater target noise and ambient noise

The first method calculates the ranging distance based on the transmission loss, (Fig. 4.23(a)). Rewriting Eq. (4.27) as $TL = SL - NL + GS + GT - DT$, the distance r which satisfies this equality is the ranging distance we are seeking. The second method calculates the ranging distance based on the FOM (Fig. 4.23(b)).

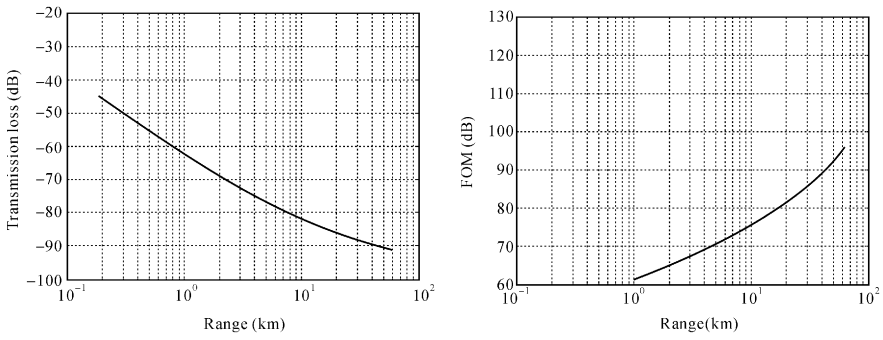


Fig. 4.23. Calculation of sonar range prediction value. (a) Range prediction by calculation of transmission loss; (b) Range prediction by calculation of FOM

In Fig. 4.23(b), we can see that if the $FOM = 84$ dB, then the ranging distance is 20 km. Fig. 4.24 shows a typical plot of sonar ranging distance prediction. Curves A, B, and C present different types of FOM vs. r. This is of course related to the frequency.

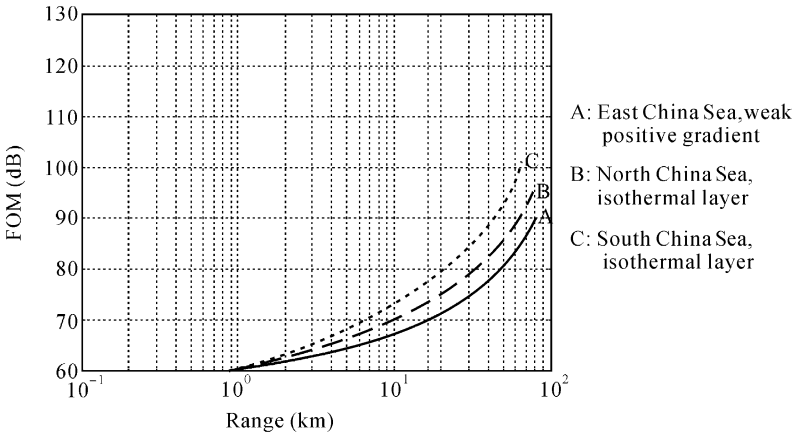


Fig. 4.24. Sonar ranging distance in different ocean environments in shallow water

Example (calculation of sonar ranging distance)

(1) Active sonar

Suppose a surface active sonar system operates at a frequency of 4 kHz, the sound power of the transmitter transducer P =1 kW, and the directional coefficient $DI = 15$ dB. Based on Eq. (4.22), we have $SL = 170.8 + 10\lg P + 15 = 215$ (dB). If the space gain of the receiver is GS = 12 dB, the time gain is GT = 8 dB. The target strength TS = 15 dB and the recognition coefficient is DT = 6 dB. The platform noise is much greater than the ambient noise, and NL = 78 dB. Therefore

$$FOM = (SL + TS - NL + GS + GT - DT)/2 = 83 \text{ (dB)}.$$

(2) Passive sonar

Suppose a passive sonar has a $FOM = 80$ dB and the working frequency band is 2.5 – 6.3 kHz. We are going to estimate the ranging distance of this sonar for a submarine. The index spectrum level of this submarine is as follows

Table 4.1 Calculatio

No.	Frequency (kHz)	Spectrum level (dB)
1	2.50	118
2	3.15	115
3	4.00	114
4	5.00	110
5	6.13	106

Since the transmission loss depends on the frequency, and the passive sonar operates in broadband, it is necessary to derive the transmission loss curve for specific broadband signal and then calculate the ranging distance of passive sonar. The calculations are listed in Table 4.1.

Table 4.1 Calculation of ranging distance of passive sonar

Range (m)	SL (dB)	f_i (kHz)	2.5	3.15	4	5	6.3	Source level 2.5 – 6.3 (kHz)
		Δf_i (kHz)	0.58	0.72	0.92	1.15	1.45	
1	$SL\,(r)$ (dB)		118	115	114	110	106	126
	$1g^{-1}[SL(r)/20]$		794,328	562,341	501,187	316,227	199,256	1,979,662
100	$SL\,(r)$ (dB)		78	75	74	70	66	86
	$1g^{-1}[SL(r)/20]$		7,943	5,623	5,012	3,162	1,995	19,796
1000	$SL\,(r)$ (dB)		66	63	60	56	51	73
	$1g^{-1}[SL(r)/20]$		1,995	1,412	1,000	631	355	4,334
3000	$SL\,(r)$ (dB)		59	58	51	46	41	65
	$1g^{-1}[SL(r)/20]$		891	794	354	199	112	1,805
5000	$SL\,(r)$ (dB)		53	51	44	39	34	58
	$1g^{-1}[SL(r)/20]$		447	354	158	89	50	834
7000	$SL\,(r)$ (dB)		49	47	39	34	29	54
	$1g^{-1}[SL(r)/20]$		281	223	89	50	28	504
15000	$SL\,(r)$ (dB)		39	36	28	23	18	43
	$1g^{-1}[SL(r)/20]$		89	64	25	14	8	148

In Table 4.1, the transmission loss value is theoretical, and the bandwidth Δf_i is calculated in terms of 1/3 octave. For FOM = 80 dB, the source level decreases from 126 dB to 46 dB and the distance is about 10 km.

4.3 Ambient Noise

As we can see from the sonar equation, the performance of a sonar system is limited by the background noise. The background noise can be categorized into three classes: ambient noise, reverberation and self-noise from the platform. From the point of view of anti-interferences, it is important to make good use of the characteristic differences between signal field and noise field. Therefore, it is very important to know the statistical performance of noise field. In this section, we will introduce the main source and performance of ocean environment noise or ambient noise. Early studies of ambient noise can be found in the literature[14-21].

4.3.1 Main Source of Ambient Noise in the Ocean

The sources of ambient noise are various, and usually include pressure waves resulting from tide and sea waves and pressure impulses resulting from turbulence.

They also include seismic activity, wind, rain, molecular thermal motion and fish school activities. In the neighborhood of a port or bay, industrial noise and vessel activity are also important sources of ambient noise.

The period of pressure change resulting from tide is about 1 – 2 days. Although the pressure is very large, the frequency is much smaller than those of interest in ocean engineering. The frequency resulting from an earthquake is also smaller than 10 Hz. Many observations have proved that distant vessels are the main source of ambient noise of less than 100 Hz. A storm at a distance is also a source of ambient noise.

In nearby sea, wind is the main source of ambient noise. The noise resulting from rain is not negligible. The shore beat of waves and the friction resulting from tide with sand on the beach is also a part of ambient noise.

Ambient noise in the ocean consists of massive irregular factors. Based on the central limit theorem in probability theory, we can consider that the ambient noise, as a random signal, obeys the Gaussian distribution. The frequency range of ambient noise is listed below:

(1) Earthquake and sea bed motion: 1 – 10 Hz
(2) Ship moving and ocean engineering: 10 – 10,000 Hz
(3) Noise of sea creatures: 10 – 1,000,000 Hz
(4) Molecular thermal motion: 10,000 – 1,000,000 Hz

4.3.2 *Frequency spectrum of ambient noise*

Fig. 4.25 shows the typical frequency spectrum of ambient noise. In semi-log plot, the spectrum can be divided into several segments, each segment being a straight line. In the 10 – 100 Hz region, the spectrum value is constant; in the 100 – 10,000 Hz region, the spectrum is a cline straight line. The attenuation of each octave is about 6 – 8 dB. With increasing sea state, the spectrum value will increase by about 3 – 4 dB for one level of increase. For sea state 3, the noise level at 1,000 Hz is 73 dB (ref. 0 dB = 1 μPa).

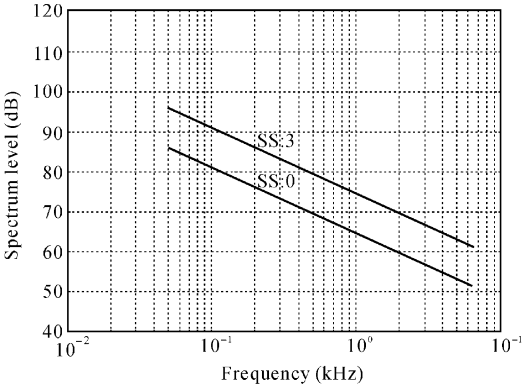


Fig. 4.25. Typical spectrum of ambient noise

Example

Sometimes it is necessary to calculate the noise power in a certain frequency band. For example, if the average power spectrum $K_n(f)$ of ambient noise can be expressed as

$$K_n(f) = K_n(f_0)[f / f_0]^{k/3} \tag{4.31}$$

where f_0 is a reference frequency, exponent k expresses the attenuation value of every octave. If $k = 6$, it indicates that $K_n(f)$ decreases by 6 dB in one octave. Taking the logarithm for Eq. (4.30), it follows that

$$\begin{aligned} N(f) &= 10 \lg K_n(f) \\ &= 10 \lg K_n(f_0) + [10k \lg(f / f_0)] / 3 \\ &= N(f_0) + [10k \lg(f / f_0)] / 3 \end{aligned} \tag{4.32}$$

In frequency band (f_1, f_2), the noise power is

$$\begin{aligned} J &= \int_{f_1}^{f_2} K_n(f) \mathrm{d}f \\ P &= 10 \lg J = N(f_0) + 10 \lg \left\{ \frac{f_0}{1 - k/3} \left[(f_2 / f_0)^{-k/3+1} - (f_1 / f_0)^{-k/3+1} \right] \right\} \quad (k > 3) \end{aligned} \tag{4.33}$$

4.3.3 Minimum Ambient Noise

It is very important to know the minimum level of ambient noise because this is the basis for determining the sensitivity requirement of the hydrophone. Clearly, the sonar hydrophone must normally work in an ocean environment with minimum ambient noise level, i.e., it must have sufficient voltage output at the minimum noise level.

The zero sea state can be considered as the minimum ambient noise level.

Example

Assume the sonar system operates in a frequency band of 500 – 5,000 Hz. From Fig. 4.25, we find that, in zero sea state, NL = 66 dB at 1,000 Hz. The power of ambient noise in this band can be calculated based on Eq. (4.32)

$$P = 10 \lg K_n(f_0) + 10 \lg f_0 - 10 \lg(k / 3 - 1) + 10 \lg[(f_1 / f_0)^{1-k/3} - (f_2 / f_0)^{1-k/3}]$$

Note when $f_1 = 500, f_2 = 5{,}000, f_0 = 1{,}000$, and $k = 6$, we have

$$P = 66 + 30 + 10\lg 1.8 = 98.5 \quad \text{(dB)}$$

This is equal to 8.4×10^4 μPa.

Suppose a hydrophone has a sensitivity of –195 dB (ref. 1V / μPa). Then the voltage sensitivity is $M = 1.9\times10^{-4}$ μV / μPa. We can find the voltage output of this hydrophone in receiving sound pressure waves of 8.4×10^4 μPa: $U = 8.4\times10^4\times1.9\times10^{-4}$ μV = 15.9 μPa. This value can be considered as the minimum input voltage of a pre-amplifier.

4.3.4 Homogeneous and Isotropic Noise Fields

The sonar array operates in sea water and regardless of whether the array is fixed or movable, background noise always exists. The space / time characteristics of ambient noise fields directly affect the detection performance of a sonar system. In sonar design, it is necessary to know the statistical behavior of the ambient noise field. Therefore, the design of a sonar array and other related parameters must be based on a reliable theoretical basis.

To discuss the space / time characteristics of a noise field, it is necessary to establish a suitable model for the ocean environment, and then compare the theoretical results with measured data.

To consider ambient noise as Gaussian random signal is the simplest model. Denote the signal by $p(x, y, z; t)$, where (x, y, z) is the location in the noise field and t is the time variable. Assume $p(x, y, z; t)$ is a stationary, ergodic random signal. If (x, y, z) is fixed, then we can calculate the time correlation function

$$R(x, y, z; \tau) = E[p(x, y, z; t)p(x, y, z; t-\tau)] \tag{4.34}$$

When t is fixed, we can calculate the spatial correlation between two points

$$R(x_1, y_1, z_1; x_2, y_2, z_2; t) = E[p(x_1, y_1, z_1; t)\, p(x_2, y_2, z_2; t)] \tag{4.35}$$

From Eqs. (4.34) and (4.35) we can see that the time correlation is the auto-correlation of the signal received in location (x, y, z) and the spatial correlation is the cross-correlation of the signal between two locations (x_1, y_1, z_1) and (x_2, y_2, z_2). If x, y, z, and t are all changeable, then we can calculate the time-space correlation function. Although the time space correlation function is only the second order moment of ambient noise, it plays a very important role in the study of noise fields.

Several useful models have been established for ambient noise fields. The most important model is the homogeneous and isotropic noise model. Suppose in the considered noise field there are no other interferences, then this field, in a local

region, can be considered as homogeneous and isotropic.

"Isotropic" means that the value of the sound pressure $p(x, y, z; t)$ depends only on $r = (x^2+y^2+z^2)^{1/2}$, and "homogeneous" means that $p(x, y, z; t)$ is stationary in space; i.e., the correlation function of $p(x_1, y_1, z_1; t)$ and $p(x_2, y_2, z_2; t)$ depends only on the difference between $r_1 = (x_1^2+y_1^2+z_1^2)^{1/2}$ and $r_1 = (x_2^2+y_2^2+z_2^2)^{1/2}$.

Eckart[15] showed that, in this noise model, the time spatial correlation function can be expressed by the time correlation.

The sound pressure is solved from Eq. (4.36)

$$\nabla^2 p(x, y, z; t) - \frac{1}{c^2}\frac{\partial^2 p(x, y, z; t)}{\partial t^2} = -4\pi\psi(x, y, z; t) \tag{4.36}$$

where $\psi(x, y, z; t)$ indicates the noise source at location (x, y, z) and c represents the sound velocity. It is proved that the time-space correlation function $R(r, \tau)$ can be expressed by Eq. (4.37)

$$R(r, \tau) = \frac{c}{2r}\int_{\tau-r/c}^{\tau+r/c} R(0, t)\mathrm{d}t \tag{4.37}$$

Therefore, we conclude that the time-space correlation can be expressed by the integral of time correlation for a homogeneous and isotropic noise field. This is a very important result in application as we need measure only the time correlation function to calculate the time-space correlation function by Eq. (4.37), and we need only one hydrophone to measure the time correlation function. In contrast, if the time-space correlation function of a noise field does not satisfy Eq. (4.37), then we can be sure that this field is not homogeneous and isotropic.

Example

Assume the auto-correlation function of a homogeneous and isotropic noise field is

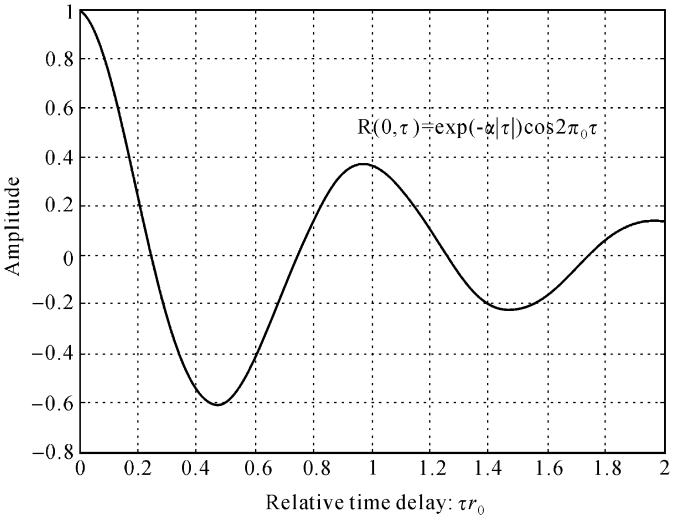
$$R(\tau) = \exp(-\alpha \mid \tau \mid)\cos(2\pi f_0\tau) \quad (\alpha > 0) \tag{4.38}$$

Substituting Eq. (4.38) into Eq. (4.37), we obtain

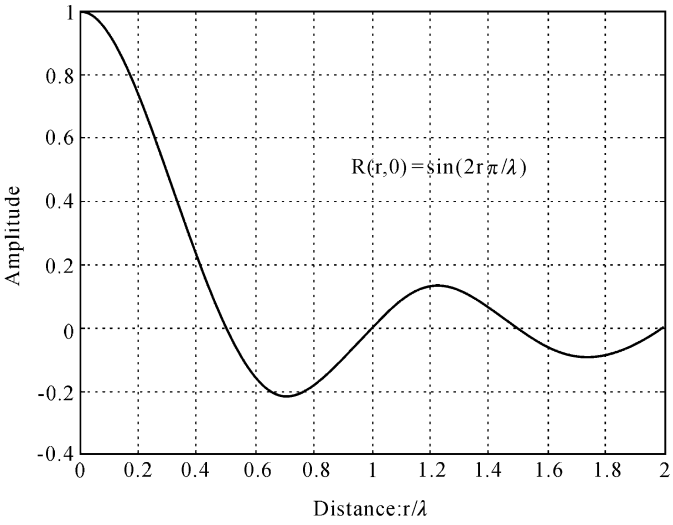
$$R(r, \tau) \approx \frac{\sin(2\pi f_0 r / c)}{2\pi f_0 r / c} \tag{4.39}$$

As we can see from Eq. (4.38), the first zero point of the time correlation function is $\tau_0 = 1/4f_0$, and the first zero point of the spatial correlation is $r_0 = \lambda/2$.

The curves of $R(\tau)$ and $R(r, 0)$ are plotted in Fig. 4.26.



(a)



(b)

Fig. 4.26. Time-space correlation function of ambient noise. (a) Time correlation function; (b) Time-space correlation function

In Eq. (4.37), if we assume $\tau = 0$, we can obtain the expression for calculating the spatial correlation function in terms of the time correlation

$$R(r,\ 0) = \frac{c}{r}\int_0^{r/c} R(0,\ \tau)\mathrm{d}\tau \tag{4.40}$$

4.3.5 Cylindrical and Spherical Model of Ambient Noise

The distribution of the spectrum and the strength of ambient noise strongly depend on the measurement position. Values from near the sea surface and from deep sea are quite different. The ambient noise of the ocean environment also depends on the depth of the sea, as shallow water and deep water will have totally different noise levels and spectrum distributions. Noise models have been studied by Cron et al.[16-22]. The main theoretical models they proposed are the cylindrical model and the spherical model, also called the surface model and the volume model, respectively.

In the spherical model, the sound pressure is expressed as

$$p(x,\ y,\ z;\ t) = \frac{A}{\mathrm{j}kr}\exp[\mathrm{j}(kr - 2\pi ft + \varphi)] \tag{4.41}$$

where $r = [x^2 + y^2 + z^2]^{1/2}$, $k = 2\pi / \lambda$, $f = c / \lambda$, and A and φ are constant. Based on Eq. (4.41), we can find the strength distribution per unit spherical area, and then take integration in spherical to obtain the spatial correlation function

$$\psi(x) = R(x,\ 0) = \sigma^2 \frac{\sin kx}{kx} \quad (k = 2\pi / \lambda) \tag{4.42}$$

It is also proved that the time-space correlation function $R(d,\ \tau)$ can be calculated from the spatial correlation function $R(d,\ 0)$

$$R(d,\ \tau) = R(d,\ 0)\cos\ (2\pi f\tau) \tag{4.43}$$

In the cylindrical model, the sound pressure is expressed by

$$p(x,\ y,\ z;\ t) = \frac{Ag(\alpha)}{\mathrm{j}kr}\exp[\mathrm{j}(kr - 2\pi ft + \varphi)] \tag{4.44}$$

where $g(\alpha)$ is called the effect factor of surface noise. The perpendicular line to the sea surface has $\alpha = 0$. Usually, we have $g(\alpha) = \cos^m(\alpha)$, Where, m is a non-negative integer. Based on Eq. (4.44), we can find the strength distribution per unit surface area, and then take integration in cylindrical surface to obtain the spatial correlation function

$$\psi(x) = R(x,\ 0) = \sigma^2 J_0(kx) \tag{4.45}$$

where $J_0(\cdot)$ is the zero order Bassel function.

The correlation functions of the cylindrical model and the spherical model are plotted in Fig. 4.27. As we can see, there are not too many differences. The positions of the first zero points are both at about 0.5λ.

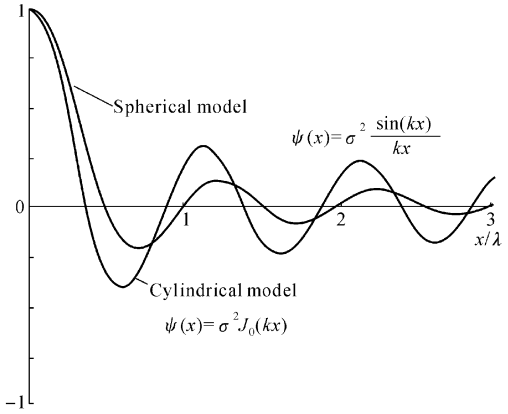


Fig. 4.27. Correlation function of cylindrical and spherical noise models

4.3.6 Vertical Directivity of Ambient Noise

Field measurement shows that ambient noise has vertical direction, especially near the sea surface. The early research work was reported by Knudson in 1948 and, therefore, sometimes the vertical directivity curve of ambient noise is called the Knudson curve[22-24]. Fig. 4.28 shows a typical deep sea Knudson curve for low frequency. In the curve, from –15° to 15° of horizontal direction, ambient noise has obvious directivity, as it is higher than in the vertical direction by about 8 – 10 dB.

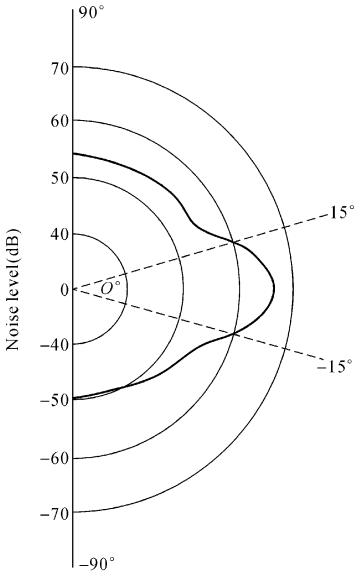


Fig. 4.28. Typical vertical distribution of deep sea ambient noise

4.4 Radiated Noise from Underwater Target and Platform Noise

The noise radiated from a surface ship, submarine or torpedo is the signal to be detected by a sonar system. However, for sonar installed in a surface ship, submarine or torpedo, self-radiated noise is undesired interference. This radiated noise is of concern to the sonar designer.

The signal radiated from a target and the interference radiated from a platform have quite different behavior. The target signal, in most cases is far from the receiver and, therefore, can be considered as a point source and has a plane wave form. However, platform interference cannot be considered as a point source because the receiver is near the source. Thus, although they are both radiated noise, each has different correlation characteristics owing to the distance of the receiver.

4.4.1 Sources of Radiated Noise

The sources of surface ship, submarine and torpedo noise can be categorized into three types[24-27].

(1) Mechanical noise: Mechanical vibration of the main engine, auxiliary machine and various parts of air equipment are the sources of mechanical noise. The vibration is radiated into the sea water via various racks and pipes, and via the hulls of vessels.
(2) Propeller noise: The blade of a propeller periodically beats sea water and the vibration of the propeller shaft results in noise, including cavitations and turbulence.
(3) Water dynamic noise: For a moving target, the friction of a vessel's hull with sea water will result in noise of a certain frequency. Movement of a vessel will yield the resonance of its hull, in which various types of accessories are installed.

In most situations, the first two sources are the main factors. The generation of cavitation noise from a propeller is randomly broken into massive bubbles, and therefore the frequency spectrum of cavitations has a continuous shape. The results of a large amount of data statistics show that in the high frequency band, the spectrum decreases by 6 – 8 dB each octave. In the low frequency band, the spectrum level shows a slight increase with frequency and therefore appears as a peak value in the whole frequency curve. For the spectrum of a surface ship or submarine, the peak value is at a certain frequency between 100 – 1,000 Hz.

Ross[25] showed the cavitation index of a submarine is

$$K \approx \frac{2}{3} \times \frac{9+h}{(N/100)^2 D^2} \tag{4.46}$$

where h represents the depth of the submarine in meters, D is the diameter of the propeller in meters, and N is the rotation speed in run / min. For $K \geq 0.5$, cavitation

is impossible, while for $0.2 \leq K \leq 0.5$, it will be moderate and for $K \leq 0.2$ it will be serious.

Apart from the continuous spectra generated by cavitations, there are also some single frequency components resulting from the water flowing from the propeller. One component is the high frequency resonance of the blade. Another is the "blade speed spectrum" with relatively low frequency, sometimes called "tone frequency" or "singing voice". It is generated by the blade of the shaft cutting the sea water that flows into the vicinity of the propeller. The frequency value is

$$f_m = n \times m \times s \tag{4.47}$$

where f_m represents the m-th resonant frequency (in Hz), n is the number of blades and s is the rotation speed of the propeller in r / s.

In Eq. (4.47), if we set $n = 1$ then it becomes the expression for calculating the axis frequency, which is the component resulting from the rotation of the axis of the propeller.

Example

Assume a propeller of a submarine has 6 blades. When it is working at 75 run / min, the axis frequency and tone frequency generated by the propeller are as follows: Axis frequency, 1.25 Hz, 2.50 Hz, 3.75 Hz, …; Tone frequency, 7.5 Hz, 15 Hz, 22.5 Hz, ….

In sea water, the radiated noise of the target is added to the background noise, and the receiving signal of a hydrophone includes the background noise and target signal, if present. Generally, if the distance of a receiver is 1,000 m from the target, then the SNR will most likely be equal to 0 dB. Clearly, if we want to study and record the "pure" radiated noise of the target, we have to put the receiver as close as possible to the target. For example, at 100 m, the recorded signal will show the characteristics of radiated noise. But note that it is impossible to obtain "pure" signal of target radiated noise because of the existence of background noise.

Fig. 4.29 shows a typical power spectrum figure of a vessel. The target noise has plentiful continuous spectrum in the frequency band from 10 – 10,000 Hz. The single frequency signal appears in the frequency band of 10 – 1,000 Hz, and the value of line spectrum is higher than that of continuous spectrum by about 5 – 7 dB. The slope of continuous spectrum in the frequency band 10 – 1,000 Hz is 3 dB per octave and in the frequency band 1,000 – 10,000 Hz is 6 dB per octave. Fig. 4.29 also indicates the suitable operation band of the sonar system, i.e., towed array sonar: 10 – 1,500 Hz; flank array sonar: 500 – 2,000 Hz; passive raging sonar: 2,000 – 10,000 Hz.

The abbreviations LOFAR (low frequency analysis record) and DEMON (demodulation on noise) indicated in this figure, are useful specific detection methods in digital sonar, which we will discuss later.

Of course, detailed information about noise radiated from military targets, for any country, is top secret. The unclassified data reported in the literature include data from World War Ⅱ (WWⅡ). Most of these vessels have been retired, but at least these data can be considered as most useful references.

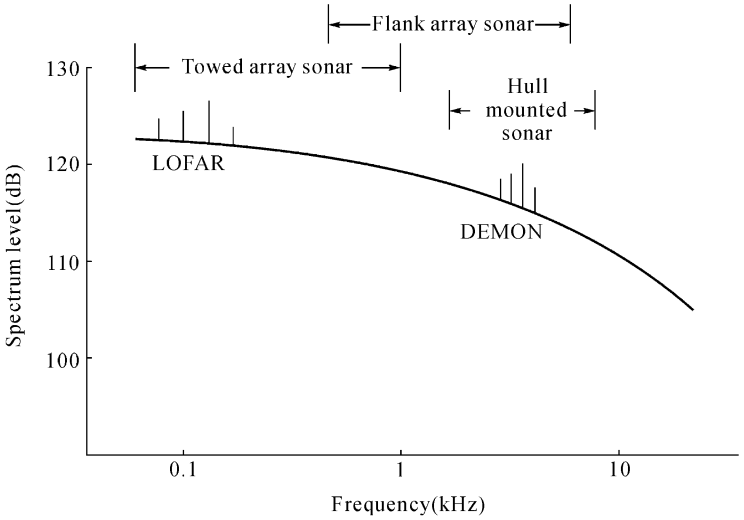


Fig. 4.29. Division of frequency bands in sonar detection

4.4.2 Radiated Noise from Submarines

There are many and various types of machines in a modern submarine. Apart from the main engine, the auxiliary equipment includes more than 50 machines, most of which are sources of noise.

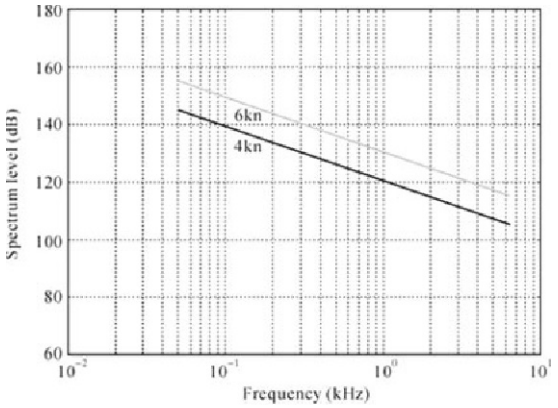
When the propeller rotates at a high speed, the sea water can not reach the rear part of the propeller and therefore cavitations are generated. This is the main source of submarine noise. The total source level from 10 – 10,000 Hz is about 170 – 200 dB.

With the old submarines of WW Ⅱ, the source level at a speed of 6 kn, under water and 12 kn on the sea surface was about 198 dB. When the submarine runs at the periscope depth, there is a critical speed. Below the critical speed, for example 4 kn, usually the noise is relatively low, but over the critical speed, the source level increases rapidly with the speed and then at some speed value, for example 8 kn, the rate of change reduces.

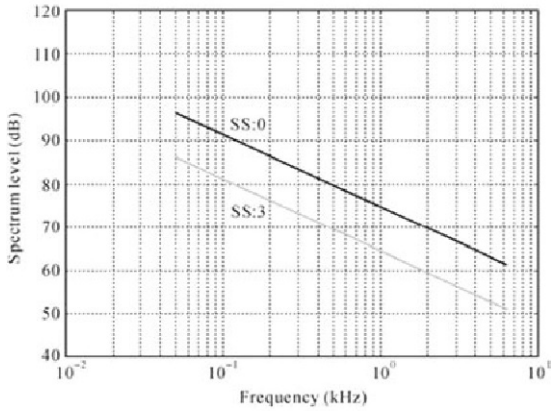
Field tests show that submarine noise has directivity; between the stern and bow direction (–20° – 20°) the measured source level will differ by about 3 dB.

Fig. 4.30 illustrates the spectrum of a US submarine, named Hake Runner[1]. For comparison, the ambient noise is also plotted.

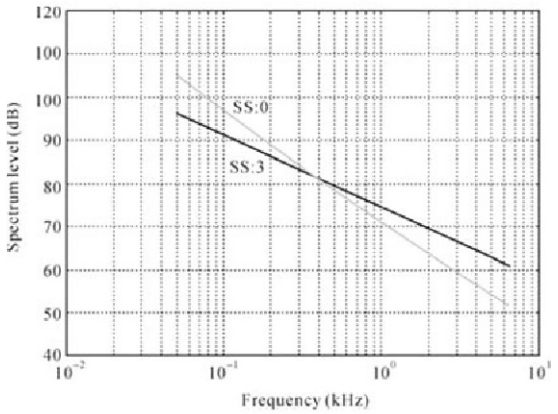
In Fig. 4.30, we can also see that the spectra of target noise and ambient noise are almost parallel. Therefore, the difference in source level is almost the same within a certain bandwidth. For example, at 1 m, the SNR is $(SNR)_{in}$ = 128–72 = 56 dB, while at 100 m, the SNR will become $(SNR)_{in}$ = 16dB owing to the spherical attenuation at 100 m.



(a)



(b)



(c)

Fig. 4.30. Some examples of submarine noise, ambient noise and tow cake noise. (a) Example of spectrum of submarine noise; (b) Example of spectrum of ambient noise; (c) Example of spectrum of tow cable noise

As the radiated noise of a submarine is the detection object of passive sonar, from a submarine design point of view, reducing the radiated noise is a very important task. The main approaches studied are the development of large push force, low rotation speed, low noise propellers and the isolation of vibration inside the submarine engine and auxiliary equipment. It is reported that the spectrum frequency of modern nuclear submarines has the form: $SL(f) = 129 - 17 \lg f$, f is in Hz. It is, of course, an approximate expression.

According to this expression, at 1,000 Hz, the spectrum level of a nuclear submarine will be SL(1,000 Hz) = 78 dB, about the same level as sea state 3 ambient noise.

In the last 30 years, the radiated noise of submarines has reduced by an average of 1 dB per year, and this trend is continuing. This is a very serious challenge for the passive sonar designer.

Miasnikov showed that the source level of modern nuclear submarine noise can be categorized into three classes: noisy, quiet and very quiet[28]. The reported spectrum level of a ballistic nuclear strategy submarine (BNSS) at 30 Hz is as follows: Noisy: SPL(30 Hz) = 140 Db; Quiet: SPL(30 Hz) = 120 dB; Very quiet: SPL(30 Hz) = 100 dB.

4.4.3 Radiated Noise of Surface Ships

The sources of radiated noise from surface ships are basically the same as those from submarines. However, since a surface ship has much less draught compared to a submarine, the mechanical vibration inside the hull of a vessel has less opportunity to propagate into the sea water. Also, in a surface ship the radiated noise generated by the propeller is proportionally larger than that from a submarine. Table 4.2 lists some measurement results obtained by a US navy surface ship using the TAG-6 system[26].

The reference sound pressure value of the original data was 1 bar at 18 m; this has been changed to 1 μPa and 1 m, respectively.

Table 4.2 Examples of frequency spectra of surface ships (ref. 1 μPa)

Frequency (Hz)	Cargo, (10 kn)	Passenger ship (15 kn)	Battle ship (20 kn)	Cruiser (20 kn)	Destroyer (20 kn)	Light cruiser (20 kn)
100	151	161	175	168	162	156
300	141	151	165	158	152	146
1000	130	140	154	147	141	135
3000	120	130	144	137	131	125
5000	116	130	140	133	127	121
10000	110	116	134	127	121	117
25000	102	112	126	119	113	107

4.4.4 Radiated Noise of Torpedo

The mechanism of noise radiation from a torpedo is similar to that from a submarine. But since the architecture of a torpedo is much simpler than that of a submarine, and because the rotation speed of its propeller is much faster than that of a submarine, the construction of torpedo noise is relatively simple. Because the torpedo is the most threatening weapon under the sea, the technique of torpedo alarming is of particular concern in the field of sonar.

The radiated noise of a torpedo has obvious directivity. If the radial direction from stern to head is taken as 0°, then the direction of the maximum value is at ±90°. The source level at 0° may be lower by 10 dB than that at ±90°. This characteristic is useful in the passive detection of torpedo noise. Fig. 4.31 shows examples of torpedo spectra, which have been smoothed[1].

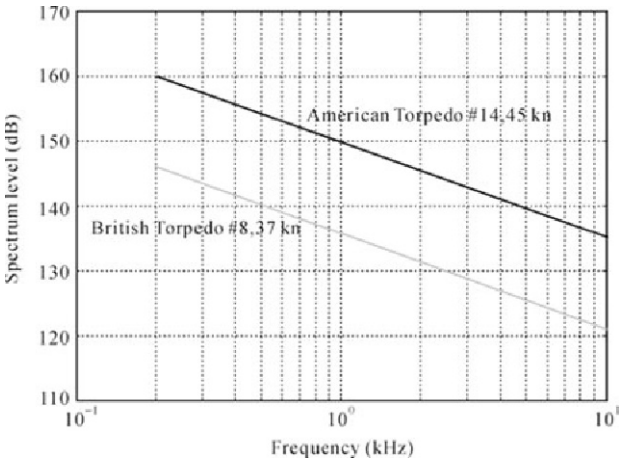


Fig. 4.31. Examples of torpedo spectra

Since the rotation speed of a torpedo is much higher than that of a submarine, the continuous part of the spectrum is shifted to the high end of the frequency band. The tone components, if present, are also moved to the high end.

4.4.5 Self-noise of Vessels

In the sonar equation, the radiated noise appears in the form of the source level *SL*, because it is considered as the target to be detected, but the target noise itself also contributes to the interference of its own sonar system. Therefore, it also may be considered as background noise *NL* in the sonar equation.

In terms of the different types of modern sonar installed in vessels, the platform noise has the following different forms.

4.4.5.1 Platform Noise Received by Hull-mounted Sonar

The hydrophones installed in hull-mounted sonar receive the platform noise, including the propeller and the hull vibration noise. Generally, the further the hydrophones are installed from the propeller, the lower is the platform noise. Fig. 4.32 shows an example measurement result of the self-noise of a US submarine. The curve shows the relationship between self-noise and the submarine speed at 1,000 Hz. As shown, the speed of 15 kn is a critical value. Below this critical value, the self-noise increases slowly with speed, but above this value, the self-noise increases quite quickly with speed. A reasonable explanation is that when the submarine speed goes above 15 kn, the machine inside the submarine produces resonance such that excessive noise is generated.

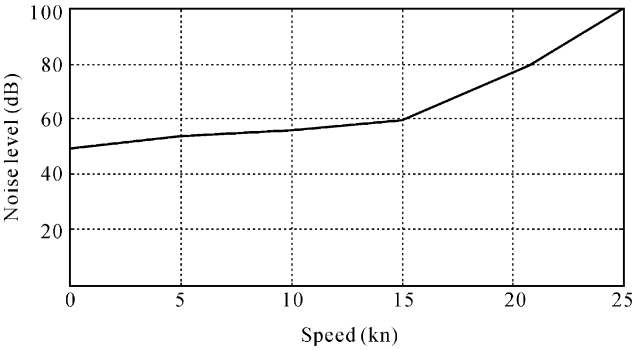


Fig. 4.32. Self-noise of a submarine (1,000 Hz)

But the platform of a surface ship does not show this phenomenon. Fig. 4.33 shows an example of a measurement result of a destroyer. From this figure, we can see the decibel value of background noise has a linear relationship with vessel speed.

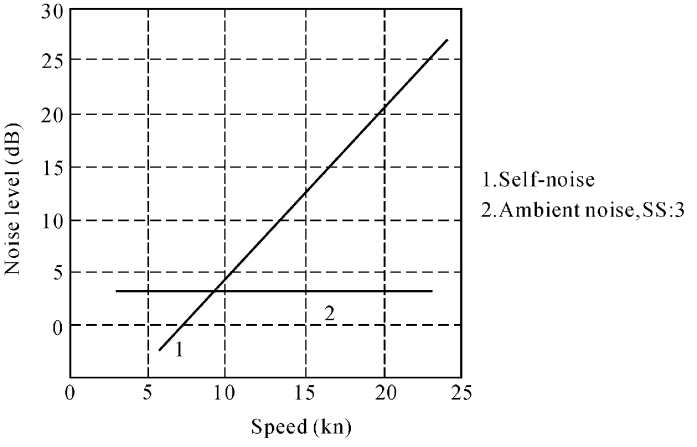


Fig. 4. 33. Self-noise of a destroyer (0.1 – 10 kHz)

Fig. 4.34 illustrates the frequency spectrum of background noise. Fig. 4.35 shows the distribution of submarine noise along the hull body. There seems to be a critical position that defines the point beyond which self-noise changes fast.

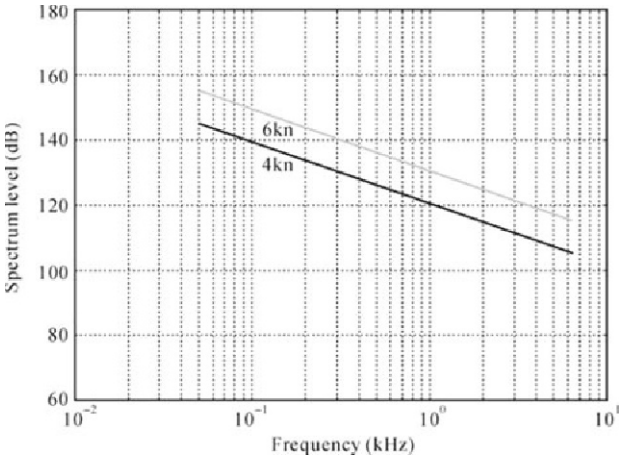


Fig. 4.34. An example of spectrum of self-noise from the bow of a US submarine

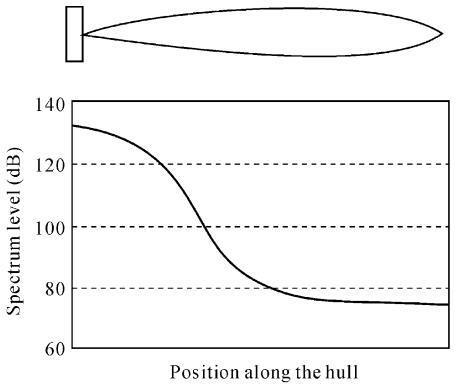


Fig. 4.35. The distribution of self-noise along the hull of a submarine (4 kHz, ref. 1 μPa)

4.4.5.2 Platform Noise Received by a Towed Array or Towed Body

If the received array is installed in a towed body, such as variable depth sonar, or in a towed array, such as towed array sonar, then the platform noise can be considered to be the same as the target noise. This is because the towed body or towed array is usually far away from the propeller. However, the total noise received by the hydrophone array also includes flow noise from a towed body or towed array. Sometimes it is difficult to distinguish between flow noise and platform noise. Fig. 4.36 illustrates some examples of measurement results of

self-noise from towed array sonar. The measurement results are in the low frequency range, which is suitable for an underwater seismic cable operation, and the diameter of the received array is $\varphi \geq 6$ cm.

4.4.5.3 Flow Noise of Acoustic Module of Towed Array

The hydrophones of towed array sonar are installed in the cable, in what is called the acoustic module. It is the most important part of the wet end of towed array sonar. The study of the flow noise of an acoustic module is an interesting topic in towed array sonar design. Flow noise is a special kind of hydraulic dynamic noise. Around the hydrophone, the hydraulic turbulences in the skin of a towed body or towed cable generate vibration, resulting in a change in sound pressure. Based on practical measurement, the spectrum level $NPL(f)$ of flow noise is proportional to f^3. This means that the spectrum level of flow noise decreases by 9 dB / Oct. For a cable with a smooth surface, the flow noise is inversely proportional to the diameter of the cable for a fixed frequency. The theoretical and experimental results show that if the diameter of the cable is doubled, then the flow noise will decrease by 5 – 10 dB. That means that the flow noise of a thin cable is serious. Flow noise appears mainly at frequencies lower than 150 Hz (Fig. 4.36)[1, 27]. If the sonar operation frequency can avoid this frequency band, then the detection performance of towed array sonar should not be degraded.

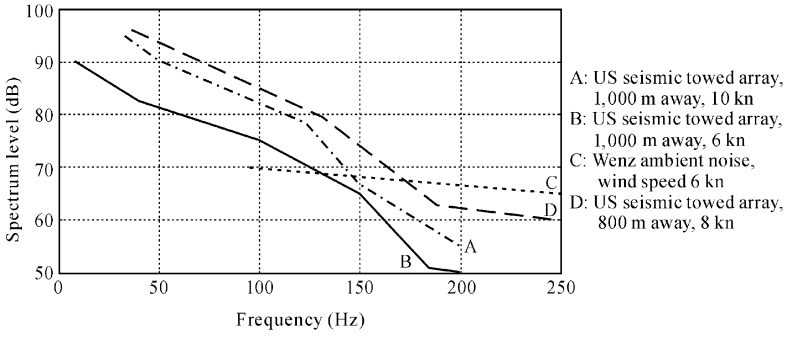


Fig. 4.36. Towed array noise

4.4.6 Auto-correlation Function of Target Noise

As the detected signal, the target noise used to be filtered via a band pass filter. The behavior of second order statistic moment is very important in sonar signal processing. The auto-correlation function is used to calculate the system gain in the sonar equation. A specific expression for the auto-correlation function of the band-limited random signal is derived here. It will be very useful in some cases,

including active sonar signal processing and narrow-band passive sonar signal processing.

Assume the power spectrum of target noise $n(t)$ is a straight line in a semi-log figure, i.e.

$$K_n(f)=\begin{cases} A\mid f\mid^{-\alpha} & f_1<\mid f\mid<f_2 \\ 0 & \text{otherwise} \end{cases} \tag{4.48}$$

where the exponent α is positive.

The auto-correlation function can then be calculated using the Wiener-Kshinchine theorem (see Eq. (2.64))

$$R_n(\tau)=\int_{f_1}^{f_2} A\mid f\mid^{-\alpha}\exp(-2\pi \mathrm{j}f\tau)\mathrm{d}\tau=2\int_{f_1}^{f_2} Af^{-\alpha}\cos(2\pi f\tau)\mathrm{d}\tau \tag{4.49}$$

Based on Eq. (4.49), the numerical calculation can show the auto-correlation function. If $\Delta f=f_2-f_1<<(f_1+f_2)/2=f_0$, then we can consider that $K_n(f)$ is a constant within $[f_1, f_2]$. In this situation, it is easy to prove that

$$R_n(\tau)=\frac{\sin(\Delta\pi f\tau)}{\Delta\pi f\tau}\cos(2\pi f_0\tau) \tag{4.50}$$

Eq. (4.50) is extensively used in active / passive sonar signal processing. Fig. 4.37 is the plot of this auto-correlation. If we choose the first zero point as the correlation radius, then the value of this radius is $\tau_0=1/(4f_0)$.

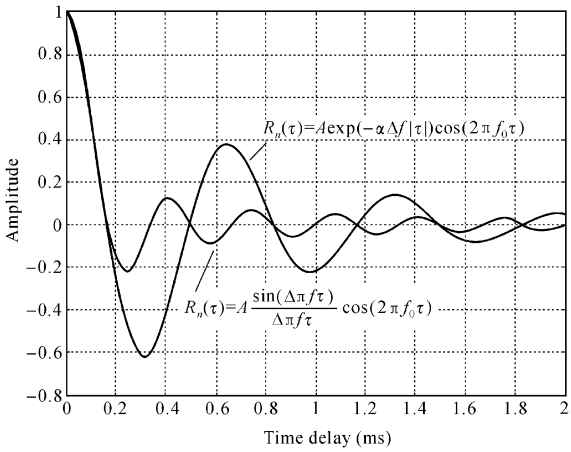


Fig. 4.37. Auto-correlation function of target noise

For comparison, a correlation function for the broad band signal is also plotted in Fig. 4.37, which is

$$R_n(\tau) = \exp(-\alpha\Delta f \mid \tau \mid)\cos(2\pi f_0 \tau) \tag{4.51}$$

In Fig. 4.37, the center frequency f_0 of the signal is 1,500 Hz.

4.5 Reverberation

Reverberation is one of the most important phenomena, which appears in active sonar signal transmitting. It is also the main concern factor in active sonar detection, particularly in reverberation limited circumstance[29-36]. Massive particles of various sizes exist, which can generate the effect of sound scattering in the ocean. Also, the non-flatness of the sea surface and sea floor will result in the re-transmission of the receiving signal. At any point in the ocean environment, the summation of the total scattering wave comprises reverberation. Reverberation is a part of the background noise of active sonar and is sometimes the main factor limiting the ranging distance of active sonar.

4.5.1 Sources of Reverberation

Reverberation can be categorized into three classes:
(1) Volume reverberation: sound wave scattering resulting from sea creatures and the non-uniform density of sea water;
(2) Surface reverberation: sound scattering resulting from scatters on the sea surface or near the surface;
(3) Bottom reverberation: sound scattering resulting from an uneven sea floor.

In the theoretical analysis of reverberation, it is assumed that the scattering sound described above is additive, the summation of total scattering comprising the reverberation. But, in practice, when the reverberation is received, it is hard to distinguish its components. A mass of experimental data validates the reverberation model of volume, sea surface and sea floor. Since the strength of reverberation is related to the source location and ocean environment, in different ocean environments repeat transmission of the same signal can verify if the mathematical model of reverberation is correct.

4.5.2 Short Distance Reverberation Theory

The reverberation behavior during a short period after signal transmission is of

concern in sonar system design. The reverberation level *RL* is a part of the background noise *NL*, but it is necessary to know that for a certain duration, the *RL* is a function of time and will dominate the *NL*. After a long period of signal transmission, the reverberation level decreases to a relatively low level and ambient noise will become the main factor limiting the detection range of the sonar system (Fig. 4.38).

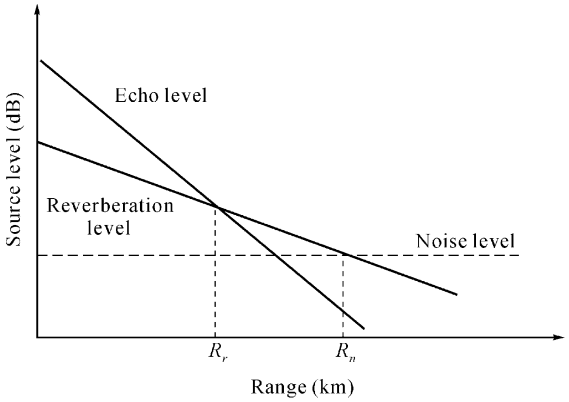


Fig. 4.38. Relationship between noise, echo and reverberation levels and sonar detection range

Scattering strength is a basic quantity for expressing the reverberation level. It is defined as the ratio of scattering intensity at 1 m above the incidental plane wave intensity, which is measured in unit volume or unit area. Denoted by dB, it is

$$S = 10 \lg \frac{I_{\text{scat}}}{I_{\text{inc}}} \tag{4.52}$$

where I_{inc} represents the incidental wave intensity and I_{scat} indicates the intensity of the scattering wave in m^2 or m^3. It is usually measured at a relatively distant point and is then transferred to 1 m.

In the sonar equation, if *RL* dominates *NL*, then we should replace *NL* by *RL* in the active sonar equation. In the calculation of *RL*, some assumptions should be made for simplification:

(1) The sound scattering propagates in a straight line form and the propagation loss can be calculated by spherical attenuation plus volume absorption.
(2) The distribution of scatters in a certain area or volume at any moment is uniform.
(3) The density of scatters is high enough such that any small area unit ds or volume unit dv has many scatters.
(4) The impulse duration of the transmitting signal is short enough such that the propagation effect inside the area or volume unit is negligible.
(5) Any secondary scattering is negligible.

Based on the above assumptions, the mathematical expressions for volume and boundary reverberation can be derived[30].

4.5.3 Volume and Boundary Reverberation Levels

This section will introduce Volume reverberation level and boundary reverberation level.

4.5.3.1 Volume Reverberation Level

Volume reverberation can be expressed as

$$RL_v = SL - 40\lg r + S_v + 10\lg V \tag{4.53}$$

where $SL = 10\lg I_0$, and I_0 is the source level of the transmit signal, r is the range of the receiver and transmitter in m, S_v is the volume reverberation defined by Eq. (4.52) and

$$V = \frac{c\tau}{2}\psi r^2 \tag{4.54}$$

where c is the sound velocity, τ is the impulse width and ψ is the equivalent synthetic beam width of the directivity $b(\theta, \varphi)$ of the transmit transducer and the directivity $b'(\theta, \varphi)$ of the receiving hydrophone. This can be expressed as

$$\psi = \int_0^{\pi}\int_0^{2\pi} b(\theta, \varphi)b'(\theta, \varphi)\mathrm{d}\theta\mathrm{d}\varphi \tag{4.55}$$

The term V expressed by Eq. (4.54) is called the reverberation volume, i.e., at any moment, it is the volume of scatters which result in reverberation.

4.5.3.2 Boundary Reverberation Level

Boundary reverberation (including surface and bottom reverberation) can be expressed as:

$$RL_b = SL - 40\lg r + S_b + 10\lg A \tag{4.56}$$

where

$$A = \frac{c\tau}{2}\Phi r \tag{4.57}$$

is the reverberation area, and Φ is the equivalent synthetic plane beam width

$$\Phi = \int_0^{2\pi} b(0,\ \varphi)b'(0,\ \varphi)\mathrm{d}\varphi \tag{4.58}$$

S_b is the scattering strength of the sea surface or sea floor as defined in Eq. (4.52). For an omni-directional transmitter and a receiving hydrophone, we have $\psi = 4\pi,\ \Phi = 2\pi$. The former is solid angle and the latter is circle angle. Table 4.3 lists the equivalent synthetic beam width of some popular transducers.

Table 4.3 Synthetic equivalent beam width of some transducers

Transducer	$10\lg\psi$ (dB)	$10\lg\Phi$ (dB)
Circle plane, unlimited baffle ($a > 2\lambda$)	$20\lg(\lambda/(2\pi a)) + 7.7$	$10\lg(\lambda/(2\pi a)) + 6.9$
Rectangle, unlimited baffle($a,\ b >> \lambda$)	$10\lg(\lambda^2/(4\pi ab)) + 7.4$	$10\lg(\lambda/(2\pi a)) + 9.2$
Horizontal linear array($l >> \lambda$)	$10\lg(\lambda/(2\pi l)) + 9.2$	$10\lg(\lambda/(2\pi l)) + 9.2$
Omni-directional	$10\lg(4\pi)$	$10\lg(2\pi)$

In calculating the reverberation level, it is necessary to remember that *RL* is a time related quantity. The value depends not only on the source level of the transmitting signal, but also on the time of the transmitting moment. In Eqs. (4.53) and (4.56) the scattering strength S_v or S_b is time related. Once the active signal is transmitted, the volume reverberation appears immediately, but the boundary reverberation is different as scattering occurs only when the front of the signal impulse arrives at the sea surface or sea floor. Therefore, the strength of volume reverberations from large to small and of boundary reverberations from small to large, reach a peak value and then decrease (Fig. 4.39).

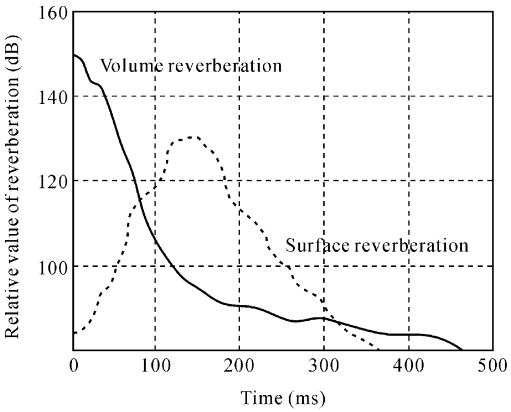


Fig. 4.39. Illustration of the average intensity of volume and surface reverberation

4.5.4 Relationship between Reverberation Strength and Impulse Duration

From Eqs. (4.53) and (4.56) we know that the reverberation strength is proportional to the impulse duration τ. But this conclusion is correct only in situations where τ is short enough (for example, $\tau \leq 100$ ms). This is because for small τ, the absorption and attenuation effect can be negligible in the scattering area or volume. For a large τ this kind of effect cannot be negligible until the reverberation strength is not related to the impulse duration and trends to saturation.

For the short duration τ, the strength of reverberation can be expressed as

$$I(t,\tau) = \frac{R_0 \tau \exp(-\beta\tau)}{t^m} \tag{4.59}$$

where R_0, β is constant. β is a coefficient which indicates the attenuation behavior of reverberation. Usually, it depends on sea water absorption and boundary loss. The value of m is related to the mechanism of reverberation and the propagation condition of the sound wave.

When the impulse duration increases, it can be considered as the summation of several contiguous short impulses. Therefore, the reverberation resulting from a long duration impulse can be expressed as

$$I(t,\tau) = K \int_{t}^{t+\tau} \frac{\exp(-\beta\tau)}{v^m} \mathrm{d}v \tag{4.60}$$

It is proved that if $\beta\tau \leq 0.1$, $I(t, \tau)$ is proportional to τ. In the case of a large $\beta\tau$, the reverberation strength reaches saturation and $I(t, \tau)$ is related only to time t. This result tells us that in active sonar design, the long time duration impulse has some advantages. Fig. 4.40 illustrates the relationship between reverberation strength and time. We can see the dynamic range is very large, even with a relatively small transmitting power. The dynamic range of reverberation is often about 80–120 dB.

This is a problem we have to face in digital sonar design, since with a general purpose pre-amplifier it is impossible to manage such large dynamics. A special treatment called the TVG (time varying gain) technique can be adopted to handle this problem. With the TVG technique, the gain of the pre-amplifier will change with time. At the moment of signal transmission the gain is very small and thereafter it becomes large until eventually remaining constant.

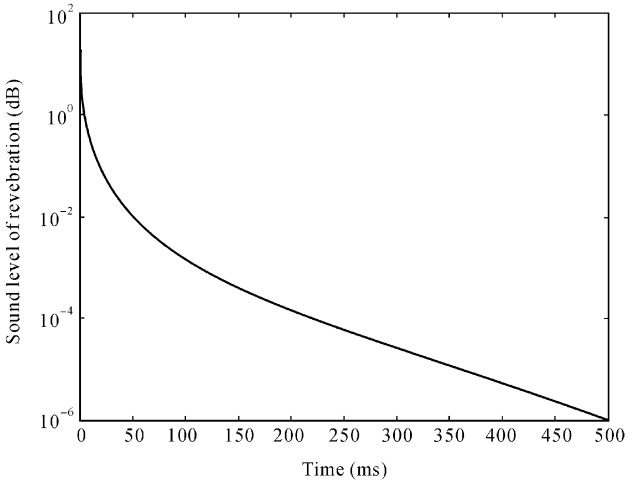


Fig. 4.40. Reverberation of short duration impulse

4.5.5 Statistical Characteristics of Reverberations

As the background noise, reverberation can be considered as a random signal. But this kind of random signal is different from ambient noise, although it also consists of the summation of massive scattering waves. This is because the scattering waves are generated from the same source. Therefore, the reverberation has specific statistical characteristics.

The statistical model of reverberation described below was proposed mainly by Middeleton[30-31] in the early 1960s. The non-uniform scatters which result in the scattering wave in the sea are considered as randomly "embedded" in the sea surface or sea bottom. The scattering coefficient of the sea surface, sea floor and sea volume are indicated by α_s, α_b, and α_v respectively. The amplitude and phase are indicated by $|\alpha_s|$, $|\alpha_b|$, $|\alpha_v|$ and φ_s, φ_b, φ_v, respectively. Suppose they are all mutually independent. All secondary scattering is negligible because the scattering wave is weak.

Suppose the transmitting signal is a narrow band impulse signal with the complex form $s(t) = |s_0(t)| \exp(\mathrm{j}\psi(t))$. The reverberation at moment t can be expressed as

$$p(t) = \sum_{n=1}^{N} g(r_n) f(r_n) |\alpha_n| |s_0(t-t_n)| \exp[\mathrm{j}2\pi f_0(t-t_n) + \psi(t-t_n) + \psi_n] \quad (4.61)$$

where $g(r_n)$ is the number of scatters within micro volume ΔV_n at distance r_n.

The possible value of $g(r_n)$ is zero or one; $f(r_n)$ is the two-way attenuation factor of scatter withinΔV_n; t_n is the moment of echo arrival; N is the total number of ΔV_n which have contributed to the scattering wave in moment t. Denote the real part of $p(t)$ as

$$\mathrm{Re}[p(t)] = x(t)\cos(2\pi f_0 t) - y(t)\sin(2\pi f_0 t) \tag{4.62}$$

Then

$$x(t) = \sum_{n=1}^{N} g(r_n) f(r_n) \mid \alpha_n \mid\mid s_0(t - t_n) \mid \cos(-2\pi f_0 t_n + \psi(t - t_n) + \varphi_n) \tag{4.63a}$$

$$y(t) = \sum_{n=1}^{N} g(r_n) f(r_n) \mid \alpha_n \mid\mid s_0(t - t_n) \mid \sin(-2\pi f_0 t_n + \psi(t - t_n) + \varphi_n) \tag{4.63b}$$

For the sea surface reverberation, because $g(r_n)$, $|\alpha_n|$, and φ_n are independent, the expectation value will be

$$\begin{aligned} &E[x_s(t)] = E[y_s(t)] = 0 \\ &E[x_s^2(t)] = E[y_s^2(t)] = \frac{1}{2}\rho_s E[|\alpha_s|^2] f^2(t) \int_S |s_0(t-s)|^2 \mathrm{d}s \end{aligned} \tag{4.64}$$

where the subscript s indicates the sea surface reverberation.

Denote the average strength of sea surface reverberation as

$$\sigma_s^2(t) = E[x_s^2(t)] = E[y_s^2(t)] \tag{4.65}$$

Because $x_s(t)$ and $y_s(t)$ obey the Gaussian probability distribution with zero mean and are independent, the amplitude of reverberation

$$r_s(t) = [x_s^2(t) + y_s^2(t)]^{1/2}$$

obeys the Rayleigh distribution (Eq. (3.32)). The probability density function is

$$p(r) = \begin{cases} \dfrac{r_s(t)}{\sigma_s^2(t)} \exp\left[-\dfrac{r_s^2(t)}{2\sigma_s^2(t)}\right] & r_s(t) \geq 0 \\ 0 & r_s(t) \leq 0 \end{cases} \tag{4.66}$$

A similar derivation shows that the amplitude of volume reverberation is also a Rayleigh distribution.

For the sea bottom reverberation, because the scatters embedded in the sea floor are fixed, the reverberation signal can be considered as a deterministic signal

$$p_b(t) = r_b(t)\exp[\mathrm{j}(2\pi f_0 t + \psi(t))] \tag{4.67}$$

The two orthogonal components are the real part and image part of Eq. (4.67)

$$\begin{aligned} x_b(t) &= r_b(t)\cos\psi_b(t) \\ y_b(t) &= r_b(t)\sin\psi_b(t) \end{aligned} \tag{4.68}$$

Now, the total reverberation can be written as

$$p_c(t) = p_s(t) + p_v(t) + p_b(t)$$

where the amplitudes of $p_s(t)$ and $p_v(t)$ obey the Rayleigh distribution, and the phases ψ_s and ψ_v are uniform distribution random variables in $(0,\ 2\pi)$. Therefore, the amplitude $r(t)$ of compound reverberation $p_c(t)$ obeys a modified Rayleigh distribution called the Rice distribution

$$p(r) = \frac{r(t)}{\sigma_c^2(t)}\exp\left[-\frac{r^2(t)+r_b^2(t)}{2\sigma_s^2+2\sigma_v^2}\right] I_0\left[\frac{r(t)+r_b(t)}{\sigma_s^2+\sigma_v^2}\right]$$

where $I_0(\cdot)$ is the modified Bessel function of zero order.

From the above analysis, we can see that unlike ambient noise, reverberation is not a stationary random signal. The decrease in reverberation strength occurs very quickly. At any fixed moment, the amplitude of reverberation obeys the Rice distribution, and, if the effect of the sea floor is neglected, the reverberation obeys the Rayleigh distribution.

However, the spectrum of reverberation is also different from that of the transmitting signal. The spectrum spread always exists, even when the platform of the transmitter is fixed. This is because of the non-homogeneous nature of scatter distribution. The magnitude of the spectrum strongly depends on the ocean environment.

Fig. 4.41 shows an example of spectrum spread. The transmitting signal is a sinusoidal wave with an impulse width of 20 ms and center frequency of 500 Hz. The bandwidth of this signal is about 100 Hz; the spectrum spread of reverberation is about 140 Hz.

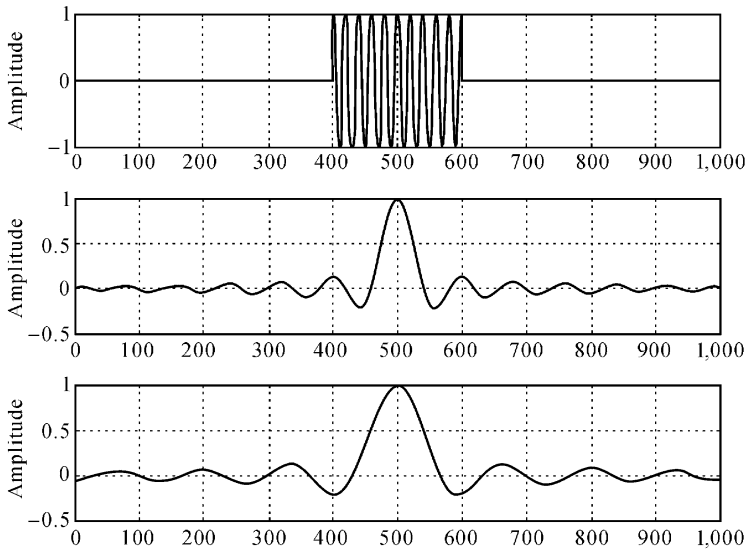


Fig. 4.41. Spectrum spread of reverberation

Fig. 4.42 is the plot of the Rayleigh probability function. The standard deviation is $\sigma = 1.0$.

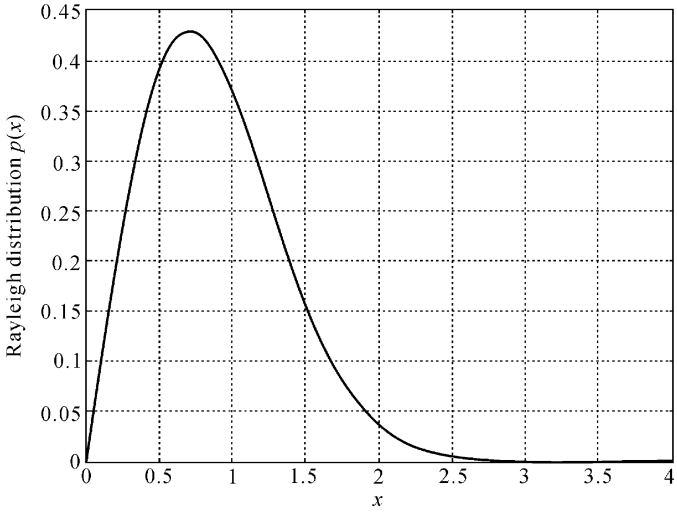


Fig. 4.42. Rayleigh distribution

In designing a sonar system, it is hard to predict the reverberation level. This is because the factors which affect reverberation strength are very complicated. From an engineering point of view, the sonar designer is concerned with the dynamic range and the echo reverberation ratio. The dynamic range is related to the design

of the dynamic range of the receiver, and the echo reverberation ratio is related to the detection performance of active sonar. Usually, it is defined by real data measurement.

Example

Assume the sound intensity of an omni-directional transducer is I kW. Therefore, $SL = 201$ dB. Suppose the background noise level is $NL = 80$ dB, then we have $SL - NL = 121$ dB. It is difficult to construct a receiver with a 121 dB dynamic range. A popular method is to set a "blind area" for the active sonar. This means that the sonar system is allowed not to function for a short time after the transmitting signal. Based on experience, in the shallow water of an isothermal layer, in Eq. (4.59) we choose $m \approx 3$. When the blind area is 500 m, the double- way signal travel time is 660 ms. If the impulse width is 30 ms, then the strength attenuation is about $30 \times \lg 22 = 46$ dB. As well as the additional exponential attenuation of the signal, the reverberation dynamic range can be reduced to 60 dB. This should be a reasonable requirement for a sonar system.

4.6 Sound Propagation in the Ocean and Underwater Acoustic Channel

A sound wave is the only physical energy media which can propagate a long distance under the sea. As the information carrier, the space and time structure of a sound wave is a basic study topic in modern underwater acoustics. In the broadest sense, the whole pathway of underwater sound wave transmission is the underwater sound channel, including the sea body, sea surface and sea floor. The behavior of sound transmission under the sea is very important for sonar design. In this section, the basic phenomena and laws of underwater sound propagation are introduced. The basic concept of an underwater acoustic channel is presented. The discussion will focus on the theoretical results of most interest to sonar designers rather than the solution of wave equations in various boundary or initial conditions[37-41].

4.6.1 Sound Wave and Vibration

The basic theory of acoustics involves the study of vibration, waves and their propagation. If the direction of particle vibration is the same as the direction of wave propagation, then the wave is called a latitudinal wave. If the direction of particle vibration is perpendicular to the direction of wave propagation, then the wave is called a transversal wave.

In air and water, mechanical vibration produces only latitudinal waves. In

solids, mechanical vibration can produce both latitudinal and transversal waves. Fig. 4.43 is an intuitive figure showing the two different kinds of waves.

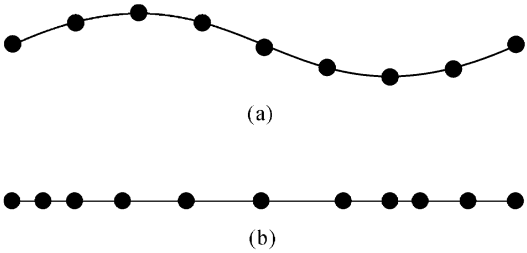


Fig. 4.43. Longitudinal and transverse waves. (a) Longitudinal wave; (b) Transverse wave

When a sound wave propagates in sea water, the structure of the water medium is changed, resulting in the spread of sound energy. The sensing of an underwater receiver for sound pressure is based on this sound pressure change. In the wave propagation process, particles in the sea water do not move from one place to another, but only vibrate around some fixed point. The acceleration speed of a particle is always proportional to the distance from a fixed point. This kind of motion is called resonance motion. It is the simplest form of periodic motion. The expression of a one-dimensional differential equation of resonance motion is

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \omega^2 x = 0 \tag{4.69}$$

The solution of this equation is

$$x(t) = A\sin(\omega t) + B\cos(\omega t) \tag{4.70}$$

where A, B are arbitrary constants and $\omega = 2\pi f$ is the cycle frequency with unit rad/s , and the unit of f is Hz.

4.6.2 Velocity of Sound in the Sea: Sound Speed Profile

The propagation velocity c of sound in the sea can be derived from the following adiabatic equation

$$c = \left(\frac{\partial p}{\partial \rho}\right)_s = \frac{1}{\rho K_a} \tag{4.71}$$

where p is the sound pressure, ρ is the density of water, and K_a is the adiabatic compression coefficient. In sea water, since ρK_a is a function of temperature,

salinity and pressure, the sound speed in sea water is also a function of temperature, salinity and pressure, but temperature is the dominant factor. In fresh water, the empirical formula for calculating sound speed is

$$c = 1410 + 4.21t - 0.037t^2 + 0.018d \ (\mathrm{m/s}) \tag{4.72}$$

where t is the temperature of sea water (℃); d is the depth (m). $c \approx 1{,}500$ m/s, for $t = 20$ ℃. The empirical formula for calculating sound speed in sea water is

$$c = 1410 + 4.21t - 0.037t^2 + 1.1S + 0.018d \ (\mathrm{m/s}) \tag{4.73}$$

where S is the salinity (‰). $c \approx 1{,}500$ m/s when $t = 14$ °C, $S = 34.5$, $d = 15$m.

Because the propagation characteristics of sound in the sea strongly depend on the sound speed, it is very important to understand the distribution of sound speed for any specific area. The relation between sound speed and depth is called the sound speed profile (SSP). The SSP is related to the latitude, season, and day / night. The SSP, which varies with many factors, is one of the most important pieces of basic data for a sonar designer. Generally, in shallow water of a continental shelf, the SSP is affected by more factors than it is in the deep sea. Fig. 4.44 is a typical SSP in the area of about $N = 40°$.

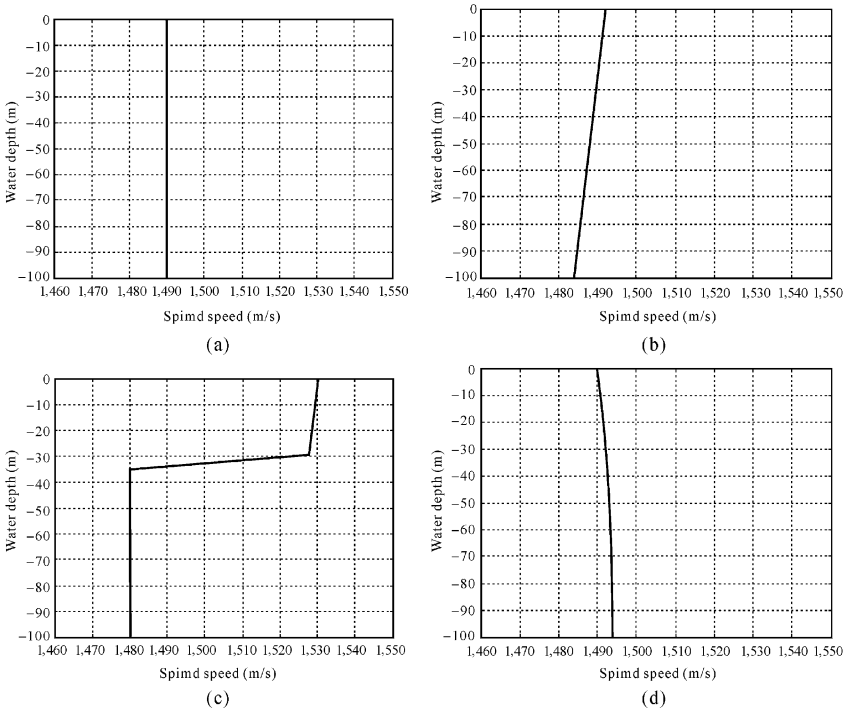


Fig. 4.44. Four types of SSP in shallow water near China Sea. (a) Isothermal layer; (b) Negative gradient; (c) Negative thermal cline; (d) Weak positive gradient

The SSP $c(z)$ is a function of the depth z. In a shallow water area, the sound speed is dominated mainly by temperature. Therefore, sometimes the term sound speed profile is replaced by the term sound temperature profile. Since the temperature can be measured directly, it is easy to understand that "isothermal layer profile" means that the sound speed profile has the same sound speed value. Similarly, it is easy to understand the terms "negative gradient" or "positive gradient" applied to temperature profiles.

The sound gradient is defined as

$$\frac{\nabla c(z)}{\nabla z} \quad (\mathrm{S}^{-1})$$

In Fig. 4.44 we can see that the SSP strongly depends on the season. From December to February, $c(z)$ is basically a constant. From March, from sea surface to sea floor, $c(z)$ changes from large to small because of the increasing sea surface temperature. This is simply a case of a negative gradient. In summer, i.e., from June to August, as the temperature of the sea surface keeps increasing, the SSP appears to suddenly change, and $\nabla c(z)/\nabla z$ may show a point of interruption at some depth. This is a case of a negative thermalcline. Then, with the change of season, the weather becomes cold and the SSP changes to a negative gradient and finally an isothermal layer again.

Fig. 4.45 shows an explanation for day/night change in SSP. In the day time, because of sunshine, the temperature of the sea surface is higher, and therefore the gradient becomes larger and sound speed increases. At night, the temperature of the sea surface decreases and therefore the gradient diminishes.

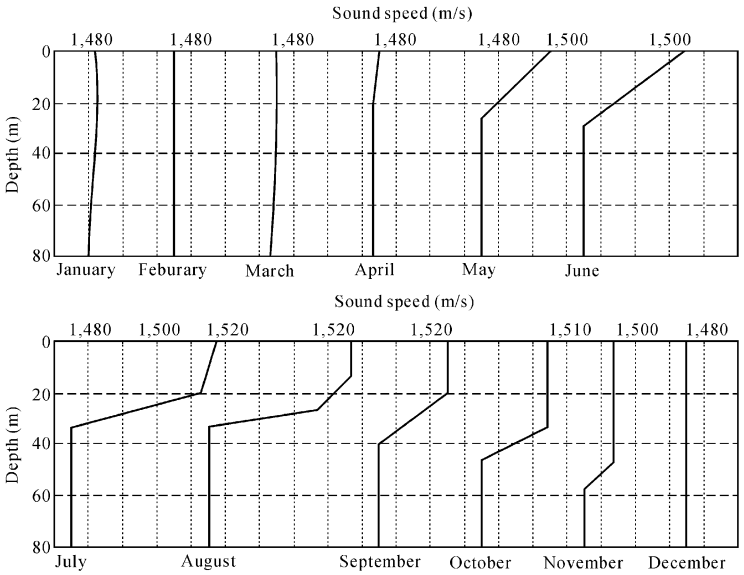


Fig. 4.45. Typical SSP over 12 months in shallow water near China Sea

Figs. 4.46 and 4.47 illustrate the SSP transition from shallow water (less than 200 m) to the middle scale of deep water (about 1,000 m). In the winter, there may sometimes be a positive gradient SSP.

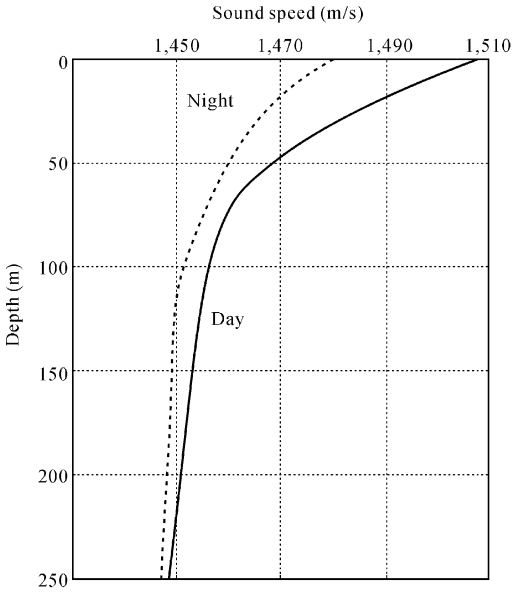


Fig. 4.46. SSP changes during day / night time

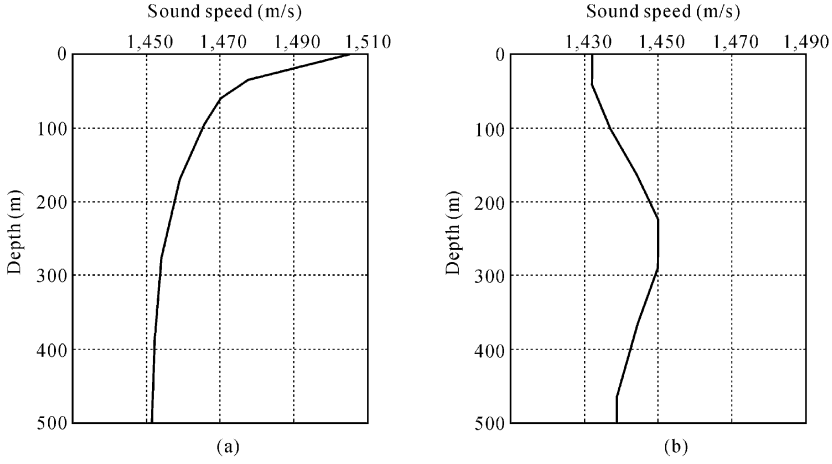


Fig. 4.47. Summer and winter SSPs in deep sea. (a) Summer SSP; (b) Winter SSP

In Fig. 4.48, a SSP typical of deep sea is illustrated. The sea surface layer is season dependent because of the relatively high temperature and the effect of wind. The SSP in this layer may be isothermal, a negative gradient or a weak positive

gradient. With increasing depth, the temperature decreases and the pressure does not have a definite affect on the sound speed. Therefore, the sound speed decreases to a minimum value. At deeper depths, where the temperature remains constant (4°C, the sound pressure will play a role in increasing sound speed. Finally a positive gradient will appear until the sea floor is reached. In terms of Eq. (4.73), the sound speed will increase by 1.7 m / s per 100 m depth, i.e., the sound speed gradient is $1.7\times10^{-2}/\mathrm{s}^{-1}$.

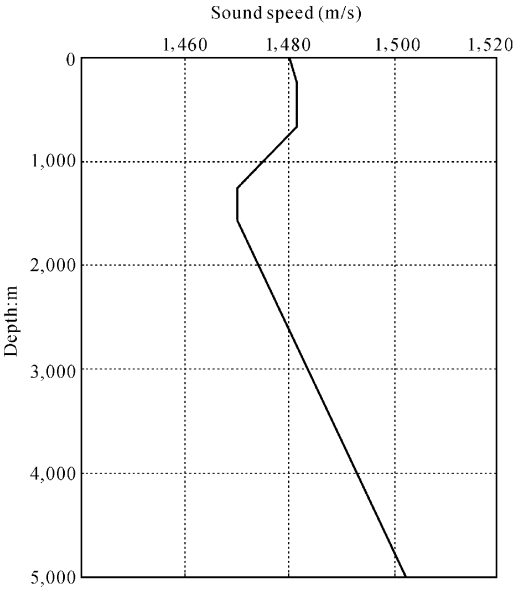


Fig. 4.48. Typical deep sea SSP at a depth of more than 3,000 m

4.6.3 Wave and Ray Theories of Underwater Sound Fields

The theory of the propagation of underwater acoustic signals has been studied since WW Ⅱ The basic theoretical framework, the wave theory and ray theory, has been completed. Wave theory considers both amplitude and phase information in sound waves whereas ray theory usually does not consider phase information. The application of these different theories depends on what kind of problems we are facing. Generally, ray theory is the approximate solution of a wave equation. If the difficulty of calculation is not taken into account in problem solving, then the propagation problems can be solved by wave theory. But, ray theory has obvious advantages in that it is simple, easy to understand and the physical meaning is quite clear. Therefore, ray theory is sometimes necessary in underwater acoustics.

We will briefly introduce the basic concepts of wave and ray theories in underwater acoustics.

Sound pressure $p(x, y, z, t)$ is a four arguments function, where x, y, z represent the coordinates of position and t indicates time. It satisfies the following wave equation

$$\nabla^2 p = \frac{1}{c}\frac{\partial^2 p}{\partial t^2} \tag{4.74}$$

where

$$\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right).$$

If $p(x, y, z, t)$ is the function of only x and t, then $p(x, t)$ is a plane wave. Eq. (4.74) shows that

$$p(x, t) = \phi(t - x/c) \tag{4.75}$$

where $\varphi(\cdot)$ can be any differentiable function, but has to satisfy the initial and boundary conditions. The most popular form of plane wave resolution is

$$p(x, t) = A\sin\omega(t - x/c) \tag{4.76}$$

where ω is the angle frequency of vibration and A is the amplitude. If $p(x, y, z, t)$ is only the function of $r = [x^2 + y^2 + z^2]^{1/2}$ and t, then we can obtain the resolution of a spherical wave. Denoting $p(x, y, z, t) = \psi(r)\exp(\mathrm{j}\omega t)$, Eq. (4.74) becomes

$$\frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial \psi}{\partial r}\right)\right] = -\frac{\omega^2}{c^2}\psi \tag{4.77}$$

It is proved that the general form of solution of Eq. (4.77) is

$$p = \frac{p_0}{r}\exp[\mathrm{j}(\omega t - kr)] \quad (k = \omega/c) \tag{4.78}$$

where p_0 is a constant and k is called the wave number. From Eq. (4.78) we see that the attenuation of a spherical wave is inversely proportional to the distance r. This is a very important fact when calculating the near range attenuation of a sound wave.

In a layered medium, the sound wave is neither a plane wave nor a spherical wave and therefore the solution of Eq. (4.74) is more complicated. In this situation, the solution of the wave equation is the summation of infinite elementary functions. From the physical point of view, the sound pressure in an acoustic field

is the summation of infinite harmonic vibrations (the acoustic mode). Since its calculation and derivation is complicated, we will not discuss it here.

In the classical ray theory of an acoustic field, the sound energy is considered to be transmitted by a sound ray, started from a source. The sound ray transmitted from the source along a certain path arrives at a receiver. The sound field in the receiver is the addition of all sound rays from various transmitters. Therefore, the ray which transmits sound energy is called a sound ray.

Generally, the shape of any SSP is not strictly a straight line. But we can always divide it into several pieces in each of which the SSP is linear. Therefore, the sound ray we are studying is the trace of ray paths in the layered medium. At first, we study the sound ray of a linear SSP. Assume the sound speed is $c(z)$, and

$$c(z) = A + gz \tag{4.79a}$$

where g indicates the gradient of sound speed, and A is a constant. The depth difference is

$$\Delta z = \frac{c_2 - c_1}{g} \tag{4.79b}$$

Similar to the optical ray, we can apply Snell's law (Fig. 4.49 (a))

$$\frac{\sin\varphi_i}{\sin\varphi_r} = \frac{\cos\theta_i}{\cos\theta_r} = \frac{c_i}{c_r} \tag{4.80}$$

where φ_i, φ_r represent the incidental angle and refraction angle respectively; θ_i, θ_r are the angles of the incident signal and refracted signal, respectively, with medium boundary; c_1, c_2 are the sound velocities of the incidental sound ray and reflected sound ray, respectively, in the medium.

The sound ray trace strongly depends on incidental angle of the sound wave and thereafter the reflection or refraction sound wave.

There are two special angles of sound propagation in underwater acoustics; one is the critical angle of transmission. When the sound wave propagates from one medium to another with different sound velocities c_1 and c_2, and $c_2 \geq c_1$ (Fig. 4.49(b)), the angle of the incidental wave such that the angle of the reflection wave is equal to $\pi/2$ is called the critical angle, i.e.,

$$\begin{aligned} &\frac{\sin\theta_{\text{crit}}}{\sin(\pi/2)} = \frac{c_1}{c_2} \\ &\theta_{\text{crit}} = \sin^{-1}(c_1/c_2) \end{aligned} \tag{4.81 a}$$

In the case of $\theta \geq \theta_{\text{crit}}$, the incidental wave will be fully reflected. This is why θ_{crit} is called critical angle. In the case of sea bottom reflection, the critical angle has another meaning (Fig. 4.49 (b)). $\pi/2-\theta_{\text{crit}}$ is called the critical grazing angle, denoted by θ_{graz}. Clearly

$$\theta_{\text{graz}} = \cos^{-1}(c_1 / c_2) \tag{4.81b}$$

From Eq. (4.79), we see that at any depth the sound velocity can be expressed as

$$c(z) = c_0 \cos\theta \tag{4.82}$$

where, $c_0 = c(0)$ is the sound velocity in a horizontal direction.

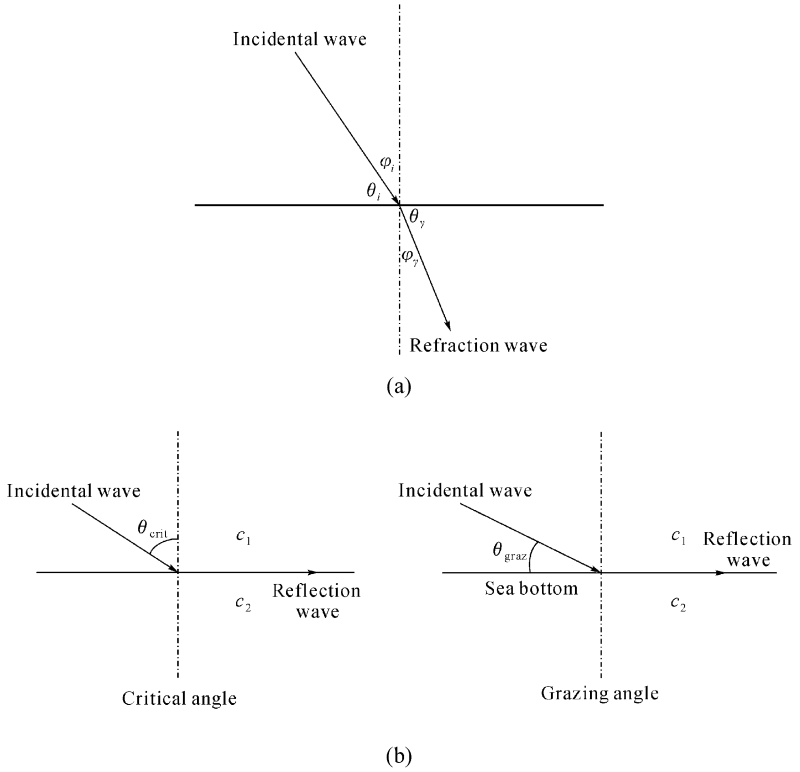


Fig. 4.49. (a) Explanation of Snell's Law; (b) Critical angle and grazing angle

Combining Eqs. (4.79) and (4.82), we obtain

$$\Delta z = -\frac{c_0}{g}(\cos\theta_1 - \cos\theta_2)$$

Integrating this equation, it follows that

$$z = -c_0(1-\cos\theta)/g \tag{4.83}$$

This is an equation of arc with radius $-c_0/g$ (Fig. 4.50).

If $g \le 0$, then $z \ge 0$, and the sound ray bends upwards. In contrast, if $g \ge 0$, then $z \le 0$, and the sound ray bends downwards. Thus, a sound ray started from a source will bend in a direction which has low sound velocity, regardless of the original incidental angle of the sound ray. Since the sound energy is brought by a sound ray, in sound propagation the sound energy is re-distributed non-uniformly. This fact plays a very important role in sonar signal detection.

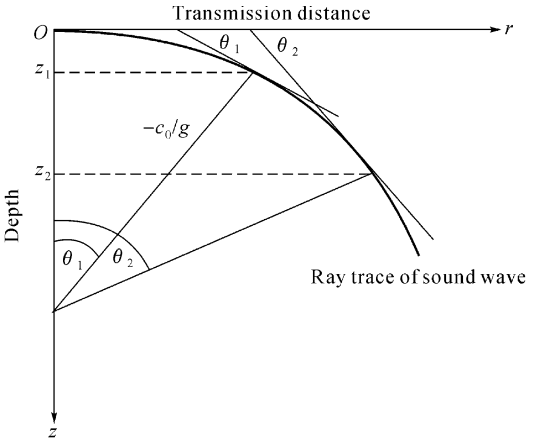


Fig. 4.50. Resolving the ray trace in a linear velocity profile

Eq. (4.83) forms the basis for sketching a sound ray path figure. In terms of SSP, we can plot the sound ray path, step by step. As an example, Fig. 4.51 illustrates the sound ray of a hull mounted sonar.

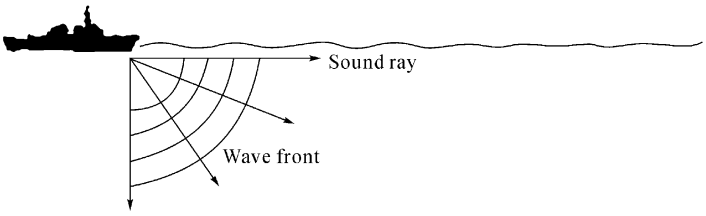


Fig. 4.51. Illustration of sound wave propagation

As mentioned before, for non-linear SSP, the segmented approximate method can be used to obtain a complete sound ray; it consists of many pieces of arc in different apertures.

4.6.4 Transmission Loss

In sonar equations, transmission loss *TL* is an important parameter in sonar design. Once the architecture of a sonar system is defined, the *FOM* of the sonar is decided. Therefore, the performance of a sonar system then depends only on the transmission loss. When a sound wave propagates in an ocean environment, the sound intensity will gradually decrease with travel distance, because of:
(1) The geometrical spread of the wave front;
(2) The loss of the sound wave at the sea surface and the sea floor;
(3) Sound absorption;
(4) Sound reflection.

Clearly, geometrical spread will result in decreasing intensity (Fig. 4.52(a)). In short range, the sound wave spreads in the spherical form and therefore the transmission loss is proportional to the square of the distance R

$$TL(R) \approx 20 \lg R \tag{4.85}$$

In this case, from 1 m to 100 m, the intensity of the sound wave will attenuate by 40 dB.

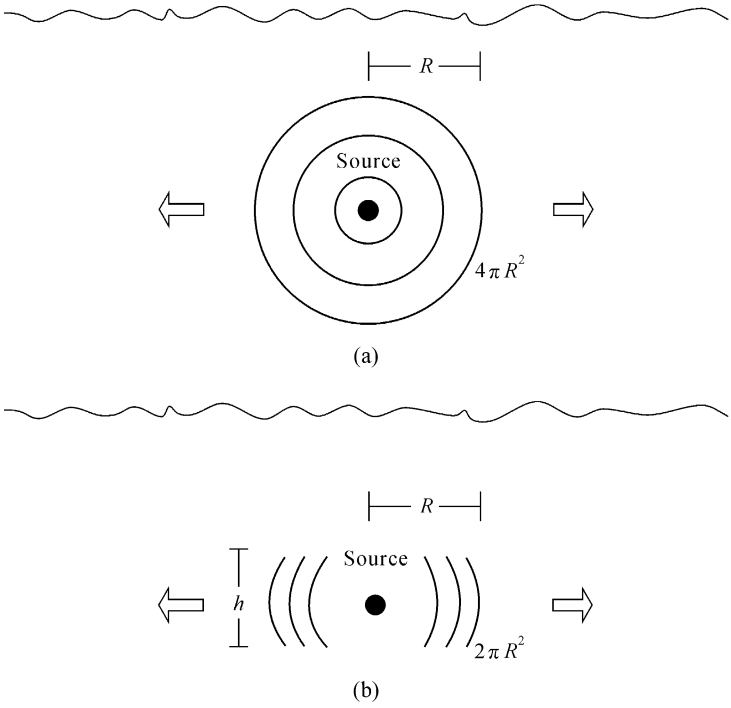


Fig. 4.52. (a) Spherical spread; (b) Cylindrical spread

In the propagation of a sound wave, if the sea surface and sea floor are relatively flat, sound reflection and absorption are negligible, so the spread of a sound wave can be considered as cylindrical (Fig. 4.52(b)). The transmission loss is proportional to the distance R

$$TL(R) \approx 10 \lg R \tag{4.86}$$

In this case, from 1,000 m to 2,000 m, the transmission loss should be 3 dB.

4.6.5 Sound Absorption in Sea Water

Sound absorption in sea water is one of the important characteristics of an acoustic channel. This is because, as the information carrier, in the propagation of a sound wave, energy dissipation is the main characteristic of a channel. Sea water is not an ideal medium for sound transmission. Viscous absorption and thermal conductor absorption result in the transition of sound energy to thermal energy. In fact, thermal conductor absorption is much lower than viscous absorption, and can be negligible. Theoretical study shows that viscous absorption results from a change in sound pressure. The compactness of the molecular architecture also changes with the sound pressure, as part of the sound energy is absorbed. Some useful formulas for calculating the absorption coefficient can be found in the literature[1, 42-44].

The absorption coefficient α is related to the frequency f. When $f \le 10$ kHz, use of the Mellen formula is recommended

$$\alpha = \frac{Af^2}{1+Bf^2} + 0.0035 f^2 \quad (\text{dB/km}) \tag{4.87}$$

where $A = 0.11 \times 10^{(\text{pH}-8)} / f_r$, $B = (1/f_r)^2$, f_r is the relaxation frequency.

$$f_r = 10^{(T-4)/100} \quad (\text{kHz})$$

T is the temperature of sea water (°C), and the pH value (density of hydrogen ions) is about 7.3 – 8.5. Thorp proposed an empirical formula for calculating sound absorption

$$\alpha_0 = \frac{0.1 f^2}{1+f^2} + \frac{40 f^2}{4.1+f^2} + 2.75 \times 10^{-4} f^2 + 0.003 \quad (\text{dB/ky}) \tag{4.88}$$

The unit f is kHz. We can transfer the unit dB/ky to dB/km by multiplying by a factor of 1.094

$$\alpha = 1.094 \times \alpha_0 \quad (\mathrm{dB/km}) \tag{4.89}$$

The results of calculations from Eq. (4.89) are shown in Table 4.4.

Table 4.4 Absorption coefficients

Frequency (kHz)	0.1	0.2	0.3	0.5	1.0	2.0	3.0	4.0	5.0	6.0	8.0	10.0
1 km absorption loss (dB)	0.001	0.004	0.010	0.024	0.065	0.130	0.194	0.273	0.370	0.487	0.780	1.150

Fig. 4.53 shows a plot of absorption coefficients produced by the US navy[43].

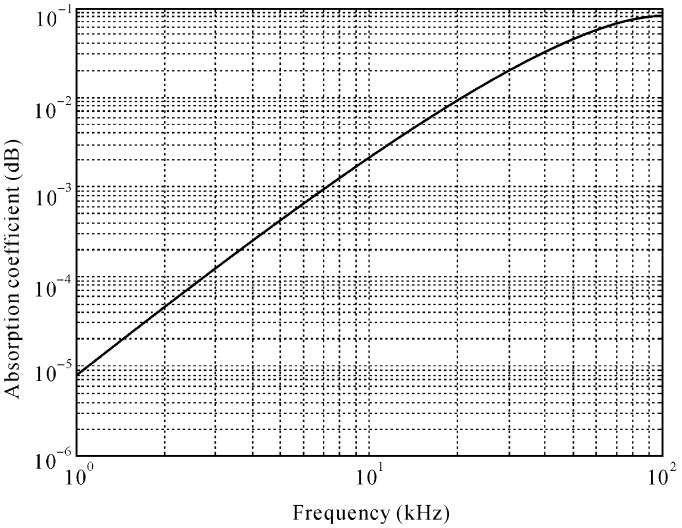


Fig. 4.53. Acoustic absorption coefficient

4.6.6 Upper Boundary of Acoustic Channel: the Sea Surface and Its Acoustic Characteristics

The sea surface is one of the boundaries of acoustic channel. When the sound wave propagates from sea water into air, it results in reflection, refraction and transmission. The reflection performance is related to the impedance of the medium. Here, the impedance is equal to the product of density and sound velocity. For the flat sea surface, since the impedance ratio of water and air is about 3,000:1, the incidental signal from water to air will all be reflected. Very little energy will be transmitted to air. That means the most parts of a sound ray going from water to air will come back to water. Generally, the path of a back sound ray is very complicated. This is because of the random fluctuation of the sea surface, affected

by wind waves. At the sea surface, if the variation of wave height with time is observed, then it is clear that it is a stationary random process. It is possible to measure the correlation function and power spectrum of a sea wave. It is reported that the correlation radius of wave signal is about 1 – 5 s. Newman gives an empirical formula for estimating the wave spectrum[37]

$$S(\omega)=\frac{c_1}{4\omega^6}\exp(\frac{-2g^2}{\omega^2u^2}) \tag{4.90}$$

where $c_1=4.8\ \text{m/s}^6$, the adjusted coefficient, $g=9.8\ \text{m/s}^2$, u is the wind speed (m/s), and $\omega=2\pi f$ is the angle frequency (rad/s). $S(\omega)$ decreases rapidly and when $f\geq 5$ Hz, $S(\omega)\approx 0$. Fig. 4.54 shows the acoustic very in case that source new see sulface.

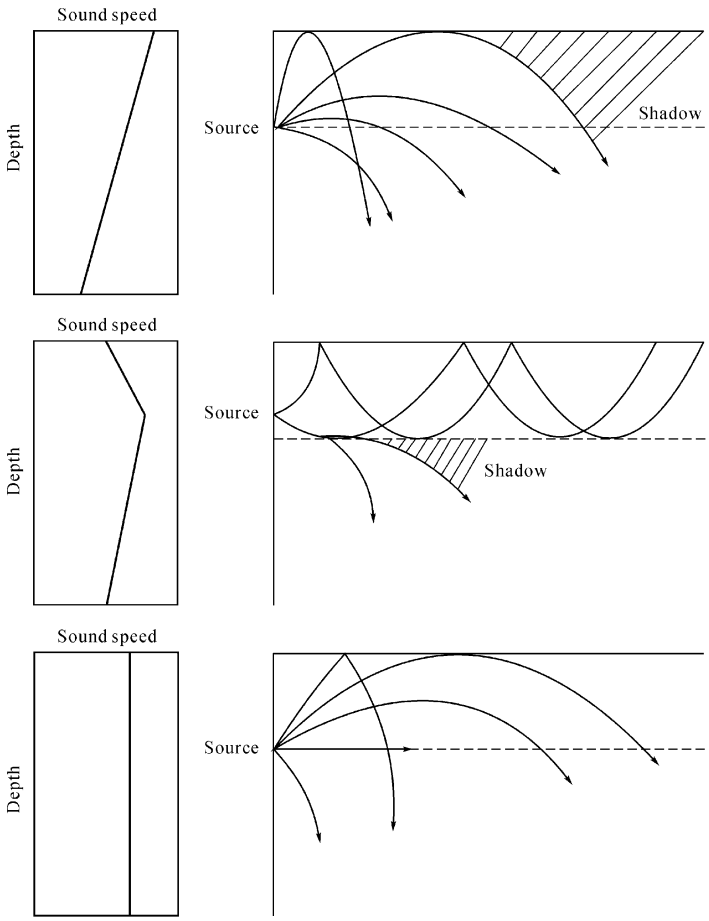


Fig. 4.54. The acoustic ray in the case that source near sea surface

4.6.7 Lower Boundary of Acoustic Channel: the Sea Floor and Its Characteristics

The crust of the sea floor is covered by a layer of non-solidified sediment. In underwater acoustics, the sea bottom boundary is defined by the boundary of sea water and this sediment. A sound wave propagated at the sea bottom has a different velocity depending on the structure of the sea floor. Experimental measurements show that most of the continental shelf belongs to a high velocity sea bottom, i.e., $c_{\text{bottom}} > c_{\text{water}}$ and most of the deep sea bottom belongs to a low velocity bottom, i.e., $c_{\text{bottom}} < c_{\text{water}}$.

A sound wave incidental to the sea bottom will result in reflection, transmission and refraction. The sonar designer is concerned with the reflection coefficient, since it will affect the sonar detection performance. The value of the reflection coefficient depends on the roughness and structure of the sea bottom. Because the sea bottom is usually uneven, the reflection and refraction of a sound wave is more complicated there than at the sea surface. In shallow water, the reflection of a sound wave is the main source of the so-called "multi path effect". For passive sonar, the multi path effect will result in a decrease in detection ability and bearing accuracy. Target recognition will also become more difficult and the passive ranging function will be harder to realize in a situation which has serious multi path phenomena. For active sonar, the multi path effect will produce signal spectrum spread and false target echoes, making echo detection difficult.

Nowadays, almost all of reflection coefficient and attenuation values are based on practical measurement or are semi-empirical. For a specific ocean environment, the converse is sometimes possible. However, the study of the sea bottom is far from complete.

Table 4.5 Reflection coefficients of the sea bottom

Sea bottom	Average particle size (mm)	Reflection coefficient
Silt	0.011 – 0.013	0.16 – 0.17
Silt plus sands	0.015 – 0.030	0.21 – 0.37
Sands plus silt	0.25 – 0.28	0.45 – 0.50
Sands	0.26 – 0.59	0.41 – 0.85
Broken stone	1 – 1.4	0.51 – 0.82

4.6.8 Use of Propagation Characteristics in Sonar Design

Because of the extreme importance of the propagation characteristics of underwater acoustic signals, the sonar designer must know the propagation behavior of underwater sound channel in various conditions. Only with a good understanding of performance propagation behavior, is it possible to use the underwater acoustic propagation law in a sound channel.

The performance of sonar may be totally different in the hands of different users and operators. In the case of detecting a hostile target, it is necessary to find a favorite position from which to extract signals to be detected. In contrast, in the case of anti-detection, it is necessary to hide in a shadow in which it is difficult to be detected by the hostile sonar. In any situation, it is important to plot the sound wave ray based on the ocean environment, and then to choose the best position for the transducer and hydrophone array, and the operating frequency band, etc.

Figs. 4.55 – 4.59 illustrate a series of examples of detection scenarios[12].

Fig. 4.55 shows the sound ray distribution of a surface ship sonar in a negative gradient SSP. As we can see, most of the sound rays bend towards to sea bottom and therefore the sonar ranging distance is limited. Fig. 4.56 shows how a submarine can hide itself in a positive gradient SSP by positioning itself in the area of a sound shadow. Fig. 4.57 illustrates a very good SSP, a sea surface duct. In this situation, the communication distance and detection distance of sonar will reach maximum values.

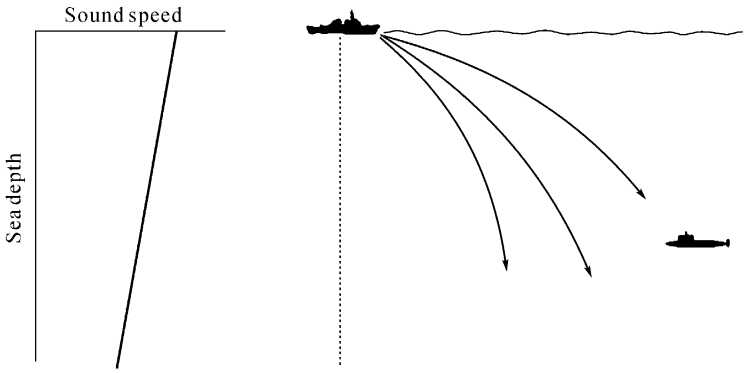


Fig. 4.55. The bow sonar of a surface ship operating in a negative gradient SSP

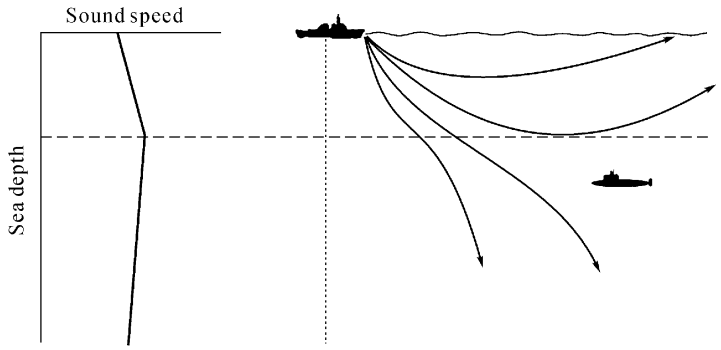


Fig. 4.56. In a weak positive gradient, a submarine hides in a sound shadow

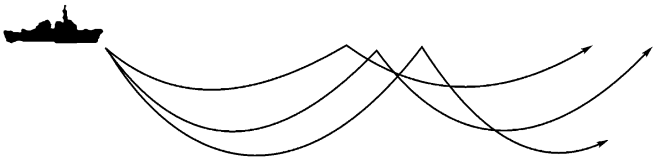


Fig. 4.57. Surface sound channel

Fig. 4.58 shows an example of deep sea propagation. At a depth of about 1,000 m, there is a channel called SOFAR (Sound Fixing and Ranging), in which the sound velocity is at a minimum. If the sound source is placed at this depth, then all sound waves transmitted from this point will bend back to this depth. Therefore, most of the sound energy will keep propagating, except that which is absorbed. It is reported that a low frequency signal transmitted in SOFAR channel can be heard thousands of kilometers away.

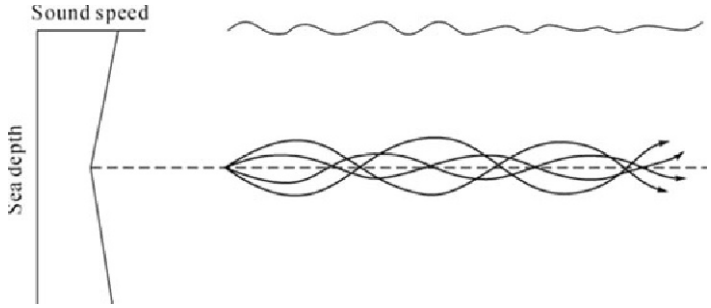


Fig. 4.58. Sound propagation in SOFAR channel

Fig. 4.59 presents another abnormal phenomenon in deep sea propagation, the convergence zone. Because of the specific focusing effect of a sound lens in deep sea, it is possible for a sound ray, under specific conditions, to focus at a certain distance. If the target appears in this zone, it will be easily detected. However, except in this convergence zone, the target can be detected only within a short distance, close to the sonar system. A convergence zone can be marked according to its order number, i.e., the first, second, and so on.

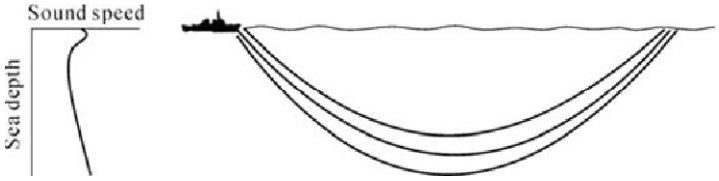


Fig. 4.59. Convergence zone effect in deep sea sound propagation

Special conditions are necessary for the formation of a convergence zone, including a suitable depth and temperature of the sea water. Tables 4.6 and 4.7 list

the relative conditions. Note that the convergence zone has a width of about 5 – 9 km. The boundary of the convergence becomes fuzzy as the order number of the convergence zone increases. The gain of the first convergence zone is about 20 dB, and the typical distance of the first convergence zone is 55 – 64 km.

Table 4.6 Conditions for forming a convergence zone

Temperature of sea surface	≤10	10–20	20–30
Depth of sea water required (km)	2	4	6

Table 4.7 Convergence zone in North Atlantic and North Pacific

Temperature in sea surface (°C)	10	15	20	25	30
Distance of first convergence zone (km)	43	54	62	66	70

4.6.9 Average Structure of a Sound Field in Shallow Water

"Shallow water" is a fuzzy concept. From different points of view, the definition is different. In geography, shallow water indicates a bay, port or inner sea, with a depth of less than 200 m. It usually extends to the margin of the continental shelf. However, in acoustics or oceanography, shallow water is considered to present a kind of acoustic channel, in which for a sound wave there are multiple reflections from the sea surface and sea bottom. Therefore, the concept of shallow water is related to the structure of the sea bottom.

Nevertheless, we often understand that a depth of less than 200 m is shallow water. Near China Sea, except in South China Sea and East of Taiwan, most of the ocean environment is shallow water. The propagation of a sound wave in shallow water shows some specific behavior, since the acoustic channel is bounded by both the sea surface and the sea bottom. Therefore, the propagation law is in a way more complicated than in deep sea. Theoretical research work and at-sea experiments completed by Chinese researchers are valuable[12, 45]. Some results from theoretical research work and experimental analysis are briefly presented here.

The sound field in shallow water is not only constrained by the sea surface and sea bottom, but also by the SSP $c(z)$. From an energy average point of view, the ability of the sea surface and sea bottom to "capture a sound wave" is determined by the sound reflection. The role of the SSP is to control the energy distribution throughout the depth, the amount of "bouncing" from the boundary, and the angle of the sound ray.

(1) Isothermal layer in shallow water (Fig. 4.60(a))

The structure of the strength of a sound field consists of several segments. Each segment has basically the same attenuation law for strength. In this plot, r_1 is about 5 – 10 times the depth; r_2 is 10 – 30 times the depth.

In area A: $I(r) \approx A / r^2$, when $r \le r_1$;

In area B: $I(r) \approx B / r$, when $r_1 \le r \le r_2$;

In area C: $I(r) \approx C / r^{3/2}$, when $r \ge r_2$.

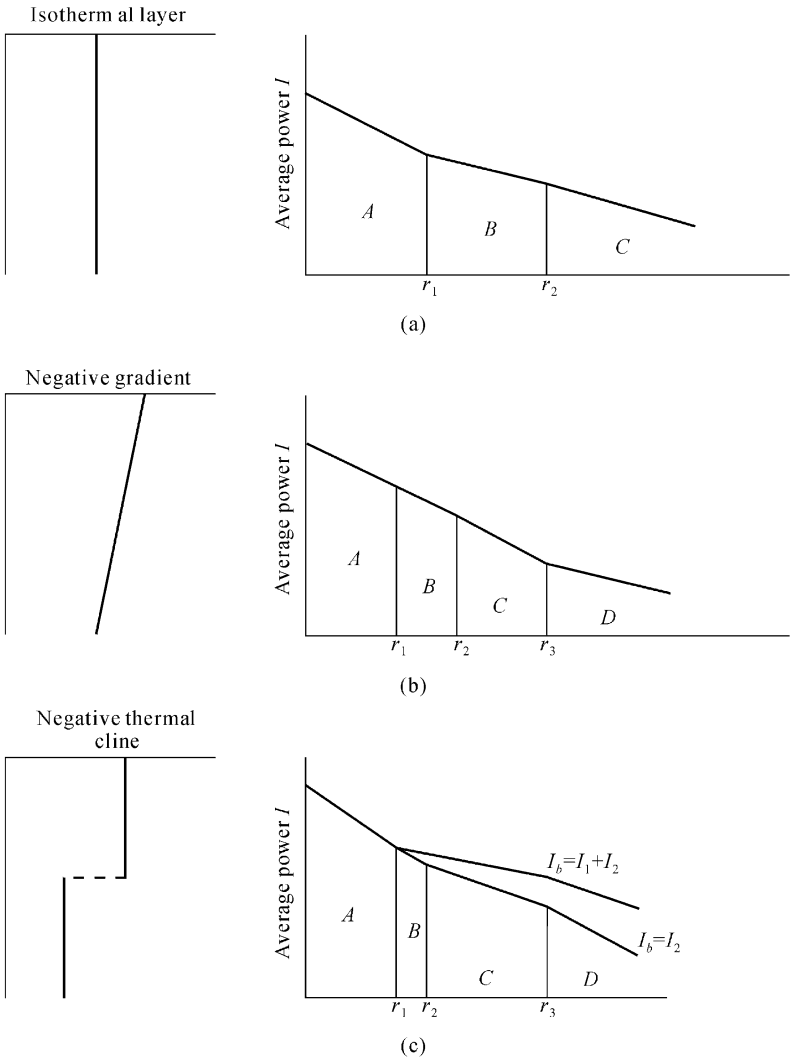


Fig. 4.60. Average smooth sound field in shallow water. (a) lsothermal layer in shadow water; (b) Negative gradient in shadow water; (c) Negative the thermal cline

2) Negative gradient in shallow water (Fig. 4.60(b))
The strength structure consists of four segments:

In area A: $I(r) \approx A / r^2$, when $r \le r_1$;

In area B: $I(r) \approx B / r$, when $r_1 \le r \le r_2$;

In area C: $I(r) \approx C\exp(-C'r)/r$, when $r_2 \le r \le r_3$;

In area D: $I(r) \approx D\exp(-C'r)/r$, when $r \ge r_3$, and $D' > C'$.

3) Negative thermal cline (Fig. 4.60(c))
In the case of a negative thermal cline, the mode can be divided into two different waves: the below layer wave, which cannot transmit to the layer and the transmitting wave, which transmits to the layer. If the average strength of the sound field of these different waves is indicated by I1 and I2, then the attenuation values of up transmit / low receiver and low transmit / up receiver are different. In this plot, in the case of up transmit / low receive, Ib = I2, and in the case of low transmit / low receive, Ib = I1+I2. The strength structure of the sound field is similar to that in the case of an isothermal SSP, but the attenuation rate is different.

4.6.10 *Use of Transmission Loss in Sonar Ranging Distance Prediction*

Based on the sonar equation, transmission loss is the important parameter in estimating sonar ranging distance. As the ocean environment is time / spatial variable, the best way to understand the characteristics of transmission loss is to measure it in sea trials. It is possible to make a theoretical estimation of transmission loss, but clearly this cannot replace the at-sea experiment. A database of measurement results is always the basis for sonar ranging distance prediction.

Fig. 4.61 illustrates a typical curve of transmission loss. As noted above, transmission loss depends on the frequency. In the range of 1 – 1000 m, we can consider that the sound wave is attenuated according to spherical spread. Also, for frequencies less than 10 kHz, the absorption by sea water can be considered negligible within a short range.

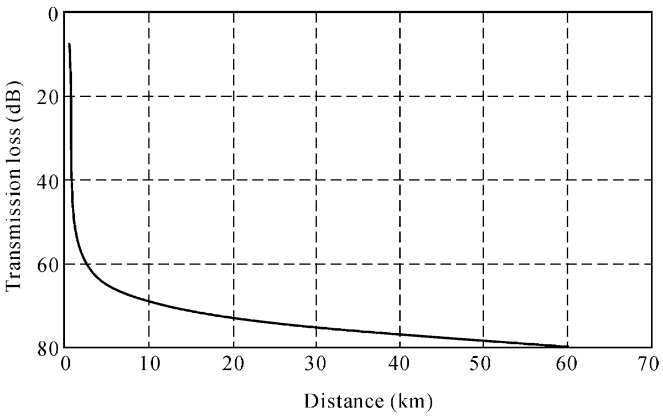


Fig. 4.61. Typical transmission loss curve of band limited signal

In discussing the sonar equation, we showed that the ranging distance of sonar is the resolution of the following equation

$$FOM = TL(r)$$

Fig. 4.62 shows an example of estimating sonar ranging distance. In Fig. 4.62 (a), the chosen estimation start point is 1 km, 60 dB. This means that as the attenuation of sound is spherical from 1 – 1,000 m, when the sound wave propagates 1,000 m, the source level decreases by 60 dB. The spherical and cylindrical attenuations are plotted in Fig. 4.62(a). The real transmission loss is somewhere between these two curves. We can use an approximation formula

$$TL(r) = 10\lg r^{\beta} + \alpha r \tag{4.91a}$$

where $1 \le \beta \le 2$, and α is defined by Eq. (4.88).

Or we can use

$$TL(r) = 60 + 10\lg r + \alpha r, \quad (r \ge 1 \text{ km}) \tag{4.91b}$$

This equation indicates that the transmission loss is spherical when the distance is less than 1 km, and cylindrical when the distance is more than 1 km.

As an example of an at-sea experiment in the Korean Strait[46], Table 4.8 shows values of transmission loss, calculated from 1 km.

Table 4.8 Transmission loss of underwater acoustic signals (Korean Strait)

Distance	TL at 200 Hz (dB)		TL at 400 Hz (dB)		TL at 800 Hz (dB)	
(km)	April	September	April	September	April	September
20 km	18 dB	22 dB	20 dB	27 dB	24 dB	39 dB
40 km	24 dB	31 dB	25 dB	37 dB	28 dB	52 dB
60 km	27 dB	36 dB	28 dB	44 dB	31 dB	65 dB

Fig. 4.62 (b) is an intuitive example where FOM =75 dB. To the right of the FOM, there is an undetectable area; to the left is a detectable area. Fig. 4.63 illustrates an example in which the same sonar has different ranging distances in different ocean environments, where FOM = 90 dB and f = 1,000 Hz.

In the case of a deep sea convergence zone, the propagation attenuation curve differs from those above in that the propagation attenuation curve decreases monotonously with distance. The propagation attenuation curve of a convergence zone will bounce in the point of convergence. When various conditions are in coincidence, the reflection from the sea bottom may result in abnormal transmission loss (Figs. 4.64 (a) and (b)).

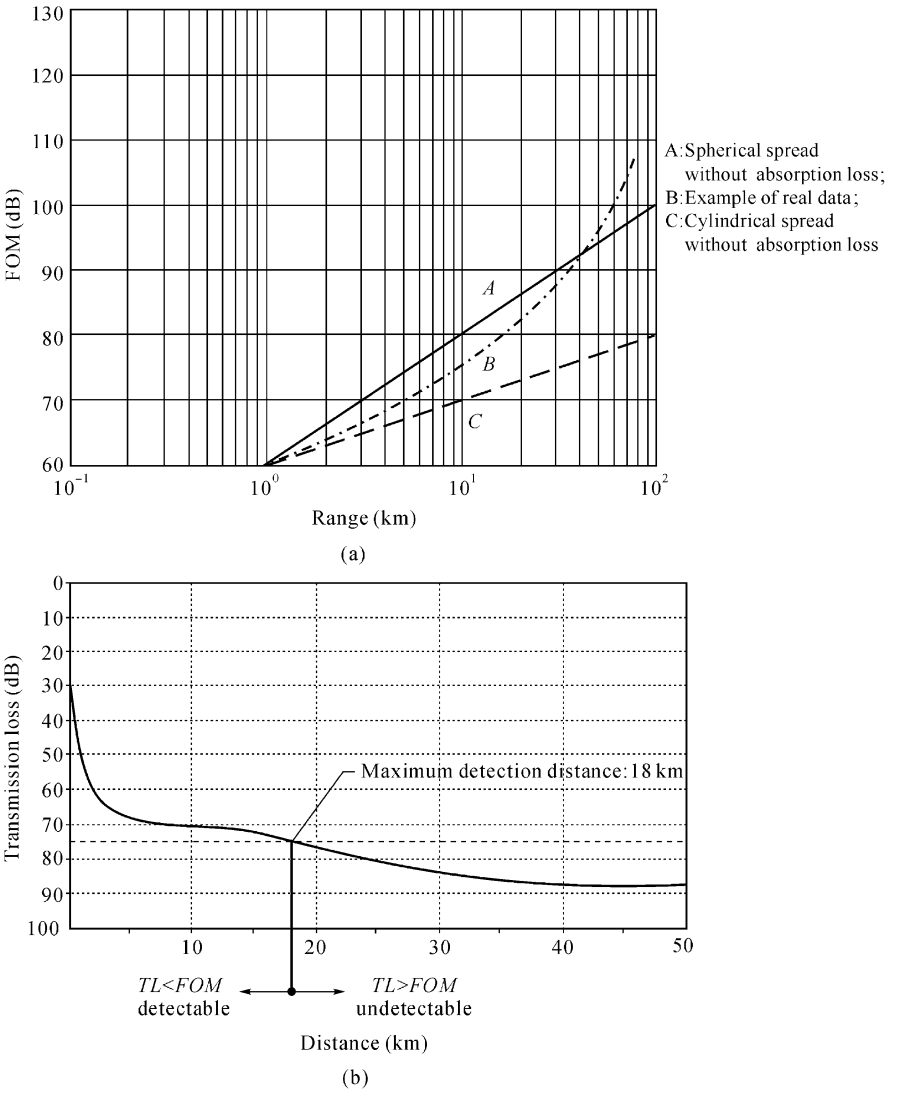


Fig. 4.62. Calculation of sonar ranging distance. (a) FOM and different transmission loss; (b) Calculation example

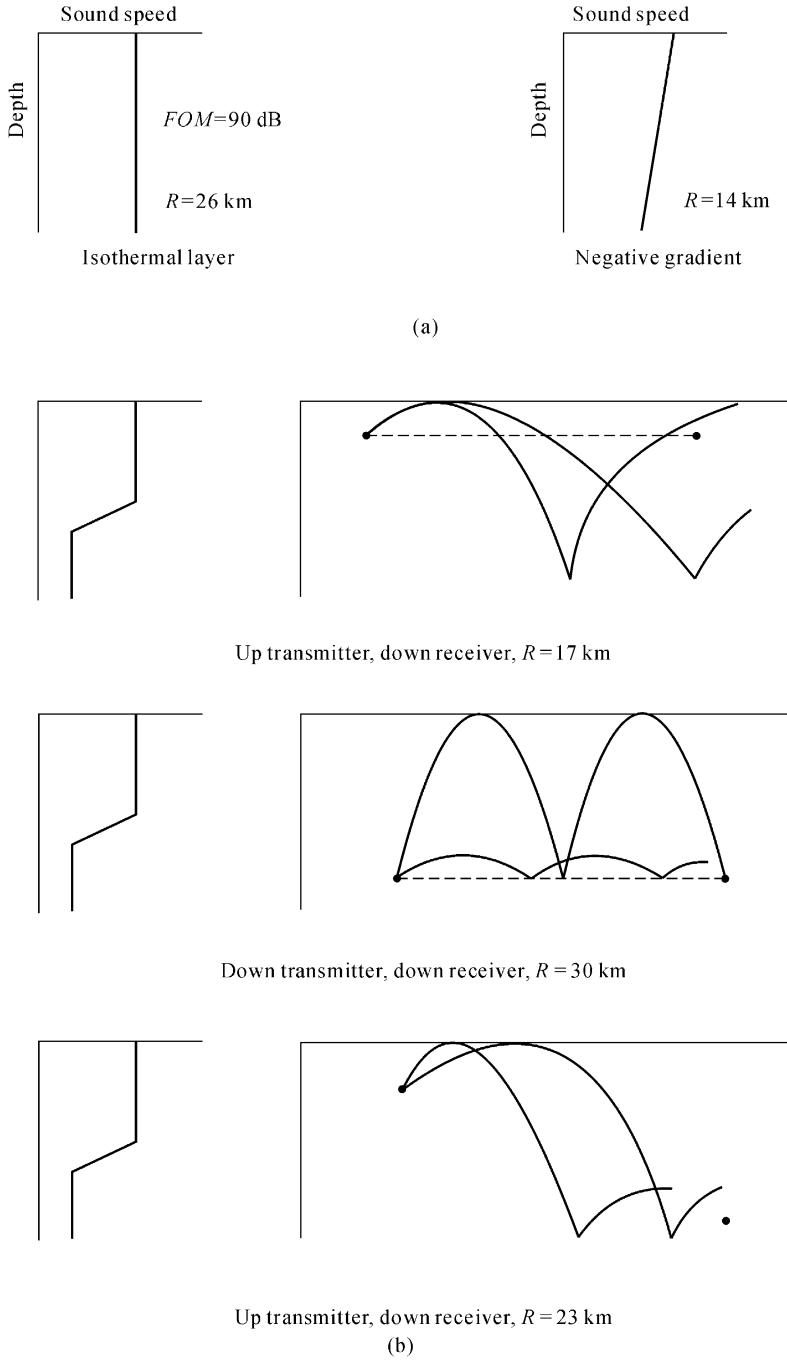


Fig. 4.63. Example of sonar ranging distance calculation. (a) Same place, different season; (b) Same place, same season, different transmitter / receiver position

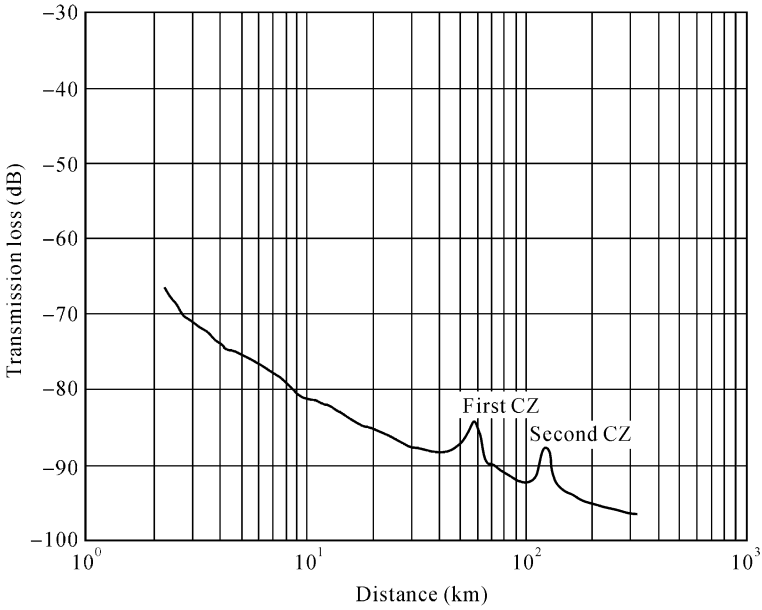


Fig. 4.64. Transmission loss curve when there is an existing convergence zone (CZ)

If one wants roughly to estimate the ranging distance of a sonar system, Eqs. (4.87) and (4.90) can be used for calculation. Table 4.9 shows an example of such a calculation. The distance is calculated in frequencies 1.0, 2.0, 4.0, 8.0 kHz and the absorption coefficients α_0 are 0.065, 0.130, 0.273, 0.780 dB / km.

Table 4.9 Example of sonar ranging distance calculation

	Isothermal layer β=1.5				Negative gradient β=1.9			
f (kHz)	1.0	2.0	4.0	8.0	1.0	2.0	4.0	8.0
Ranging distance (km), FOM = 85 dB	32	25	20	12	18	16	13	9
Ranging distance (km), FOM = 90 Db	49	39	30	16	30	25	20	12
Ranging distance (km), FOM = 95 dB	80	59	39	20	45	38	28	16

As noted above, the transmission loss of an underwater acoustic in a channel depends on the ocean environment. The best way to measure the value of transmission loss is by means of an at-sea experiment, but there is some useful software for calculating transmission loss in certain conditions. The most important use of this kind of software is for modeling propagation. A frequently used software package is called the Acoustic Toolbox[47-51], written by Mike Porter. The original software was written in the FORTRAN program language, which is no longer commonly used. Duncan provided a Matlab front-end[52] for the Acoustic Toolbox.

The software of the Acoustic Toolbox provides estimates for acoustic propagation of signal through an underwater channel by numerically solving the

propagation wave equation. It includes the following underwater models:
(1) Kraken normal mode model: a normal mode code for range-varying environments.
(2) Kraken C complex normal mode model: a complex normal mode code for range-varying environments.
(3) Scooter fast-field model: a finite element code for computing acoustic fields in range-independent environments based on direct computation of the spectral integral.
(4) Bellhop ray and Gaussian beam tracing model: a program which computes acoustic fields in ocean environments via beam tracing.

4.7 Hydrophone Array and Beamforming

An omni-directional hydrophone in the ocean environment can determine only if the target is present, but it is impossible to determine the incidental angle of the target. When we configure a group of hydrophones together, consisting of a hydrophone array, then it will be possible to determine the incidental angle of the target, provided a subsequent signal processor is connected. Target bearing, multi-target resolution or even target ranging is the main capability of a hydrophone array. The key technique in completing these tasks is the beamforming, which steers each hydrophone of the array to the target using time delay, and uses time accumulation to obtain system gain.

Beamforming is the kernel part of the signal processor of a modern sonar system, regardless of whether it is passive or active sonar. It can be considered as a spatial filter: it can allow the incidental signal from some direction to pass through, but it rejects all interferences and noise from other directions. Beamforming theory has been studied thoroughly since the 1940s, particularly in the field of radar antennas [53-96]. These studies included the analysis of directivity function, system gain calculation of arrays, optimum weighting, and array baffling. Many research results and technical principles in radar theory can be applied to sonar systems with some kind of modification.

4.7.1 Directivity Function (Beam Pattern)

There are many different kinds of array shapes, depending on the platform in which the array is to be installed. Fig. 4.65 presents five basic array shapes of sonar systems. Three are plane arrays: the equal-spaced line array, the uniform circle array and the conformal array. The other two are three-dimensional arrays: the spherical array and the cylindrical array.

The line array is used extensively in side scan sonar, towed array sonar and flank array sonar. The circle array is used in bow hull mounted sonar and sonar

buoys. Conformal sonar is used in submarine and torpedo sonar. Cylindrical and spherical sonar are used in large submarine or surface ship sonar systems.

To analyze the performance of any kind of sonar array, a three-dimensional spherical or right angle coordinate system can be established.

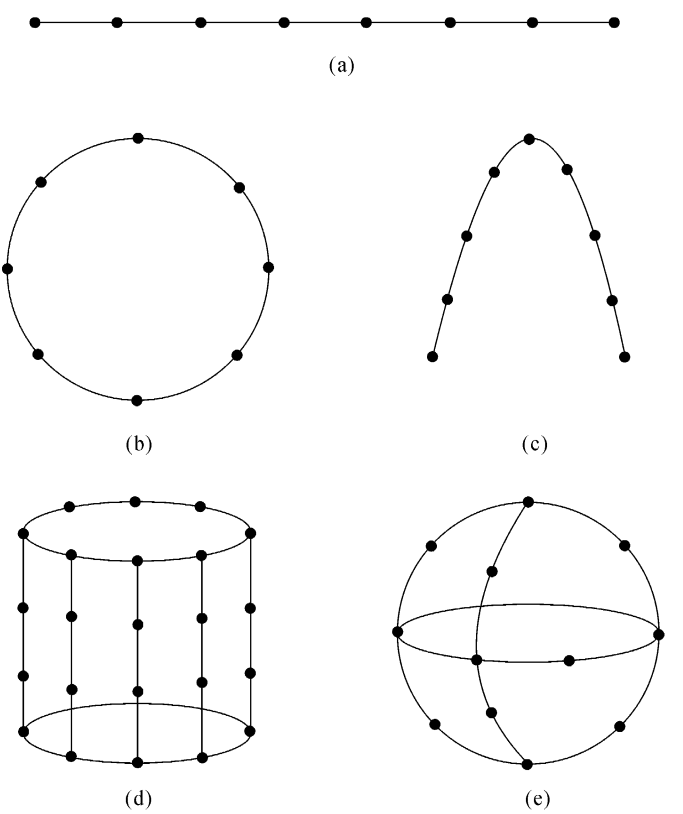


Fig. 4.65. Different types of discrete distributed element arrays. (a) Line array; (b) Circle array; (c) Conformal array; (d) Cylindrical array; (e) Spherical array

Suppose the incidental direction of a signal is (θ, φ), if the intensity of the signal remains constant then the system $D(\theta, \varphi)$ response $D(\theta, \varphi)$ indicates the sensitivity of this array for the spatial incidental signal. $D(\theta, \varphi)$ is called the directivity function or beam pattern. Intuitively, the directivity function indicates the following abilities of an array:

(1) To focus the signal energy in a specific direction;

(2) To suppress the interferences and noise from other directions.

A sonar designer should be concerned about the performance of the directivity function and needs to take care in its design. It is really the key problem requiring attention.

Fig. 4.66 presents three types of directivity function in a plane: right angle (Cartesian) coordinate, logarithm coordinate and polar coordinate. These three kinds of plot are all useful, each in a different situation.

The main characteristics of the directivity function can be described by the following parameters. Assume that the plane directivity function is $D(\theta)$, the maximum value of $D(\theta)$ is obtained at $\theta = 0$, and $D(0) = 1$. We will say this directivity is the normalized function.

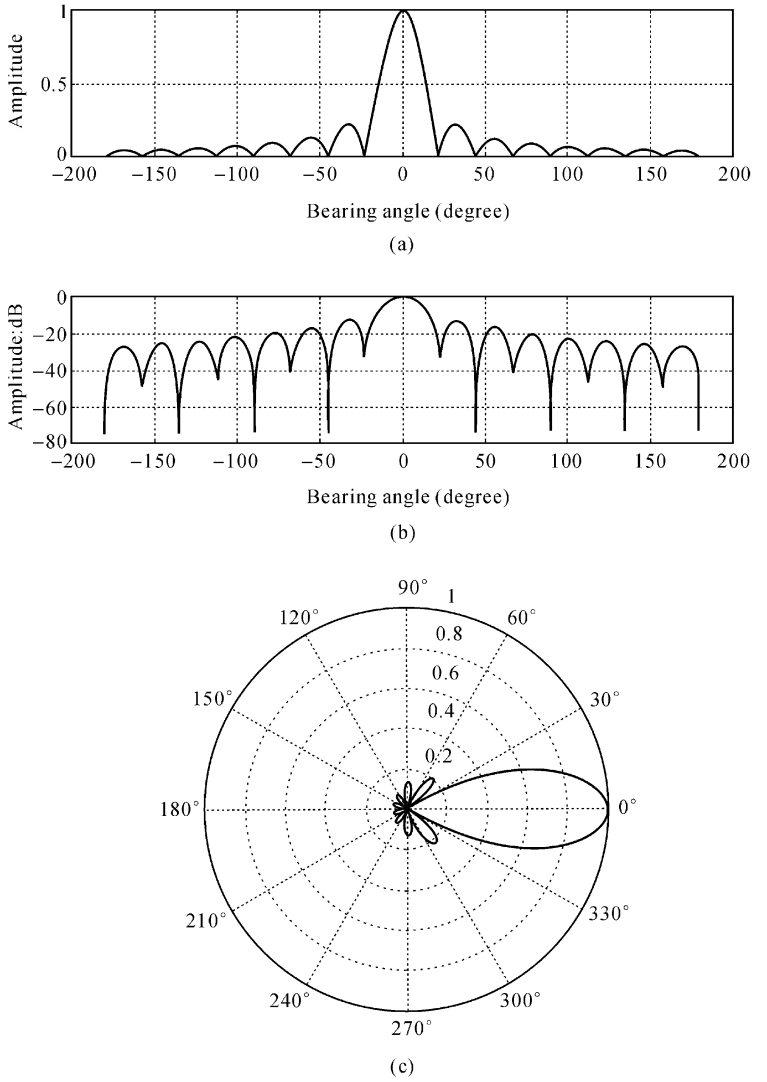


Fig. 4.66. Beam pattern in different coordinate systems. (a) Right angle coordinate; (b) Semi-log coordinate; (c) Polar coordinate

(1) Direction of main lobe

The direction of the maximum value is called the main lobe direction, where $\theta = 0°$.

(2) Beam width of main lobe

The beam width in which the maximum value decreases to the half power value is called the beam width of the main lobe. Let $D(\theta_{3\text{dB}}) = 0.707, \Delta\theta = 2\theta_{3\text{dB}}$ is the value of the beam width of the main lobe. It is also called the 3 dB beam width.

Sometimes, if the directivity has zero points, then the first pair of nearest zero points to the main lobe direction is called the beam width of the main lobe. In the following, unless otherwise stated, the beam width of the main lobe refers to the 3 dB beam width.

(3) Height of side lobe

For the directivity function, except the maximum value, any other local maximum value is called a side lobe. The logarithmic magnitude of the ratio of a side lobe to a main lobe is called the height of the side lobe. For example, if $D(0) = 1$, and $D(\theta_1) = 0.1$ is the maximum side lobe, then the height of the side lobe will be

$$20\lg\frac{D(\theta_1)}{D(0)} = -20 \text{ (dB)}$$

(4) Directivity index

The ratio of 2π over the integration value of the square of $D(\theta)$ is called the directivity index. It represents the ability of an array to suppress isotropic and homogeneous noise.

$$\gamma = \frac{2\pi}{\int_0^{2\pi} D^2(\theta)\mathrm{d}\theta} \tag{4.92}$$

In three-dimensional coordinates, the directivity index should be

$$\mu = \frac{4\pi}{\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} D^2(\theta,\phi)\cos\theta\mathrm{d}\theta\mathrm{d}\phi} \tag{4.93}$$

$DI = 10\lg\gamma$ or $DI = 10\lg\mu$ is the decibel value of the directivity index. As we supposed that the direction function is normalized, $DI \geq 0$. Note that 2π is the length of a unit circle and 4π is the value of a spherical solid angle.

4.7.2 Conventional Beamforming

The main function of beamforming is to steer the sonar system to the target and estimate the incidental angle of the target. When the radiated noise of the target, the signal in which we are interested, propagates to the receive array, because the sound path is different in each element of the array, the signal from each element is different. If we take a suitable time delay for the signal, then the output signals will be "synchronized", and thus will have no time differences relative to each other. This is the basic concept of beamforming.

Suppose an array consists of N elements and the incidental signal is plane wave. Refer to some point; assume the i-th signal is $s(t+\tau_i(\theta_0))$, where θ_0 is the incidental angle of the signal. If we take time delay for this signal, then all signals from the array are the same. Summing up all N signals obtains $N_s(t)$. From the integration of the square value of $N_s(t)$, we obtain $N^2\sigma_s^2$, and σ_s^2 is the power value of the signal. If the incidental angle changes to θ, and the i-th signal is delayed by $\tau_i(\theta_0)$, we obtain $s(t+\tau_i(\theta)-\tau_i(\theta_0))$. The system output should be

$$D(\theta)=\left\{E\left[\left(\sum_{i=1}^{N}s(t+\tau_i(\theta)-\tau_i(\theta_0)\right)^2\right]\right\}^{1/2} \tag{4.94}$$

The normalized result of this function is the directivity function.

Conventional beamforming is called a "delay-sum-square" model. It is adopted by almost all sonar signal processing systems. Why should we take the square and integration? The reason is, as discussed in Chapter 3, that energy detection is the optimum algorithm for detecting Gaussian signal in Gaussian background noise, and the integration operation is the means to obtain the necessary time gain.

It is easy to calculate the system gain for the beamforming schema in Eq. (4.94).

Suppose there is background noise $n_i(t)$ in the i-th signal, and $n_i(t)$ are independent with zero mean. The power of the system output is

$$E[(\sum_{i=1}^{N}s(t+\tau_i(\theta)-\tau_i(\theta_0)))^2]$$

For $\theta=\theta_0$, the value is $N^2\sigma_s^2+N\sigma_n^2$. Therefore, in the case when a sonar system steers to the target, the signal is enhanced N^2 times and noise is enhanced only N times. The output signal-to-noise ratio is $(SNR)_{\text{out}}=N\sigma_s^2/\sigma_n^2$. Note that the input signal-to-noise ratio is $(SNR)_{\text{in}}=\sigma_s^2/\sigma_n^2$. Therefore, beamforming gives a system gain of

$$G_s=\frac{(SNR)_{\text{out}}}{(SNR)_{\text{in}}}=N \tag{4.95}$$

We usually call this the spatial gain of an array system. The logarithm value, which appears in sonar equation, is

$$GS = 10\lg G_s = 10\lg N \tag{4.96}$$

We conclude that the spatial gain of an array with N elements is $10\lg N$, provided that the signal received by the elements satisfies the necessary conditions. The greater the number of elements is, the greater the system gain will be.

Example

Assume the radius of a circle array is 3 m and the length of the circle is 18.84 m. From section 4.3 we know that the spatial correlation radius of an isotopic noise field is the half wave length. Therefore, if the operation frequency is 1 kHz, then the half wave length is 0.75 m. Thus, the element number of this circle array is $18.84/0.75 \approx 25$ and the spatial gain is $10\lg 25 \approx 14$ dB.

The basic principle of conventional beamforming can be understood in a different way. As we know, the system output is

$$y(t) = \sum_{i=1}^{N} s(t + \tau_i(\theta) - \tau_i(\theta_0)) + \sum_{i=1}^{N} n_i(t - \tau_i(\theta_0))$$

When $\theta = \theta_0$

$$y(t) = Ns(t) + \sum_{i=1}^{N} n_i(t - \tau_i(\theta_0))$$

In terms of the large number law described in Chapter 2, we know that $\frac{1}{N}\sum_{i=1}^{N} n_i(t - \tau_i(\theta_0))$ tends to the mean value of $n_i(t)$, because of the independence of $n_i(t)$. Since $E[n_i(t)] = 0$, $(i = 1, \ldots, N)$, $y(t)/N$ will have the limit $s(t)$. Thus, the noise in the output signal of an array system with N elements will be totally suppressed provided the number N is large enough. From Eq. (4.96) we know that $G_s = N$, but this does not mean that the system gain does not depend on the incidental signal. This is because the assumption of noise independence is already inherent in the constraint for signal frequency, i.e., the displacement of elements and the wave length have to satisfy some constraint condition.

Note that in most cases, the value of system gain G_S Is almost equal to the directivity index μ and γ. Therefore, sometimes we replace GS by DI in the sonar equation.

4.7.3 Equal-Spaced Line Array

Among all sonar arrays, the equal-spaced line array is the simplest and most basic array shape. Humans have the ability to find the bearings of a sound source. When we want to define the incidental angle of a sound source, we turn our head to make the two ears "steer" the source, i.e., we align the sound source in the perpendicular bisector of our two ears. This operation is actually the maximum value steering method in source bearing.

Fig. 4.67 illustrates an equal-spaced line array. From left to right, the hydrophones are numbered H_1, …, H_N. The distance between adjacent elements is d. For convenience of calculation, the reference point is chosen at point H_1. Suppose the incidental signal is sinusoidal signal $A\cos 2\pi ft$. The incidental angle is θ. Therefore, the difference of the sound path of the i-th element is:

$$H_iP_i = (i-1)\sin\theta \tag{4.97}$$

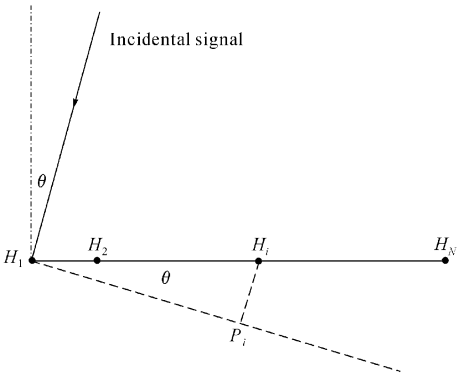


Fig. 4.67. Equal-spaced linear array

The signal received by H_i is

$$s_i(t) = A\cos\left\{2\pi f\left[t + \frac{(i-1)\mathrm{d}\sin\theta}{c}\right]\right\}$$

where c is the sound velocity. The phase difference between H_i and H_1 is

$$\varphi_i = 2\pi(i-1)\frac{\mathrm{d}\sin\theta}{\lambda} = (i-1)\psi$$

where $\lambda = c/f$ is the wave length of the signal. Summation of $s_i(t)$ gives

$$s(t) = \sum_{i=1}^{N} s_i(t) = \sum_{i=1}^{N} A\cos[2\pi ft + (i-1)\psi] \tag{4.98}$$

By using the following equality of triangle function

$$\sum_{i=1}^{N} \cos[\alpha + (i-1)\beta] = \frac{\cos(\alpha + (N-1)\beta/2)\sin(N\beta/2)}{\sin(\beta/2)}$$

we have

$$s(t) = A\cos[2\pi ft + (N-1)\psi/2]\frac{\sin(N\psi/2)}{\sin(\psi/2)}$$

In terms of the definition of the directivity function

$$D(\theta) = \left|\frac{\sin(N\pi d\sin\theta/\lambda)}{N\sin(\pi d\sin\theta/\lambda)}\right| \tag{4.99}$$

where N in the denominator is a normalization factor, i.e., $D(0) = 1$.

Similarly, if we want to steer in the direction θ_0, it is necessary to take time delay for the i-th signal by

$$\tau_i(\theta_0) = 2\pi(i-1)\mathrm{d}\sin\theta_0/\lambda \tag{4.100}$$

In this situation, it is proved that the directivity function is

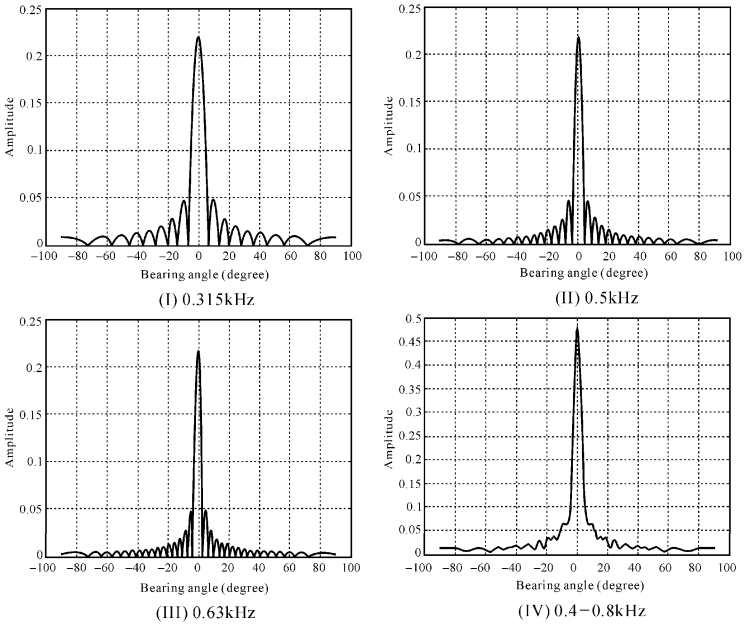
$$D(\theta) = \left|\frac{\sin[N\pi d(\sin\theta - \sin\theta_0)/\lambda]}{N\sin[\pi d(\sin\theta - \sin\theta_0)/\lambda]}\right| \tag{4.101}$$

From the above discussion we know that the directivity function of an equal-spaced line array has the form $\sin(Nx)/(Nx)$ and is easy to calculate. If $\theta_0 = 0$, then $D(\theta)$ is symmetrical to $\theta = 0$. This direction is called the broadside direction. In the case of $0 < \theta_0 < \pi/2$, $D(\theta)$ is not symmetrical to $\theta = 0$. In the extreme situation, $\theta_0 = \pi/2$, the line array is degraded to one point called the end fire direction.

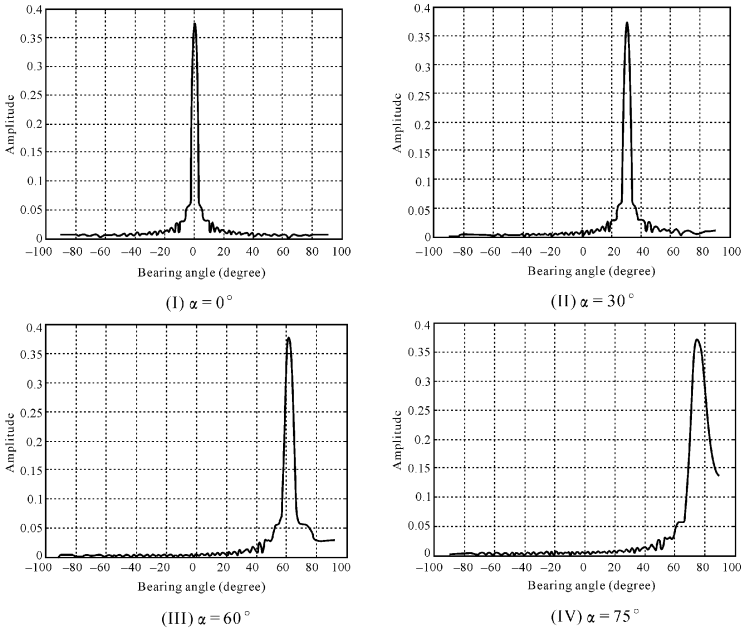
A series of examples of directivity functions are illustrated in Figs. 4.68(a) and (b). Fig. 4.68(a) presents some directivity functions where the incidental angle is in the broadside direction. The simulation signal is a sinusoidal signal or broadband Gaussian signal. Fig. 4.68(b) presents scenarios in which the incidental angle of the signal is in different directions. As we can see, when the target bias is in the broadside direction, the main lobe of directivity becomes wider.

From Eq. (4.100) we see that d/λ is an important parameter for an equal-spaced line array because for the same d/λ, the directivity is the same.

If the adjacent distance of line array d is much less than the wave length of incidental signal λ, i.e., $d << \lambda$, then the discrete line array becomes a continuous array. The directivity function of a continuous line array can be derived from Eq. (4.101).



(a)



(b)

Fig. 4.68 Examples of beampatlern of linear array. (a) Beam pattern of broad side for a linear array ($N = 80$, $d = 1$ m); Beam pattern of a linear array with different bearing angles ($N = 80$, $d =$ 1 m, 0.6 – 1.0 kHz)

Actually, the length of a line array is $L = (N-1)d$. In Eq. (4.101), letting $N \to \infty$, $d \to 0$ and keeping L constant, we obtain

$$\frac{\sin[N\pi d(\sin\theta - \sin\theta_0)/\lambda]}{N\sin[\pi d(\sin\theta - \sin\theta_0)/\lambda]} \to \frac{\sin[L\pi(\sin\theta - \sin\theta_0)/\lambda]}{L\pi(\sin\theta - \sin\theta_0)/\lambda}$$

Therefore, the directivity function of a line array is

$$D(\theta) = \left| \frac{\sin[L\pi(\sin\theta - \sin\theta_0)/\lambda]}{L\pi(\sin\theta - \sin\theta_0)/\lambda} \right| \tag{4.102}$$

This is $\sin x / x$ form function.

Theoretical analysis shows that for $d/\lambda = 0.5$ and $N \geq 32$ the differences in directivity function between a discrete line array and a continuous line array are very small (less than 0.05). Fig. 4.69 shows the directivity functions of a discrete line array and a continuous line array, where, N = 32, $d/\lambda = 0.5$ and L/λ = 15.5 (L is the length of the continuous line array).

It is necessary to show that the directivity function of a line array strongly depends on the parameter d/λ. From Eq. (4.100) we know that in some cases, except the maximum value when $\theta = \theta_0$, the directivity function may be the second or even third maximum value, if the value d/λ satisfies some conditions. Fig. 4.70 presents some examples. In the first example, where d = 1 m, f = 0.8 kHz ($d/\lambda = 0.625$) and the incidental angle is 60°, a large side lobe can be seen at –70°. The second and third examples are similar. In sonar design, these situations must be avoided, so it is necessary to choose the parameters d/λ, N and L/λ very carefully.

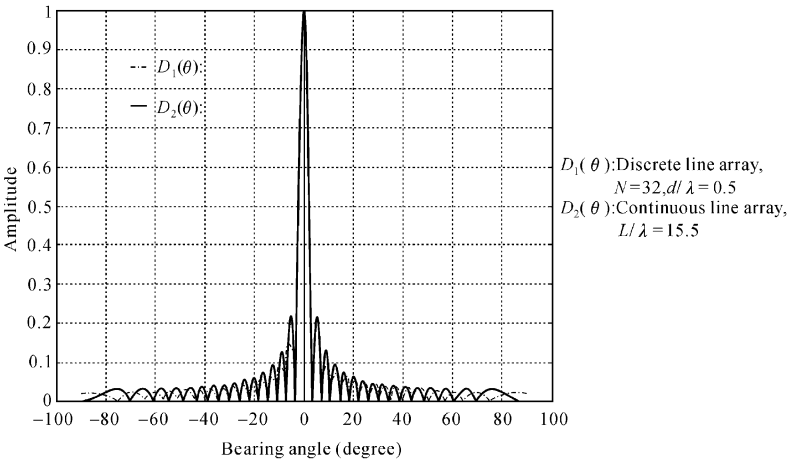


Fig. 4.69. Comparison of the directivity functions of discrete and continuous line arrays

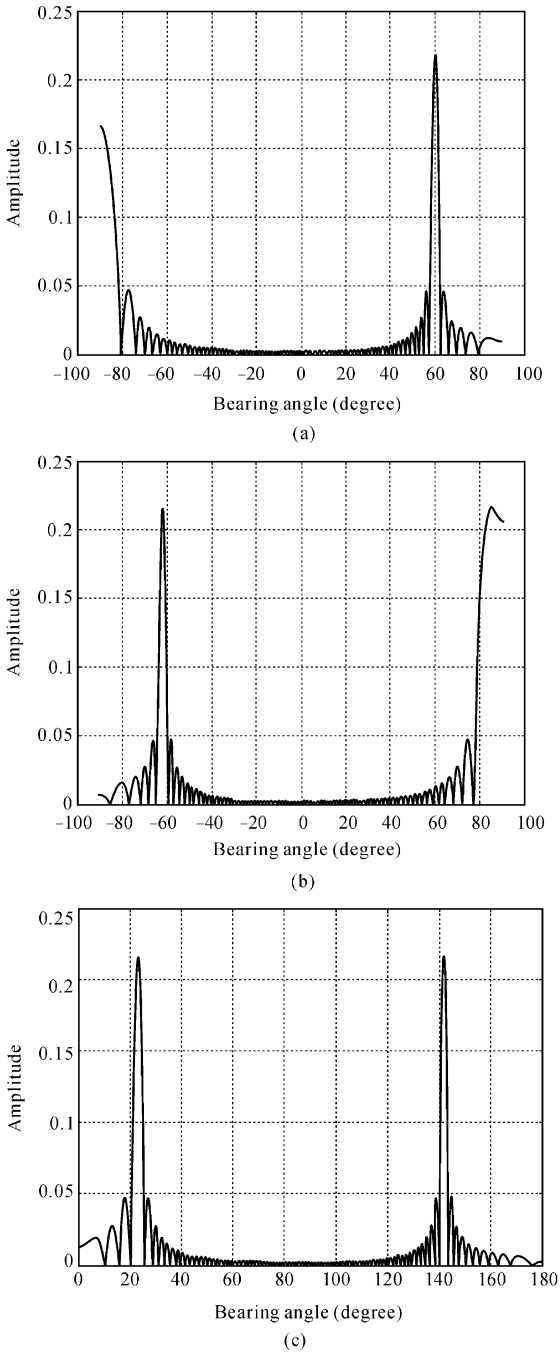


Fig. 4.70. Abnormal phenomena of linear array beam pattern ($N = 80$, $d = 1$ m). (a) $f = 0.8$ kHz, $\alpha = 60°$; (b) $f = 0.8$ kHz, $\alpha = 85°$; (c) $f = 0.88$ kHz, $\alpha = 23°$

Now we are going to study the main numerical characteristics of the directivity function for a line array. The analysis is based on the continuous line array but, as noted above, it is easy to change the conclusion to suit a discrete equal-spaced line array.

(1) Height of side lobe

It is easy to calculate that $|\sin x / x|$ has a second local maximum value 0.22 at $x = 4.50$, and therefore the height of the side lobe is 22%. The position of the second maxima of the directivity function of broadside direction can be calculated using the following equation:

If $L\pi \sin\theta / \lambda = 4.50$, we obtains

$$\theta = \arcsin\left(\frac{4.50\lambda}{\pi L}\right) = \arcsin(1.43\lambda / L) \tag{4.103}$$

If $\lambda / L << 1$, then

$$\theta \approx 1.43\lambda / L \tag{4.104}$$

(2) Beam width of the main lobe

Let the right hand side of Eq. (4.102) be equal to 0.707, i.e.,

$$\frac{\sin[L\pi(\sin\theta - \sin\theta_0)/\lambda]}{L\pi(\sin\theta - \sin\theta_0)/\lambda} = 0.707$$

It is easy to show that the solution of $\sin x / x = 0.707$ is $x = 1.39$. Therefore $L\pi(\sin\theta - \sin\theta_0)/\lambda = 1.39$, or

$$\sin\theta - \sin\theta_0 = 0.44\lambda / L \tag{4.105}$$

We can solve θ from this equation, and find that $\theta_{3\text{dB}} = 2(\theta - \theta_0)$. It is proved that, in the case of $L/\lambda >> 1$, if $\theta_0 \approx 0$ (broadside direction), we obtain

$$\theta_{3\text{dB}} \approx 0.88\frac{\lambda}{L}\frac{1}{\cos\theta_0} \tag{4.106a}$$

And if $\theta_0 \approx \pi/2$ (end fire direction)

$$\theta_{3\text{dB}} \approx 2(0.88\lambda / L)^{1/2} \tag{4.106b}$$

(3) Directivity index

Based on Eq. (4.93), the directivity index of a line array is

$$\mu = \left[\frac{1}{4\pi} \int_0^{2\pi} \mathrm{d}\phi \int_{-\pi/2}^{\pi/2} \frac{\sin^2(\pi L \sin\theta / \lambda)}{(\pi L \sin\theta / \lambda)^2} \cos\theta \mathrm{d}\theta \right]^{-1}$$

$$= \left\{ \frac{\lambda}{\pi L} [s_i(2\pi L / \lambda) - \lambda \sin^2(\pi L / \lambda) / \pi L] \right\}^{-1}$$

where $s_i(x) = \int_0^x \frac{\sin u}{u} \mathrm{d}u$. For $\lambda / L << 1$

$$\mu \approx 2L / \lambda \tag{4.107}$$

Note that in the case of $d / \lambda = 0.5$, the directivity index is equal to the spatial gain described in Eqs. (4.95) and (4.96). Actually, in Eq. (4.107), if $L = (N-1)d$, and $d / \lambda = 0.5$, we get $\mu = N - 1$. This is almost the same value as with G_s. It is important to remember the condition of $d / \lambda = 0.5$, as it is useful in digital sonar design.

Note that the directivity function for a line array has a very regular distribution of zero points. Sometimes, the width of the first pair of zero points is defined as the beam width of the main lobe.

Simple calculation shows that, in broadside direction, if $\lambda / L << 1$, then

$$\Delta\theta \approx 2\lambda / L \tag{4.108a}$$

In the end fire direction

$$\Delta\theta \approx 2\sqrt{2\lambda / L} \tag{4.108b}$$

4.7.4 *Uniformly Distributed Discrete Circle Array*

Fig. 4.71 presents a discrete circle array with N uniformly distributed elements. Assume the radius of this array is r. Arrange the hydrophones clockwise as H_1, ..., H_N. The line connecting the center of the circle and hydrophone H_1 has the direction of 0° and the angle of adjacent elements should be $\alpha = 2\pi / N$.

In calculating the directivity function, the center of the circle is chosen as the reference point of time delay. Suppose the incidental angle of the target signal is θ. The signal received at point O is supposed to be $A\cos(2\pi ft)$. The signal received by i-th hydrophone H_i is

$$s_i(t) = A\cos[2\pi f(t + \tau_i(\theta))]$$

where $\tau_i(\theta)$ is the relative time delay of H_i and O, i.e.,

$$\tau_i(\theta) = r\cos[\theta-(i-1)\alpha]/c \quad (i=1,\ ...,\ N) \tag{4.109}$$

where c is the sound velocity in the ocean.

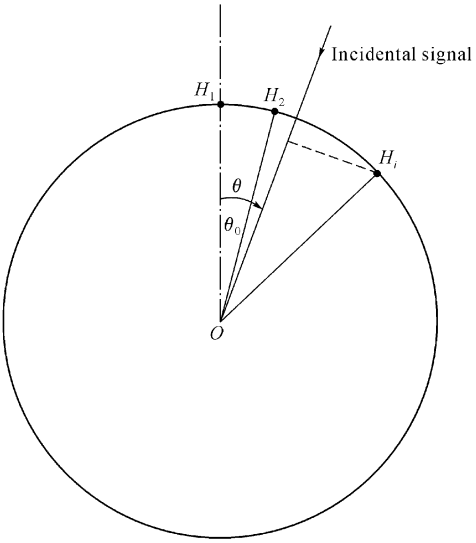


Fig. 4.71. Calculation of time delay for a uniformly distributed circle array

If we use the wave length of signal $\lambda = c/f$, then

$$s_i(t) = A\cos[2\pi ft + \varphi_i]$$
$$\phi_i = 2\pi r\cos[\theta-(i-1)\alpha]/\lambda\text{'}$$

To steer the array in the direction θ_0, it is necessary to make a time delay $\tau_i(\theta_0)$ for the received signal $s_i(t)$. Now the i-th signal, after time delay becomes

$$s_i(t-\tau_i(\theta_0)) = A\cos[2\pi f(t+\tau_i(\theta)-\tau_i(\theta_0))] = A\cos[2\pi f(t+\Delta_i(\theta))]$$

where $\Delta_i(\theta) = \tau_i(\theta)-\tau_i(\theta_0)$.
Taking the summation of N element delayed signals, obtains

$$s(t) = \sum_{i=1}^{N} s_i(t-\tau_i(\theta_0)) = \sum_{i=1}^{N} A\cos[2\pi f(t+\Delta_i(\theta))] \tag{4.110}$$

To calculate the directivity function of a circle array, it is necessary to find the value of $E[s^2(t)]$. Expending each term in Eq. (4.110) and combining the same class terms obtains

$$\begin{aligned} D(\theta) &= \frac{1}{N}[E(s^2(t))]^{1/2} \\ &= \frac{1}{2}\left[\left(\sum_{i=1}^{N}\cos 2\pi f \Delta_i(\theta)\right)^2 + \left(\sum_{i=1}^{N}\sin 2\pi f \Delta_i(\theta)\right)^2\right]^{1/2} \end{aligned} \tag{4.111}$$

This is the final expression of the directivity function of a circle array. Generally, the value of $D(\theta)$ depends on $\Delta_i(\theta)$, which is quite complicated. Therefore, there are no closed forms of expression of directivity function for a circle array. The numerical calculation can show the results of the directivity function.

If the parameters of a circle array satisfy the following condition, then the directivity function of the array can be approximately expressed as a Bessel function.

Denote the arc length of adjacent elements by s_0, i.e., $s_0 = 2\pi r / N$. If

$$N \geq \frac{4\pi r}{\lambda} + 2 \quad \text{or} \quad \frac{s_0}{r} < \frac{1}{2} - \frac{1}{N} \tag{4.112}$$

Then

$$D(\theta) \approx \left| J_0\left(\frac{4\pi r}{\lambda}\sin\frac{\theta - \theta_0}{2}\right)\right| \tag{4.113}$$

where $J_0(\cdot)$ is the zero order Bessel function:

$$J_0(x) = \sum_{k=0}^{\infty}(-1)^k\left[\left(\frac{x}{2}\right)^{2k} \Big/ (k!)^2\right]$$

For a circle array applied in a sonar system, if the operation frequency is narrow band, then Eq. (4.112) usually can be satisfied. But, if the sonar system works in a broad frequency band, sometimes at the high frequency end, Eq. (4.112) is hard to satisfy. It is also possible to write Eq. (4.112) in terms of frequency f

$$f \leq 0.24(N/2 - 1)/r \tag{4.114}$$

where the unit of f is kHz and the unit r is m.

Unlike for a line array, generally, the directivity function for a circle array is quite uniform. Particularly when Eq. (4.112) is satisfied, the directivity in all directions is similar. This result can be found in Eq. (4.113), where the directivity function depends only on $\theta - \theta_0$.

In the following, suppose Eq. (4.112) is satisfied and the main characteristics of the directivity function of a circle array are analyzed.

(1) Height of side lobe

The second maxima of $J_0(x)$ is 0.40, therefore, the height of the side lobe is 40% (about –8 dB).

(2) Beam width of main lobe

Based on Eq. (4.113), one should let

$$J_0\left(\frac{4\pi r}{\lambda}\sin\frac{\theta-\theta_0}{2}\right)=0.707$$

It is well known that $J_0(x)=0.7073,$ when $x=1.126$. Therefore

$$\frac{4\pi r}{\lambda}\sin\frac{\theta-\theta_0}{2}=1.126$$

In the case of $r/\lambda >> 1$,

$$\theta_{3\mathrm{dB}}=2\times(\theta-\theta_0)\approx 1.12\lambda/(\pi r) \tag{4.115}$$

By comparison with Eq. (4.105), we can see that πr is the aperture of the circle array. Note that the unit of the right hand side of Eq. (4.115) is the radian.

(3) Directivity index

The calculation of the spatial directivity index is quite complicated. For example, the plane directivity index is

$$\nu=\frac{1}{\pi}\int_0^{\pi}\left(J_0\left(\frac{4\pi r}{\lambda}\sin\frac{\theta}{2}\right)\right)^2\mathrm{d}\theta$$

$$DI=10\lg\nu$$

The circle array described above is extensively used in sonar systems, but in some cases the elements used in beamforming vary depending on the practical situation. Sometimes, it is not necessary to have N elements of signal in beamforming. For example, in a bow array in a submarine in which the bearing beam around 0° degrees steers straight ahead of the submarine, it is not necessary to add the signal from the back direction of the submarine, as this would add excessive noise and interference that would reduce the sonar performance. An ideal method is to baffle the circle array, such that each hydrophone has some kind

of directivity instead of an omni-directional function. For a circle array with baffling, the beamforming elements comprise an arc array.

4.7.5 Circle Array Baffling and Arc Array

A circle array used in a submarine or surface ship is installed in the bow of the vessel. Since propeller noise comes from the aft of a vessel, it is the main source of interference in sonar operation. Therefore, it is necessary to reduce the noise from the aft direction. An effective method is to baffle the circle array such that a part of the noise from the back direction is reduced. Also, when we use beamforming technique, we choose only those elements that are around the steering direction. Thus, a bow sonar consists of a multiple beam system; each beam collects only a part of the elements in a circle array.

The baffling technique is used not only in circle arrays, but also in other kinds of sonar arrays, for example, the flank array and side scan array. The directivity function of a baffled array is different from the original array and depends on the method of baffling.

Array baffling technique is a very important topic in sonar design. Array baffling can be of the sound absorption type or the sound reflection type, depending on the frequency in which we are interested and the possibility of manufacture. But, in this book, we are not going to study the mechanism of array baffling. We are concerned only with the characteristics of a baffling array from a signal processing point of view.

The directivity function of an array consisting of omni-directional hydrophones was discussed above. Once the array is baffled, the directivity function of each hydrophone in the array is no longer omni-directional, therefore, the synthetic beam pattern of the array must be re-calculated.

In the following, we will firstly study the effect of baffle on the directivity function of hydrophones. In Fig. 4.72, the distance of hydrophone H to the baffler is α. Assume the aperture of the baffler is much larger than the signal wave length λ.

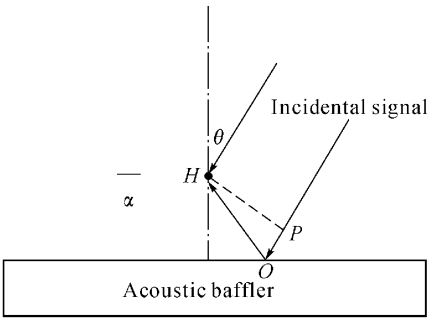


Fig. 4.72. Illustration of the role of an acoustic baffler

Suppose the received direct arrival signal at H is $A\cos(2\pi ft)$ and the reflect coefficient of the baffler is α. Then the reflection wave is $A\alpha\cos(2\pi ft-\delta)$, where δ is the phase difference resulting from the reflection. This consists of two parts, one resulting from the sound path difference $PQ+QH$, and the other resulting from the acoustic characteristics of the baffler. It is easy to calculate the first part, i.e.,

$$\delta_1 = 2\pi f\frac{a}{c\cos\theta}(1+\cos 2\theta)$$

where c is the sound velocity. The second part, denoted by δ_2, is related to the incidental angle θ of the signal. But, in a certain frequency range, δ_2 can be considered as a constant. Therefore

$$\delta = 2\pi f\frac{a}{c\cos\theta}(1+\cos 2\theta)+\delta_2$$

The reflection wave is

$$A\cos\left[2\pi ft-\delta_2-2\pi f\frac{a}{c\cos\theta}(1+\cos 2\theta)\right] \tag{4.116}$$

The synthetic wave received by hydrophone H is

$$u(t) = A\cos 2\pi ft + A\alpha\cos(2\pi ft-\delta)$$

We can rewrite this expression in a more compact form:

$$u(t) = p\cos(2\pi ft-\gamma) \tag{4.117}$$

where

$$p = A[1+2\alpha\cos\delta+\alpha^2]^{1/2}\quad \gamma = \arctan[(\alpha\sin\delta)/(1+\alpha\cos\delta)] \tag{4.118}$$

From Eqs. (4.117) and (4.118) we can see that, when the incidental angle θ of the signal is varied, the amplitude response of the received signal will also vary. Therefore, the directivity function of a single hydrophone in the array will vary with the incidental angle of the signal.

The distance a from the hydrophone to the baffler can be adjusted. In the case of so called "hard reflection", i.e., the reflection wave and incidental wave have the same phase: $\delta_2 = 0$; when a is small enough, then the reflection wave will

have the same phase as the incidental wave and, therefore, the sensitivity of the hydrophone will increase. Thus, placing the hydrophone close to the baffler will give some advantage in sensitivity.

In the case of so-called "soft reflection", i.e., where the reflection wave has 180° phase difference compared with the incident wave, when $\delta_2 = \pi$, the reflection wave will cancel the incidental wave, and so it is not appropriate to place the hydrophone close to the baffler.

Careful analysis of Eq. (4.115) shows that the directivity function of a hydrophone has some kind of transforming relationship with the sensitivity. In fact, when $\theta = 0$, a change in f obtains the frequency response of the hydrophone; for a fixed f, when we change θ, we get the directivity function. A change in f and θ will result in a change of $f \propto (1+\cos 2\theta)/\cos\theta$. This conclusion seems very simple, but it is an important reference fact in wet end design of sonar because, based on this fact, it can be seen that the frequency response curve approximately derives the directivity function and vice versa.

Obviously, $(1+\cos 2\theta)/\cos\theta$ is the monotonic decreasing function of θ. When θ changes from small to large, $(1+\cos 2\theta)/\cos\theta$ changes from large to small, and correspondingly f changes from large to small. Therefore, to obtain a cardinal directivity function, it is necessary to design a baffler with a flat or positive gradient frequency response.

The calculation of the directivity function for a baffled array becomes complicated compared to that for an array with no baffler. But, for a line array, the following Bridge product theorem explores an important fact in directivity calculation.

4.7.6 Product Theorem of Directivity Function of a Line Array

The directivity function of a baffled array is equal to the product of the directivity function of a line array without baffling and the directivity function of a single hydrophone.

To prove this theorem, suppose the directivity function of a single hydrophone with baffling is $\mathrm{d}(\theta)$. The signal received by the i-th hydrophone is $\mathrm{d}(\theta)\cos(2\pi ft+\varphi_i)$. The total array response should be

$$
\begin{aligned}
D_1(\theta) &= \left\{ E\left[\sum_{i=1}^{N} \mathrm{d}(\theta)\cos(2\pi ft+\varphi_i) \right]^2 \right\}^{1/2} \\
&= d(\theta)\left\{ E\left[\sum_{i=1}^{N} \cos(2\pi ft+\varphi_i) \right]^2 \right\}^{1/2} \\
&= d(\theta)D(\theta)
\end{aligned}
$$

where $D(\theta)$ is simply the directivity function without baffling.

For an arc array, the calculation of the directivity function is more complicated because there is no simple theorem like the product theorem for a line array. Each single hydrophone of an arc array has its own directivity function, as the direction of maximum value (i.e., the direction of the sound axis) differs between hydrophones. The angle of the sound axis between two adjacent elements is equal to $\alpha = 2\pi / N$. If the incidental angle of the signal is θ, the signal received by the i-th element H_i is

$$A\xi(\theta - (i-1)\alpha)\cos(2\pi ft + \phi_i)$$

where $\xi(\theta)$ is assumed to be the directivity function of a single hydrophone of an arc array. Fig. 4.73 illustrates an arc array. If the steering direction of beamforming is in the middle of two adjacent elements, the number of working elements is $2M$.

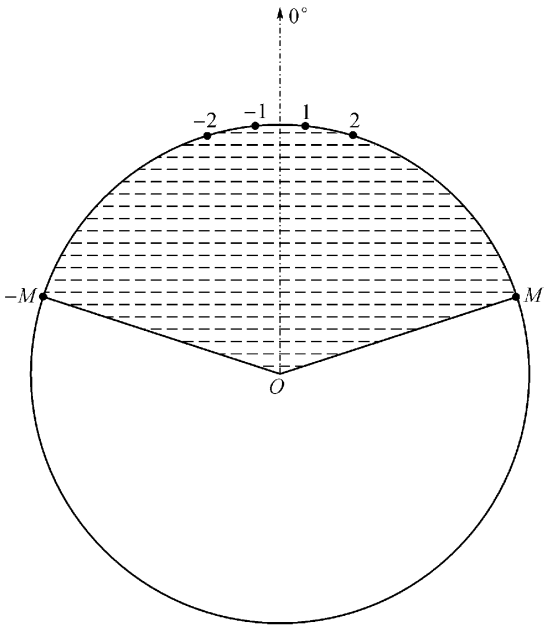


Fig. 4.73. Illustration of the working sector of a circle array

The ideal directivity function of a single hydrophone is shown in Fig. 4.74. This is a symmetrical cardinal curve with a certain beam width to guarantee a sufficiently large working section of the arc array. Clearly, the direction function of a single hydrophone in a baffling array is related to the frequency. It is very important to design a baffler for which the directivity function is good enough over a wide frequency range. Here is an example of the directivity function of a single hydrophone.

$$\xi(\theta) = \exp[-(\theta / 90^\circ)^2] \tag{4.119}$$

The unit of θ in Eq. (4.119) is degree.

When the working sector includes $2M$ elements, the directivity function is

$$D(\theta) = \frac{1}{2M}\{[\sum_{i=1}^{M}(a_i(\theta)\cos\Delta_i(\theta) + a_i^{'}(\theta)\cos\Delta_i^{'}(\theta))]^2$$

$$+[\sum_{i=1}^{M}(a_i(\theta)\sin\Delta_i(\theta) + a_i^{'}(\theta)\sin\Delta_i^{'}(\theta))]^2\}^{1/2} \tag{4.120}$$

where $\Delta_i(\theta) = \tau_i(\theta) - \tau_i(\theta_0)$, $\Delta_i^{'}(\theta) = \tau_i^{'}(\theta) - \tau_i^{'}(\theta_0)$,

$$\tau_i(\theta) = r\cos[\theta - (i-0.5)\alpha]/c\ ,\ \tau_i^{'}(\theta) = r\cos[\theta + (i-0.5)\alpha]/c$$

and $a_i(\theta) = \xi(\theta - (i-0.5)\alpha)$, $a_i^{'}(\theta) = \xi(\theta + (i-0.5)\alpha)$, $i = 1,\ ...,\ M$

If the working sector includes $2M$+1 elements, then the beam pattern of this arc array should be

$$D(\theta) = \frac{1}{2M+1}\left\{\left[\sum_{i=-M}^{M} a_i(\theta)\cos\Delta_i(\theta)\right]^2 + \left[\sum_{i=-M}^{M} a_i(\theta)\sin\Delta_i(\theta)\right]^2\right\}^{1/2} \tag{4.121}$$

where $\Delta_i(\theta) = \tau_i(\theta) - \tau_i(\theta_0)$, $\tau_i(\theta) = r\cos[\theta - i\alpha]/c$ $i = -M,\ ...,\ 0,\ ...,\ M$.

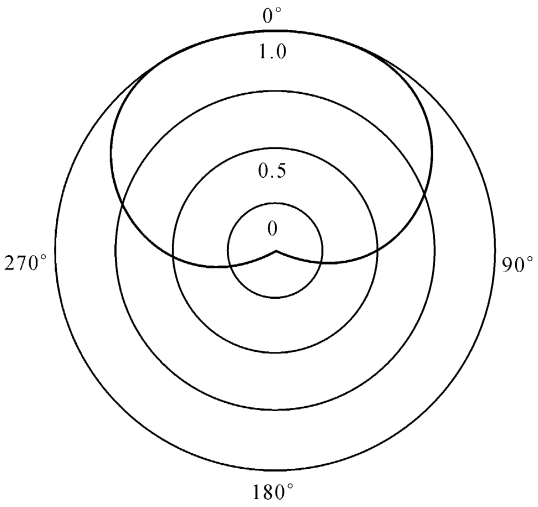


Fig. 4.74. Ideal directivity of a single hydrophone in a circle array

Fig. 4.75 shows an example of an arc array with a total of 32 hydrophones. The working sector is 2×7 (i.e., $M = 7$).

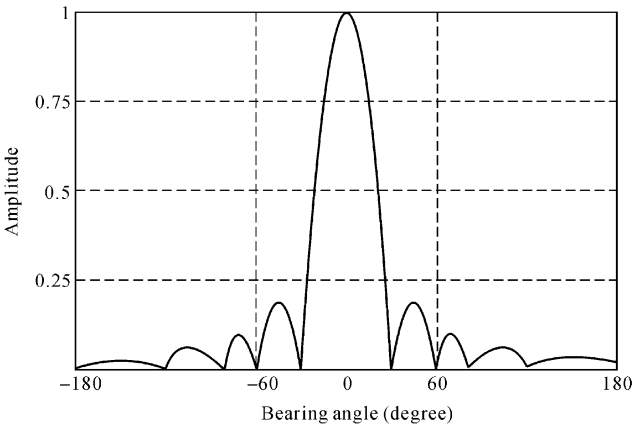


Fig. 4.75. A typical beam pattern of a linear array with a baffler ($N = 32$, $M = 14$)

A successful baffler design should have a cardinal directivity function for each hydrophone in the array. If the directivity function is not cardinal, or even has the shape of a butterfly, then the beam pattern of the array will have some problems and the performance of the bearing and the target resolution will be degraded.

Table 4.10 shows the numerical results from a baffling circle array with 32 elements.

Table 4.10 Beam pattern of a baffling array ($N = 32$)

Frequency (kHz)	Number of elements	Beam width of main lobe (degree)	Height of side lobe (dB)
5.0	12	10	-12
5.0	14	9	-12
5.0	16	8.2	-11
White noise 2 – 10	12	10.0	-13.2
Radiated noise 2 – 10	12	20.0	-14.9

As noted above, the height of the side lobe of a circle array is about 40%. However, for a baffling array, the height of the side lobe can usually be improved, to even less than 10%.

4.7.7 Weighting of an Array

Weighting is a method for improving the performance of the beam pattern of an array. The basic algorithm for weighting is to multiply a weight coefficient by the output signal of each array element. The concept of weighting is totally different

from baffling. This is because, unlike baffling, the weighting coefficient is independent of the incidental angle of the signal.

Suppose an array has N elements and the i-th received signal is $x_i(t)$. Before weighting, the output of beamforming is

$$y(t) = \sum_{i=1}^{N} x_i(t - \tau_i(\theta_0)) \tag{4.122}$$

If we multiply a weight coefficient w_i by the i-th received signal then, after weighting, the output of the signal is

$$y(t) = \sum_{i=1}^{N} w_i x_i(t - \tau_i(\theta_0))$$

where, $x_i(t) = x(t + \tau_i(\theta))$ (Fig. 4.76).

The synthetic directivity function of a baffling and weighting array is

$$D(\theta, \theta_0) = \left\{ E\left[\sum_{i=1}^{N} w_i a_i(\theta) x(t + \tau_i(\theta) - \tau_i(\theta_0)) \right]^2 \right\}^{1/2} \tag{4.123}$$

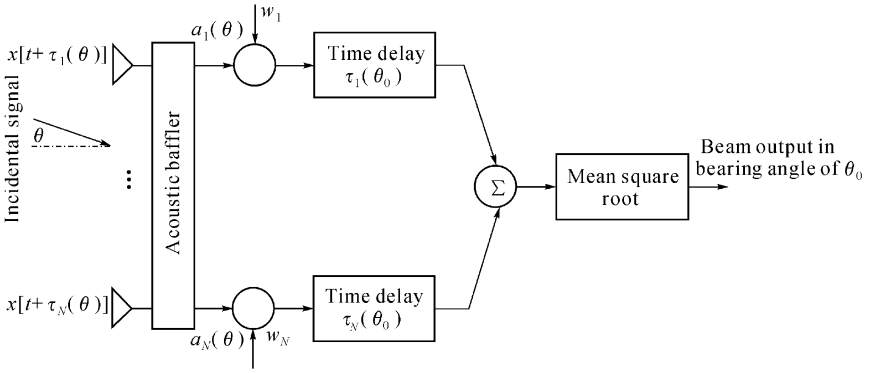


Fig. 4.76. Conventional beamforming with baffling and weighting

The aim of weighting is to improve the beam pattern or to suppress some specific interferences. Several weighting algorithms are used in sonar design.

(1) Dolph-Chebychev weighting of a discrete line array

In deducing the directivity function of a discrete line array, we know that the value of directivity is the summation of some kind of function $\cos k\varphi$, where $\varphi = (\pi d / \lambda)\sin\theta$. Therefore, the directivity function of a line array, after weighting, should have the form

$$y = \sum_k A_k \cos k\varphi$$

From the theory of mathematical analysis, we know that for any k, $\cos(k\phi)$ can be expended as the power series of $\cos\phi$, the so-called Chebychev polynomial. Thus, the directivity function has the form

$$y = \sum_k A_k \cos k\varphi = \sum_n B_n (\cos\varphi)^n \tag{4.124}$$

The relation, between A_k and B_n can be found using the Chebychev polynomial, i.e.

$$\begin{aligned} \cos(2k+1)\varphi &= \sum_{m=0}^{k} c_{2m+1} (\cos\varphi)^{2m+1} \\ \cos(2k\varphi) &= \sum_{m=0}^{k} c_{2m} (\cos\varphi)^{2m} \end{aligned} \tag{4.125}$$

where

$$c_{2m+1} = (-1)^{k-m} \sum_{p=k-m}^{k} \binom{p}{p-k+m} \binom{2k+1}{p}$$

$$c_{2m} = (-1)^{k-m} \sum_{p=k-m}^{k} \binom{p}{p-k+m} \binom{2k}{2p}$$

In Eq. (4.124), if we consider $x = \cos\varphi$ as a variable, then $|x| \leq 1$ and

$$y = \sum_n B_n x^n$$

Therefore, the directivity function of a line array after weighting is a polynomial in segment [–1, 1], and the order of this polynomial is defined by the number of elements in the array. Based on the approximation theory, among all polynomials in segment [–1, 1] with order N and coefficient 1.0 of maximum power term, the Chebychev polynomial has maximum value. Dolph applied this theory to line array weighting and derived two important conclusions[36]:

1) For any given height of a side lobe, the Chebychev weighting has the narrowest main lobe width (where the width of the main lobe means the width of the first pair of zero points).
2) For any given width of a main lobe, the Chebychev weighting has the lowest height of the side lobe.

Since the mathematical deduction of Dolph-Chebychev weighting is complicated, we will not prove it here. As an example, a six-element line array with Dolph-Chebychev weighting is presented. The requirement of the height of the side lobe is –30 dB. It is easy to see that the ratio of the main lobe to the side lobe is $r = 31.6$, since $20 \lg r = 30$.

The directivity function is

$$\begin{aligned} F(\varphi) &= A_1 \cos\varphi + A_2 \cos 3\varphi + A_3 \cos 5\varphi \\ &= 16A_3 x^5 + (4A_2 - 20A_3)x^3 + (A_1 - 3A_2 + 5A_3)x \\ &= G(x) \end{aligned}$$

we use the Chebychev polynomial Eq. (4.125) to expend $\cos(3\varphi)$, $\cos(5\varphi)$ as the function of $\cos\varphi$. Based on the principle of Chebychev weighting, it is necessary to force the coefficient of the 5-th term to equal 1. Denoting it by $T_5(z_0 x)$, we have

$$T_5(z_0 x) = 16(z_0 x)^5 - 20(z_0 x)^3 + 5z_0 x$$

Letting $T_5(z_0 x) = G(x)$, we obtain

$$16A_3 = 16z_0^5,\ 4A_2 - 20A_3 = -20z_0^3,\ A_1 - 3A_2 + 5A_3 = 5z_0$$

The next step is to find z_0 in terms of r. When $\theta = 0$, $F(\varphi)$ reaches the maximum value 1; for $\varphi = 0$, we have $x = 1$. Therefore, $T_5(z_0) = 31.6$, and we get $z_0 = 1.35$. Now we have $A_1 = 15.60$, $A_2 = 10.70$, $A_3 = 4.66$.

The closed form formula for calculating the Chebychev weighting has already been published[39]. For an $N = 2M$ element, equal spaced line array, from left to right the coefficients are indicated by $A_M / 2$, ..., $A_1 / 2$, $A_1 / 2$, ..., $A_M / 2$; for a $N = 2M + 1$ element array, from left to right the coefficients are indicated by $A_M / 2$, ..., $A_1 / 2$, A_0, $A_1 / 2$, ..., $A_M / 2$. We have

$$\begin{aligned} &A_M = z_0^{2M-1}, \\ &A_{M-m} = \sum_{p=0}^{m-1} \frac{(N-1)(m-1)!(N-p-2)!}{(m-p)!\,p!(m-p-1)!(N-m-1)!} z_0^{N-(2m+1)} (z_0^2 - 1)^{m-p} \end{aligned} \quad (4.126)$$

($m = 1,\ ...,\ M$ and $T_M(z_0)$ = main lobe /side lobe)

Note that Chebychev weighting is an equal side lobe weighting, as determined by the characteristics of the Chebychev polynomial. In fact, it is impossible to obtain a perfect weighting which has the narrowest width of the main lobe and the lowest height of the side lobe. This fact has not yet been proved but Elliot has at least shown that if one narrows the main lobe, then the side lobe must become higher; in contrast, if one suppresses the side lobe, then the main lobe must become wider[39].

Fig. 4.77 illustrates the relationship between the width of the main lobe and the height of the side lobe. Fig. 4.78 is an example of Dolph-Chbychev weighting for a line array with 12 equal-spaced elements.

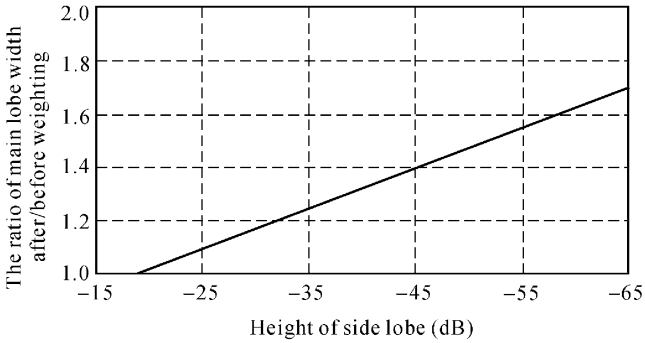


Fig. 4.77. Characteristics of Chebychev weighting for an equal spaced line array

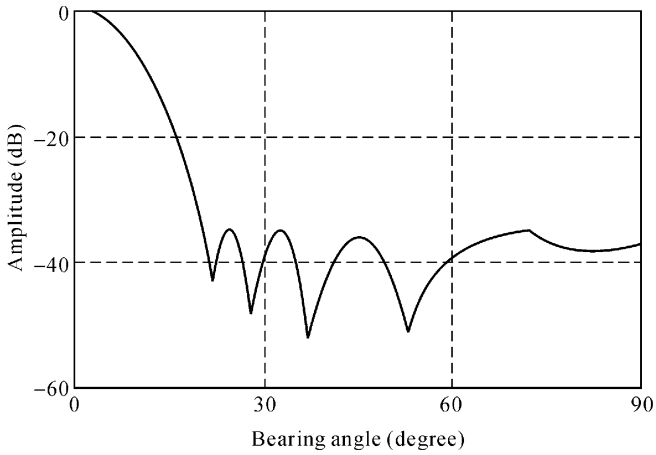


Fig. 4.78. An example of Dolph-Chebychev weighting for a line array ($N = 12$), ($d/\lambda = 0.4$, $A_1 = 0.327$, $A_2 = 0.428$, $A_3 = 0.733$, $A_4 = 0.983$, $A_5 = 1.184$, $A6 = 1.295$)

(2) Similar weighting

Similar weighting is suitable for any kind of plane array. The criterion of this weighting is maximizing the output signal-to-noise ratio in the direction of the

main lobe. As an example, we are going to discuss similar weighting for a circle array with N elements. Assume the weight coefficient of the i-th signal is w_i. When the incidental angle of signal is θ_0, then the signal received in the i-th element is $a_i s(t+\tau_i(\theta_0))$, where $a_i = \xi(\theta_0 - (i-1)\alpha)$. After weighting, the signal becomes $w_i a_i s(t+\tau_i(\theta_0))$. The output noise is $w_i n_i(t)$. Suppose the noises are independent, and the system is steering in the direction of the signal. The signal component in beamforming output is

$$y_s(t) = \sum_{i=1}^{N} w_i a_i s(t)$$

The noise component in beamforming output is

$$y_n(t) = \sum_{i=1}^{N} w_i n_i(t - \tau_i(\theta_0))$$

The output signal-to-noise ratio is

$$(SNR)_{\text{out}} = \frac{E[y_s(t)]^2}{E[y_n(t)]^2} = \frac{\left(\sum_{i=1}^{N} w_i a_i\right)^2}{\sum_{i=1}^{N} w_i^2}$$

In terms of Schwartz inequality, it follows that

$$\left(\sum_{i=1}^{N} w_i a_i\right)^2 \le \left(\sum_{i=1}^{N} w_i^2\right)\left(\sum_{i=1}^{N} a_i^2\right)$$

The equality sign holds if, and only if $w_i = k a_i$ $(i = 1, \ldots, N)$ where k is a constant.

Therefore, we have

$$(SNR)_{\text{out}} \le \frac{\left(\sum_{i=1}^{N} w_i^2\right)\left(\sum_{i=1}^{N} a_i^2\right)}{\sum_{i=1}^{N} w_i^2} = \sum_{i=1}^{N} a_i^2$$

The equality sign holds if, and only if

$$w_i = ka_i = k\xi[\theta_0 - (i-1)\alpha] \tag{4.127}$$

A weighting, which satisfies Eq. (4.127), is called a similar weighting, in the sense that the weighting coefficient is proportional to the response of the incidental signal. a_i $(i = 1, ..., N)$ is actually a discrete sampling of the directivity function of a single hydrophone.

Example

Assume a baffling circle array has 36 elements and the directivity function of a single hydrophone is

$$\xi(\theta) = \exp[-(\theta / 90^\circ)^2]$$

The beamforming system has a working sector of $M = 12$ elements. Then the similar weighting should be

$$w_i = \xi[(i-0.5)\alpha] \ (\alpha = 360^\circ / 36, \ \ i-1, \ 2, \ ..., \ 6)$$

It is easy to calculate

$$w_1 = 0.99, \ w_2 = 0.97, \ w_3 = 0.93, \ w_4 = 0.86, \ w_5 = 0.78, \ w_6 = 0.69$$

The similar weighting improves the signal-to-noise ratio in the steering direction, but has almost no effect on the side lobe.

Besides the weighting algorithm described above, there are other kinds of weighting. For example, the Taylor weighting has similar criteria to Dolph Chebychev weighting, but is suitable not only for the equal-spaced line array, but also for other kinds of array[37].

4.7.8 General Expression of Directivity Function

From the procedure for calculating the directivity function of an array we know that the key step is to calculate the difference in the sound path. For a spatial discrete array the simplest method is to use a spherical coordinate (Fig. 4.79). H_1 represents the *i*-th hydrophone, the rectangle coordinate is (x_i, y_i, z_i) and the spherical coordinate is $(\theta_i, \ \varphi_i, \ r_i)$. The relationship between these coordinates is

$$x_i = r_i \sin\theta_i \cos\varphi_i, \quad y_i = r_i \sin\theta_i \sin\varphi_i, \quad z_i = r_i \cos\theta_i$$

Assume the angles between OH_i and the three coordinate axes OX, OY, OZ are α_i, β_i, γ_i. Then

$$\cos\alpha_i = x_i / r_i = \sin\theta_i \cos\varphi_i$$
$$\cos\beta_i = y_i / r_i = \sin\theta_i \sin\varphi_i$$
$$\cos\gamma_i = z_i / r_i = \cos\theta_i$$

Suppose the incidental angle of the signal is (θ, φ); the angles between the incidental direction and the three coordinates' axes OX, OY, OZ are (α, β, γ). It follows that

$$\cos\alpha = \sin\theta\cos\varphi,$$
$$\cos\beta = \sin\theta\sin\varphi,$$
$$\cos\gamma = \cos\theta.$$

The angle between the incidental direction and OH_i is indicated by δ_i. The cosine of δ_i is given by

$$\cos\delta_i = \cos\alpha\cos\alpha_i + \cos\beta\cos\beta_i + \cos\gamma\cos\gamma_i$$

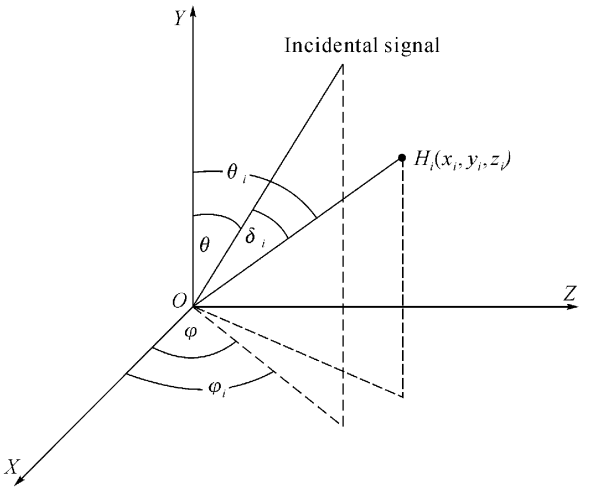


Fig. 4.79. Calculation of the beam pattern of a spatial distributed array

In Fig. 4.79, we can calculate the difference H_iO in sound path $d_i = r_i \cos\delta_i = x_i \cos\alpha + y_i \cos\beta + z_i \cos\gamma = x_i \sin\theta\cos\phi + y_i \sin\theta\sin\phi + z_i \cos\theta$.

Therefore, the phase difference can be derived as

$$\xi_i = 2\pi(x_i \sin\theta\cos\varphi + y_i \sin\theta\sin\varphi + z_i \cos\theta)/\lambda .$$

If we want to steer the direction of $(\theta_0,\ \varphi_0)$, then the signal in H_i should be compensated

$$\xi_{i0} = 2\pi(x_i \sin\theta_0\cos\varphi_0 + y_i \sin\theta_0\sin\varphi_0 + z_i \cos\theta_0)/\lambda .$$

The signal after phase compensation becomes

$$s_i(t) = A_i \cos(2\pi ft + \Delta\xi_i) .$$

where $\Delta\xi_i = \xi_i - \xi_{i0}$. The summation of N elements of signal is

$$y(t) = \sum_{i=1}^{N} A_i \cos(2\pi ft + \Delta\xi_i) \tag{4.128}$$

Therefore, the directivity function is

$$D(\theta,\ \varphi) = \left[\left(\sum_{i=1}^{N} A_i \cos\Delta\xi_i\right)^2 + \left(\sum_{i=1}^{N} A_i \sin\Delta\xi_i\right)^2\right]^{1/2} \tag{4.129}$$

This is the general expression for calculating the directivity function. Sometimes, for convenience, we can adopt the complex expression:

$$\begin{aligned} u(t) &= \sum_{i=1}^{N} A_i \exp[\,j(2\pi ft + \Delta\xi_i)], \\ D(\theta,\ \varphi) &= \| u(t) \| \end{aligned} \tag{4.130}$$

The result of Eq. (4.128) is two-dimensional functions. We usually observe it in one section. In sonar design in particular, the spatial distributed array almost always has some kind of regular shape. In calculating the directivity function, it is very important to use the characteristics of element distribution when choosing a suitable coordinate system. Sometimes it is possible to use product theorem to simplify calculation. The simplification procedure is similar to this: assume an array consists of $N = M \times L$ elements. Divide it into M sub-arrays, each with L elements.

The phase compensation of the (m, l)-th element is $\Delta\xi_{ml}$ if the structure of the sub-arrays is "separable", i.e., $\Delta\xi_{ml} = \xi_m + \eta_l$. At the same time, the response of

the (m, l)-th $A_{ml} = A_l$ is independent of m. This gives

$$\begin{aligned} y(t) &= \sum_{m=1}^{M}\sum_{l=1}^{L} A_{ml}\cos(2\pi ft + \Delta\xi_{ml}) \\ &= \sum_{m=1}^{M}\sum_{l=1}^{L} A_l\cos(2\pi ft + \xi_m + \eta_l) \\ &= \sum_{m=1}^{M} A(\theta,\ \varphi)\cos(2\pi ft + \xi_m + \eta) \end{aligned}$$

where

$$A(\theta,\ \varphi) = \left[\left(\sum_{l=1}^{L} A_l\cos\eta_l\right)^2 + \left(\sum_{l=1}^{L} A_l\sin\eta_l\right)^2\right]^{1/2}$$

$$\eta = \arctan\frac{\sum_{l=1}^{L} A_l\sin\eta_l}{\sum_{l=1}^{L} A_l\cos\eta_l}$$

From Eq. (4.68), we have

$$\begin{aligned} y(t) &= A(\theta,\ \varphi)\sum_{m=1}^{M}\cos(2\pi ft + \xi_m + \eta) \\ &= A(\theta,\ \varphi)B(\theta,\ \varphi)\cos(2\pi ft + \xi + \eta) \end{aligned}$$

where

$$B(\theta,\ \varphi) = \left[\left(\sum_{m=1}^{M}\cos\xi_m\right)^2 + \left(\sum_{m=1}^{M}\sin\xi_m\right)^2\right]^{1/2}$$

$$\xi = \arctan\frac{\sum_{m=1}^{M}\sin\xi_m}{\sum_{m=1}^{M}\cos\xi_m}$$

Clearly, $A(\theta,\ \varphi)$ is the directivity function of one class of sub-arrays and $B(\theta,\ \varphi)$ is that of another. The combination of these two classes of directivity function results in the directivity function of the whole array

$$D(\theta,\ \varphi) = A(\theta,\ \varphi)B(\theta,\ \varphi) \tag{4.131}$$

The above method is frequently used in calculating the directivity function of large arrays. Quite often $A(\theta, \varphi)$ and $B(\theta, \varphi)$ are one variable function; in this case the calculation can be simplified further. For a fixed φ, $D(\theta, \varphi)$ is called the directivity function in the meridian plane, and for a fixed θ, $D(\theta, \varphi)$ is called the directivity function in the horizontal plane. In particular, when $\theta = 0$, $D(0, \varphi)$ is called the directivity function in the equatorial plane.

Example

A discrete cylindrical array and a spherical array are shown in Fig. 4.65. We can calculate the directivity function as follows.

1) Discrete cylindrical array

We can think of a cylindrical array as being the combination of two classes of sub-arrays: one is the line array, which is distributed along generatrix, and the other is the horizontal circle array. Suppose a line array in generatrix has L elements, and a circle array has M elements. Now, using the product theorem, the directivity of a cylindrical array is

$$D(\theta, \varphi) = D_0(\theta) D_1(\theta, \varphi) \tag{4.132}$$

where $D_0(\theta)$ is the directivity function of a line array

$$D_0(\theta) = \frac{\sin[\pi \mathrm{d}(\sin\theta - \sin\theta_0)/\lambda]}{\pi \mathrm{d}(\sin\theta - \sin\theta_0)/\lambda}$$

$D_1(\theta, \phi)$ is the directivity function of an arc array compensated by spatial time delay

$$D_1(\theta, \varphi) = \left[\left(\sum_{i=1}^{M} \cos \Delta\xi_i \right)^2 + \left(\sum_{i=1}^{M} \sin \Delta\xi_i \right)^2 \right]^{1/2}$$

$$\Delta\xi_i = 2\pi r[\sin\theta \cos(\varphi_i - \varphi) + \sin\theta_0 \cos(\varphi - \varphi_0)]/\lambda$$

2) Discrete spherical array

If we choose the center of a spherical array as the reference point of time delay, it is easy to prove that

$$\begin{aligned} \Delta\xi_i = {} & 2\pi r[\sin\theta_i \cos\theta \cos(\varphi_i - \varphi) + \cos\theta_i \cos\theta]/\lambda \\ & - 2\pi r[\sin\theta_i \cos\theta_0 \cos(\varphi_i - \varphi_0) + \cos\theta_i \cos\theta_0]/\lambda \end{aligned}$$

Substituting it into Eq. (4.129) we can find the directivity function of the discrete spherical array.

4.7.9 Continuous Distributed Array

In the practice of sonar application, when the wavelength of the signal of interest is much larger than the distance between adjacent elements, then the array can be considered as a continuous array. Therefore, the ratio of the wavelength of the signal over the distance between adjacent elements is an important parameter which determines if an array can be considered to be continuous or discrete.

Fig. 4.80 illustrates five types of continuous array. The directivity function of each of these arrays can be derived from the corresponding discrete array. The meaning of related parameters expressed in the equation can be found in Fig. 4.80.

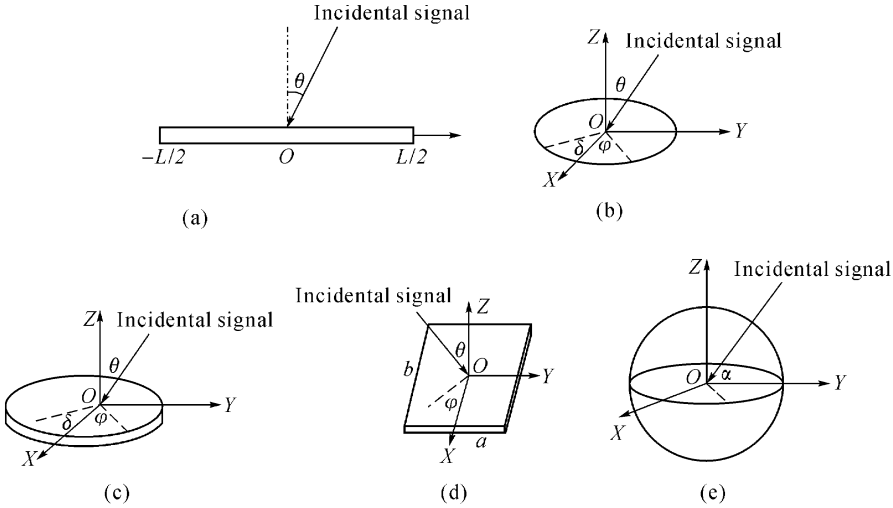


Fig. 4.80. Examples of continuous distributed arrays. (a) Continue linear array; (b) Ring array; (c) Circle plate array; (d) Rectangular plate array; (e) Spherical array

(1) Linear array

$$D(\theta) = \frac{\sin(\pi L \sin\theta / \lambda)}{\pi L \sin\theta / \lambda} \tag{4.133}$$

(2) Ring array

$$D(\theta) = J_0(2\pi r \sin\theta / \lambda) \tag{4.134}$$

where $J_0(\cdot)$ is a zero order Bessel function. For $r >> \lambda$, the directivity index is $\mu \approx 4\pi r / \lambda$.

(3) Circle plate array

$$D(\theta) = \frac{2J_1(2\pi r \sin\theta / \lambda)}{2\pi r \sin\theta / \lambda} \tag{4.135}$$

where $J_1(\cdot)$ is the first order Bessel function. For $r >> \lambda$, the directivity index is $\mu \approx (2\pi r / \lambda)^2$.

(4) Rectangular plate array

$$D(\theta,\varphi) = \frac{\sin[\pi b \sin\theta\cos\varphi / \lambda]\sin[\pi a \sin\theta \sin\varphi / \lambda]}{[\pi b \sin\theta\cos\varphi / \lambda][\pi a \sin\theta \sin\varphi / \lambda]} \tag{4.136}$$

(5) Spherical array

$$D(\alpha) = \frac{\sin[4\pi r \sin(\alpha / 2) / \lambda]}{4\pi r \sin(\alpha / 2) / \lambda} \tag{4.137}$$

4.8 Calculation of Sonar System Gain

To effectively extract the signal from background noise, it is necessary to operate a processing series for the received signal. The processing can be categorized into two steps.

(1) Beamforming: by using the differences in statistical characteristics between the signal and noise, the hydrophone array obtains spatial gain GS.

(2) Post-processing of the output signal of the beamformer and this obtains the time gain GT for the sonar system.

Fig. 4.81 shows the change in the signal-to-noise ratio in a sonar system.

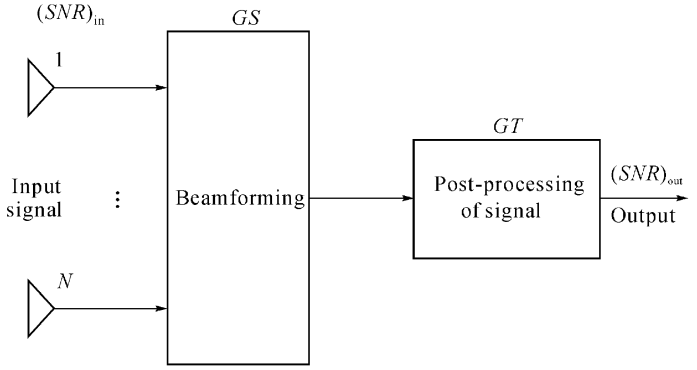


Fig. 4.81. Calculation of sonar system gain

In Fig. 4.81, the signal-to-noise ratio of a single hydrophone is $(SNR)_{in}$, spatial gain *GS* and time gain *GT*. In the output end of a sonar system, the signal-to-noise ratio is $(SNR)_{out} = (SNR)_{in} + GS + GT$. The signal processing of a sonar system includes some techniques in time or frequency domain that aim to extract

the signal of interest from the background noise. Note that the differences between the signal and noise in their statistical characteristics are important. The choice of signal processing algorithms strongly depends on these differences.

4.8.1 Spatial Gain of Sonar System

As shown above, the gain of an array with N elements is

$$GS = 10\lg N \tag{4.138}$$

provided that the noise received by each element is independent.

In using Eq. (4.137), it is necessary to understand that the value N indicates the number involved in beamforming rather than the total number of elements in an array. For example, for an arc array, if the number of total elements is N, the working sector of each beam is $2\pi \times 2M / N$ (rad), and the spatial gain is

$$GS = 10\lg 2M + 10\lg \mu \tag{4.139}$$

where μ is the gain of a single hydrophone in the baffling array. If $\xi(\theta)$ is the directivity function of a single hydrophone, we obtain

$$\mu = \frac{2\pi}{\int_0^{2\pi} \xi^2(\theta)\mathrm{d}\theta} \tag{4.140}$$

Example

Suppose $\xi(\theta) = \exp[-(\theta / 90^\circ)^2]$, we obtain $10\lg \mu \approx 4.7$ dB. If $M = N/2$, then $GS \approx 10\lg N + 1.7$ (dB). We find that for a baffling array, although the number of elements in beamforming decreases to half the total number, the spatial gain increases due to the baffling.

For an arc array,

$$MD = 10\lg 2M - 10\lg N + 10\lg \mu \tag{4.141}$$

is called the modified factor. Therefore, the spatial gain of an arc array sonar is

$$GS\,|_{\text{arc array}} = 10\lg N + MD \tag{4.142}$$

Generally, it is reasonable to choose $GS = 10\lg N$. But we have to remember that it is constrained by some conditions. For example, if the array is working in

an isotropic, homogeneous sound field, the distance between adjacent elements has to be greater than the half wave length of the incidental signal. When the practical situation differs from previous conditions, it is necessary to make some modifications.

4.8.2 Calculation of Time Processing Gain of Passive Sonar

It is necessary to run a post-processing series for the beamformed signal. In the post-processing, time accumulation is necessary. This is because the fluctuation in the random signal must be smoothed by time integration and because the signal energy can be distinguished from the background noise only after long time accumulation.

Indicating ambient noise by $n(t)$ and radiated noise (i.e., signal) by $s(t)$, the received signal may be $x(t) = n(t)$ or $x(t) = s(t) + n(t)$. Therefore, we face a hypothesis test problem:

H_0 : noise only, $x(t) = n(t)$;

H_1 : signal plus noise, $x(t) = s(t) + n(t)$.

As we have shown in sections 4.3 and 4.4, the ambient noise and radiated noise of a target can be considered, at least approximately, as Gaussian stationary random signal. Therefore, the signal processing problem described here is based on the optimum detection of Gaussian signal in a Gaussian noise background. Fig. 4.82 shows a simplified model of passive sonar signal processing.

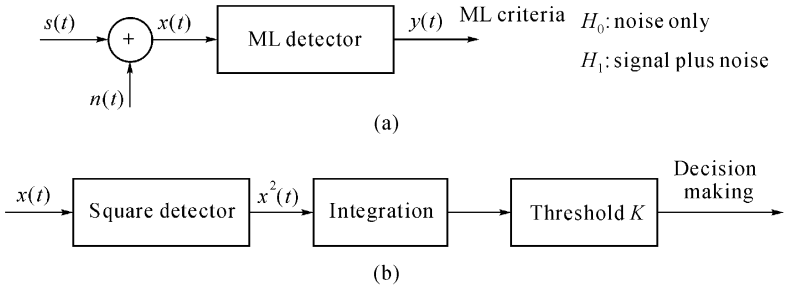


Fig. 4.82. Optimum detector for passive sonar. (a) Block diagram; (b) Realization

The maximum likelihood criterion is adopted as the optimum detection standard, i.e., in the case that a false alarm probability is less than or equal to a given positive value, find a test which maximizes the detection probability.

Suppose the N observations are $x(t_1)$, ..., $x(t_N)$. The joint probability function is $p_N(x_1, ..., x_N)$. Suppose $x(t_i)$ $(i = 1, ..., N)$ are independent. Then $pN(x_1, \ldots,$

$x_N) = p(x_1) \ldots p(x_N)$.

where $p(x)$ is a one-dimensional Gaussian probability density function. In the case of noise only, we have

$$p(x_i)\,|_{n(t)} = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{x_i^2}{2\sigma_n^2}\right)$$

Now indicate

$$p_0(x_1,\ ...,\ x_N) = p_N(x_1,\ ...,\ x_N)\,|_{H_0} = \frac{1}{(2\pi\sigma_n^2)^{N/2}} \exp\left(-\frac{x_1^2 + ... + x_N^2}{2\sigma_n^2}\right) \tag{4.143}$$

Similarly, when the signal presents, the probability density function of N samples is

$$p_1(x_1,\ ...,\ x_N) = p_N(x_1,\ ...,\ x_N)\,|_{H_1} = \frac{1}{(2\pi(\sigma_n^2 + \sigma_s^2))^{N/2}} \exp\left[-\frac{x_1^2 + ... + x_N^2}{2(\sigma_n^2 + \sigma_s^2)}\right] \tag{4.144}$$

where σ_s^2 and σ_n^2 represent the signal and noise power, respectively. Therefore, the likelihood ratio function is

$$\begin{aligned}\Lambda(x_1,\ ...,\ x_N) &= \frac{p_1(x_1,\ ...,\ x_N)}{p_0(x_1,\ ...,\ x_N)} \\ &= \frac{1}{(2\pi(\sigma_n^2 + \sigma_s^2))^{N/2}} \exp\left[-\frac{\sigma_s^2}{2(\sigma_s^2 + \sigma_n^2)\sigma_n^2}(x_1^2 + ... + x_N^2)\right]\end{aligned}$$

The accepting region can be calculated as

$$X_1 = \{(x_1,\ ...,\ x_N) \mid \Lambda(x_1,\ ...,\ x_N) \geq Z_0\} = \left\{(x_1,\ ...,\ x_N)\left|\frac{x_1^2 + ... + x_N^2}{N} \geq M\right.\right\}$$

where

$$M = \frac{2(\sigma_s^2 + \sigma_n^2)\sigma_n^2}{N\sigma_n^2} \ln\left[Z_0\left(\frac{\sqrt{\sigma_s^2 + \sigma_n^2}}{\sigma_n}\right)^N\right]$$

Clearly, the optimum detector of signal from noise is a square detector, which performs the operation of

$$\xi_N = (x_1^2 + ... + x_N^2) / N \geq M \tag{4.145}$$

The detector expressed in Eq. (4.145) is also called an energy detector. It actually sums up the N-sample values and takes the average.

The above deduction for an optimum detection system is based on the independent assumption of samples. We can think of it as a sparse sampling of the signal. When the sampling becomes finer then the independence among samples cannot be guaranteed. However, in some conditions, it is still proved that the energy detector is optimum. That is

$$y(t) = \frac{1}{T} \int_{t-T}^{t} x^2(u)\mathrm{d}u \geq M \tag{4.146}$$

where T is the integration time, and is equivalent to the product NT_s of observation number N and sampling duration T_s in the discrete situation.

Each beam output of a conventional beamforming system is necessary to perform post-processing of integration; the comparison of multi-beam values gives the information of the incidental angle of the target.

In the following, we use Eq. (4.146) to explain how to calculate the time gain GT of passive sonar.

Suppose the array consists of N elements, the i-th signal is

$$x_i(t) = s(t + \tau_i(\theta)) + n_i(t)$$

where θ is the incidental angle of the signal, and $\tau_i(\theta)$ is the time delay of i-th signal. After time delay $\tau_i(\theta_0)$ for the i-th signal $x_i(t)$, we obtain

$$y_i(t) = x_i(t - \tau_i(\theta_0)) = s(t + \tau_i(\theta) - \tau_i(\theta_0)) + n_i(t - \tau_i(\theta_0))$$

Summing up all $y_i(t)$, we obtain

$$y(t) = \sum_{i=1}^{N} y_i(t) \tag{4.147}$$

For $\theta_0 = \theta$, the beam will steer to the target. The output of the square detector is

$$D(\theta,\ \theta_0) = E\left[\sum_{i=1}^{N} y_i(t)\right]^2 \tag{4.148}$$

Since the noises are independent, it is easy to see that in the case of noise only, we have

$$D|_n = N\sigma_n^2$$

When the signal presents, we have

$$D|_{s+n} = N^2\sigma_s^2 + \sigma_n^2$$

The direct current jump of the beam steered at the target is

$$\Delta = \sqrt{D|_{s+n}} - \sqrt{D|_n} = (N^2\sigma_s^2 + N\sigma_n^2)^{1/2} - (N\sigma_n^2)^{1/2} \tag{4.149}$$

The normalized value is

$$\Delta_0 = \frac{\Delta}{\sqrt{N\sigma_n^2}} = [N \times (SNR)_{\text{in}} + 1]^{1/2} - 1 \tag{4.150}$$

where $(SNR)_{\text{in}} = \sigma_s^2 / \sigma_n^2$ indicates the input signal to noise ratio.

Δ_0 expresses the fact that the amplitude changes in the direction of signal in a multi-beam background due to the presence of the signal. We can also transfer the value of Δ_0 to dB value, i.e.

$$20\lg\Delta_0 = 20\lg[(N \times (SNR)_{\text{in}} + 1)^{1/2} - 1]$$

Fig. 4.83 shows the relationship between Δ_0 and $(SNR)_{\text{in}}$.

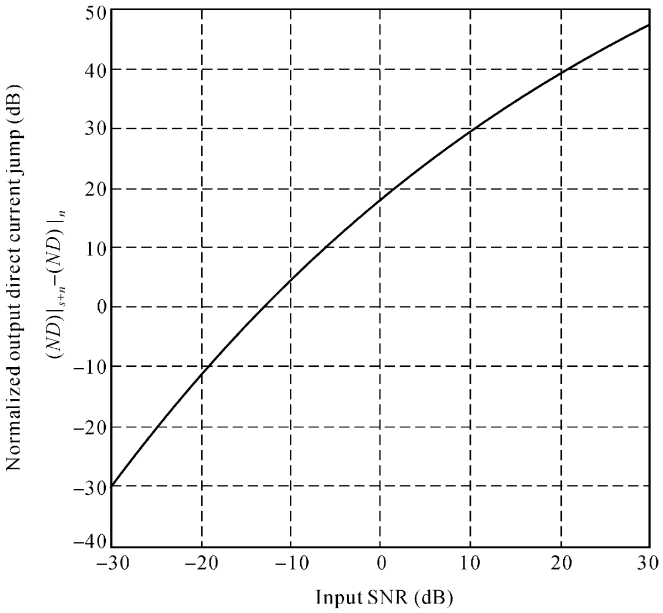


Fig. 4.83. Relationship between input signal-to-noise ratio and output direct current jump ($N = 64$)

Eq. (4.150) is very important in estimating sonar performance. This is because, if we know the value of Δ_0, we can deduct the value of $(SNR)_{\text{in}}$ and this is important. We usually do not know the exact signal power but the noise background can be monitored in real time. Once we obtain $(SNR)_{\text{in}}$, it is equivalent to obtaining σ_s^2. This is important in sonar range prediction, target identification and torpedo alarming.

Now we turn to calculating the time gain of a sonar system. As mentioned above the input signal-to-noise ratio is $(SNR)_{\text{in}} = \sigma_s^2 / \sigma_n^2$. The output signal-to-noise ratio is defined as the ratio of direct current jump of the beam steered target over the value of output fluctuation of noise, i.e.

$$(SNR)_{\text{out}} = \frac{E[y(t)]\,|_{s+n} - E[y(t)]\,|_n}{[Var(y(t))\,|_n]^{1/2}} \tag{4.151}$$

$$GT = \frac{(SNR)_{\text{out}}}{(SNR)_{\text{in}}} \tag{4.152}$$

Based on Eq. (4.145), it is easy to show that

$$E[y(t)]\,|_{s+n} = E[\frac{1}{T}\int_{t-T}^{t}(s(u)+n(u))^2\,\mathrm{d}u] = \sigma_s^2 + \sigma_n^2$$

$$E[y(t)]\,|_n = \sigma_n^2$$

$$\mathrm{Var}[y(t)]\,|_n = \mathrm{Var}[\frac{1}{T}\int_{t-T}^{t} n^2(u)\mathrm{d}u]$$

$$E\left[\left(\frac{1}{T}\int_{t-T}^{t} n^2(u)\mathrm{d}u\right)\right]^2 = \left[E\left(\frac{1}{T}\int_{t-T}^{t} n^2(u)\mathrm{d}u\right)\right]^2 \tag{4.153}$$

The second term in Eq. (4.153) is equal to σ_n^4. To calculate the first term, it is necessary to use the formula for calculating the expectation value of four Gaussian signal products

$$E[x_1x_2x_3x_4] = E(x_1x_2)E(x_3x_4) + E(x_2x_3)E(x_1x_4) + E(x_1x_3)E(x_2x_4) \tag{4.154}$$

Therefore,

$$E\left[\frac{1}{T}\int_{t-T}^{t} n^2(u)\mathrm{d}u\right]^2 = E\left[\frac{1}{T^2}\int_{t-T}^{t}\int_{t-T}^{t} n^2(u)n^2(v)\mathrm{d}u\mathrm{d}v\right]$$
$$= \frac{2}{T^2}\int_{t-T}^{t}\int_{t-T}^{t} R_n^2(u-v)\mathrm{d}u\mathrm{d}v + \sigma_n^4$$

Here, the first integration is two fold, and we can change it to one fold by using the following transform: $u - v = \tau$. We have

$$\frac{1}{T^2}\int_{t-T}^{t}\int_{t-T}^{t} R_n^2(u-v)\mathrm{d}u\mathrm{d}v = \frac{2}{T}\int_0^T R_n^2(\tau)\left(1-\frac{\tau}{T}\right)\mathrm{d}\tau \tag{4.155}$$

where, $R_n(\tau)$ is the auto-correlation function of noise. Substituting these results into Eqs. (4.150) and (4.151), we obtain

$$\begin{aligned}(SNR)_{\text{out}} &= (SNR)_{\text{in}}\left[\frac{T}{2\int_0^T\left(1-\frac{\tau}{T}\right)\rho_n^2(\tau)\mathrm{d}\tau}\right]^{1/2}\\ GT &= \left[\frac{T}{2\int_0^T\left(1-\frac{\tau}{T}\right)\rho_n^2(\tau)\mathrm{d}\tau}\right]^{1/2}\end{aligned} \tag{4.156}$$

where $\rho_n(\tau)$ is the auto-correlation coefficient of noise.

The denominator

$$\Delta\tau = 2\int_0^T\left(1-\frac{\tau}{T}\right)\rho_n^2(\tau)\mathrm{d}\tau \tag{4.157}$$

in Eq. (4.155) is defined as the correlation radius of noise because this value indicates the magnitude of the nonzero segment of the correlation function.

Now, Eq. (4.155) becomes

$$GT = (T/\Delta\tau)^{1/2} \tag{4.158}$$

This expression is very comprehensive and easy to understand. It says that the time gain of a passive sonar system is equal to the ratio of integration to the radius of noise auto-correlation. It is necessary to show, for a broadband noise, $\Delta\tau$ is roughly equal to the inverse of bandwidth Δf. Therefore, $GT \approx (T \times \Delta f)^{1/2}$.

Note that, from a theoretical point of view, the longer the integration time is, the larger the time gain is. But in fact, the increase in time gain is not unlimited. When

the observation time becomes long, the stationarity of noise cannot be guaranteed.

In realizing square detection, it is important to note that the square operation is not always convenient. The cost of hardware is much higher than that of an absolute operation, for example. Therefore, sometimes we prefer to use absolute value detection instead of square detection. For the Gaussian signal, the absolute detector is also the optimum detector.

It is easy to prove this conclusion. Suppose $x(t)$ is Gaussian signal, and the probability density function is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{\sigma^2}\right)$$

The output of a square detector is

$$D_0 = E[x^2(t)] = \int x^2 p(x) \mathrm{d}x = \sigma^2 \tag{4.159}$$

and the output of an absolute value detector is

$$D_1 = E[|\, x(t)\,|] = \int |\, x\,|\, p(x) \mathrm{d}x = \sqrt{\frac{2}{\pi}}\sigma \tag{4.160}$$

Comparing Eqs. (4.159) and (4.160), we can see that the difference between $\sqrt{D_0}$ and D_1 is a constant factor of $\sqrt{2/\pi}$.

4.8.3 Calculation of Time Processing Gain of Active Sonar

The detection problem of active sonar is different from that of passive sonar, although they both involve dealing with signal detection problems in a noise background. However, we have to treat active sonar signal detection problem from a different point of view.

Because the time length of the transmitting signal is limited, the time length of the echo signal is also limited such that the integration time of active signal processing is not unlimited. However, active signal processing must satisfy the specific requirements of echo frequency structure. As shown in sections 3.4 and 4.5, the matched filter is one of the optimum detectors for extracting signals from background noise.

Fig. 4.84 illustrates a block diagram of optimum processing of the active sonar signal.

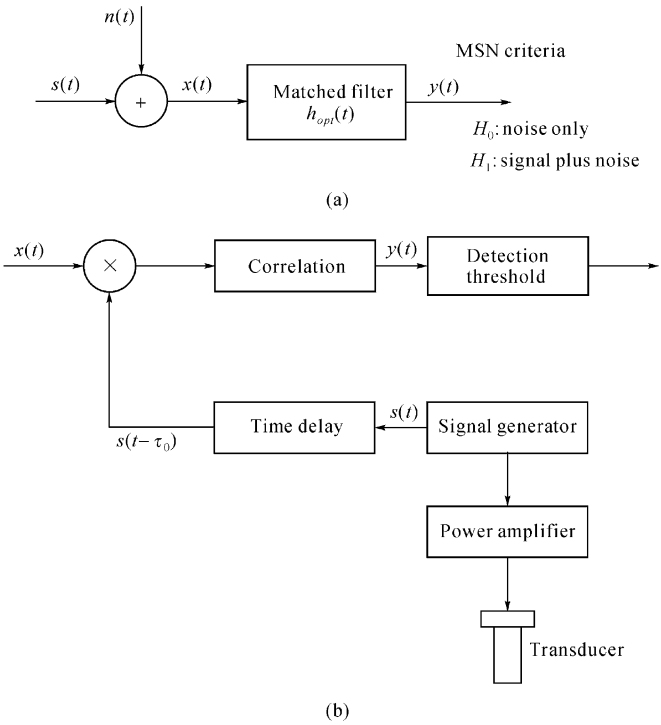


Fig. 4.84. Optimum signal detector of active sonar. (a) Block diagram; (b) Realization

According to Eq. (4.161), the time gain of a matched filter is

$$GT = \frac{(SNR)_{\text{out}}}{(SNR)_{\text{in}}} = 2TW \tag{4.161}$$

where T is the impulse duration and W is the band width. For a rectangular wave modulation signal, $2TW = 1$. Therefore, the time gain is 0 dB.

Examples

(1) Suppose the integration time of an energy detector of passive sonar is $T = 1$ s, and the correlation radius of input noise is $\Delta\tau = 500\ \mu s$, then $T / \Delta\tau = 2000$. The time gain of this sonar system is $GT = 5\lg 2,000 = 16.5$ dB.

(2) Suppose the impulse width of the transmitting signal, which is frequency modulated, of an active sonar is 40 ms. The carrier frequency is 4 kHz, and the band width of modulation frequency is $\Delta f = \pm 100$ Hz, then $GT = 10\lg 2TW = 10\lg 16 = 12$ dB.

Fig. 4.85 summarizes the calculation of spatial and time gain for passive and active sonar systems.

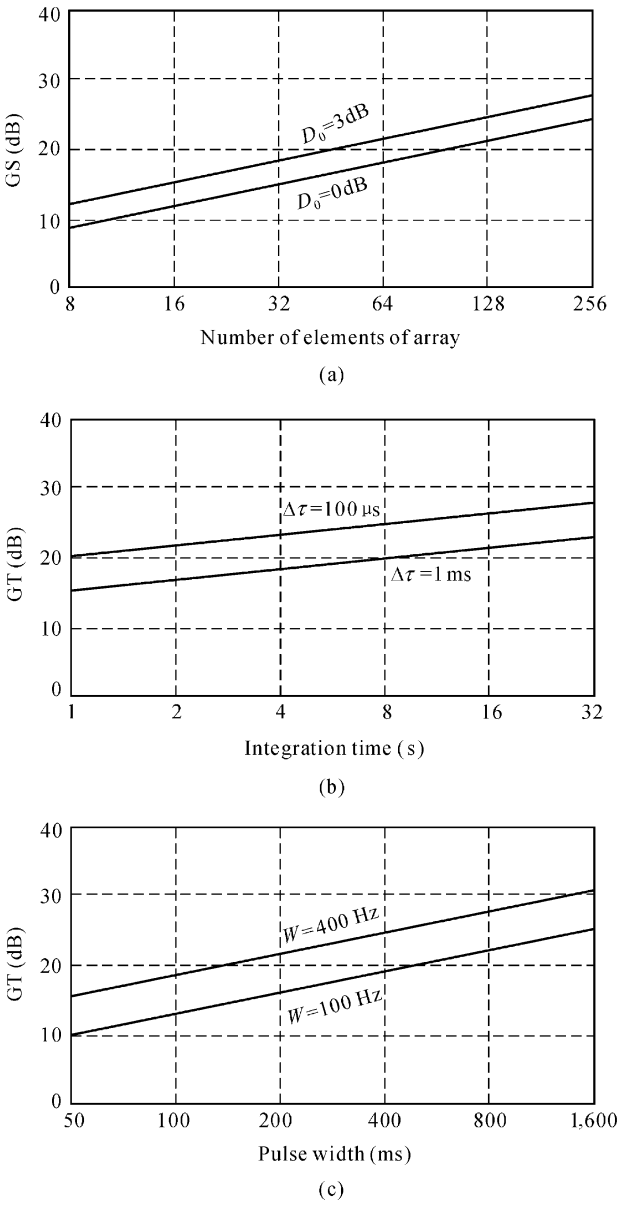


Fig. 4.85. Calculation of space and time gains for sonar systems. (a) Space gain (suitable for active / passive sonar; (b) Time gain (suitable for passive sonar; (c) Time gain (suitable for active sonar)

Fig. 4.85(a) shows the spatial gain suitable for both passive and active sonar; Fig. 4.85(b) shows the time gain suitable for passive sonar and Fig. 4.85(c) shows the time gain suitable for active sonar.

4.9 Gain Loss of a Sonar System in the Interface of Various Sub-systems

The system gain of sonar is the key parameter which determines sonar performance. The greater the system gain is, the lower the detectable minimum signal-to-noise ratio is. It is difficult to guarantee the system gain of an entire sonar system if the sonar system consists of several sub-systems. This is because the interface of sub-systems will result in loss of system gain. The gain obtained in one sub-system may be lost in the next sub-system due to a mismatch of the two sub-systems.

Winder showed[10] that the system gain is easy to lose at the interface of two sub-systems. That means that the entire system gain may be less than the sum of its sub-system gains. Therefore, the correct merging of two systems is very important. A well-designed sonar system will realize "seamless connection", based on a knowledge of sonar modeling, human / machine interfaces, data fusion, etc. The goal of this section is to describe the interface of sonar systems and how to minimize the loss of gain at the interface.

4.9.1 Relationship between Sonar System Gain and Input signal-to-noise Ratio

In calculating sonar system gain, we have to pay attention to the value of the input signal-to-noise ratio. Inherent in the derivation procedure is an assumption that the system gain is independent of the input signal-to-noise ratio. However, in fact, sonar system gain depends on the input signal-to-noise ratio, and this dependency is usually not constant. For a small $(SNR)_{in}$, the system gain is large. With an increase in the input signal-to-noise ratio, the system gain will decrease. As we are concerned with the situation of weak signal detection, we always use the gain value calculated in small $(SNR)_{in}$ as the system gain. Fig. 4.86 shows a typical relationship between relative system gain and input signal-to-noise ratio. Of course, in application, the system gain should be calculated based on the structure of the sonar system.

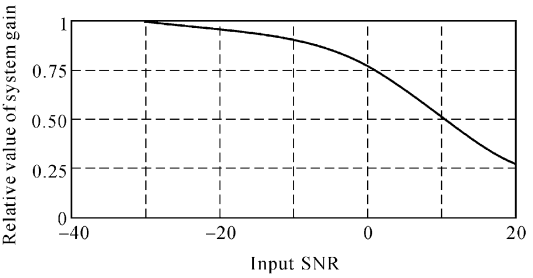


Fig. 4.86. Relationship between system gain and input signal to noise ratio

Example

Suppose a sonar array consists of N elements and the signal system adopts a square detector. Based on Eqs. (4.149) and (4.150) it is proved that the normalized direct current jump of the beam which steers in the target is

$$\Delta_0 = \frac{\Delta}{\sqrt{N\sigma_n^2}} = [(1 + N(SNR)_{\text{in}}]^{1/2} - 1 \tag{4.162}$$

where $(SNR)_{\text{in}} = \sigma_s^2 / \sigma_n^2$. Therefore, the system gain is

$$G_s = \frac{\Delta_0}{(SNR)_{\text{in}}} = \frac{[(1 + N(SNR)_{\text{in}}]^{1/2} - 1}{(SNR)_{\text{in}}} \tag{4.163}$$

It is clear that the smaller is the value of $(SNR)_{\text{in}}$, the greater is the value of G_s.

4.9.2 Gain Loss at the Interface of a Hydrophone and an A / D Converter

The hydrophone array is placed in an environment of ambient noise. As the receiver of sound pressure, the hydrophone has its frequency band and sensitivity. It must be set to guarantee the receiving of the minimum sea noise, e.g., the 0 sea state noise. As we have shown in section 4.1, if a hydrophone has the sensitivity –194 dB (ref. 0 dB = 1V / μPa), then for a sound pressure of 5×10^4 μPa, the output sound pressure of the hydrophone is about 10 μV. This means that the self-noise of a subsequent pre-amplifier must be much less than 10 μV. Statistical data of ambient noise show that in the frequency band of 100 – 10,000 Hz, the ambient noise level of 0 sea state is about 10 – 20 μV. Suppose in this environment of ambient noise, there is a target with a signal to noise ratio of 0 dB, then the output voltage of the hydrophone will be 14 – 28 μV. Usually, in the case of passive detection, it is not often that the signal-to-noise ratio is greater than 0 dB, unless the target is very close to the receiving array. In active sonar detection, at the moment of signal transmission, the signal-to-noise ratio will be much greater than 0 dB, and the pre-amplifier may be saturated. This is a kind of interface loss. To avoid this loss, two methods can be considered. The first is to use a TVGA (time varying gain amplifier). The gain of this amplifier will vary with time. The second method is to use a hard limiter, by which the input signal will be automatically limited when it is greater than a fixed voltage.

In digital sonar, the analogue signal should first be converted to digital. We have to guarantee that the A / D converter is working in good conditions. That is the operation of the A / D converter must be in a suitable working point of the

dynamic range. As shown in section 2.1, the signal-to-noise ratio of quantization is

$$SQNR = 10\lg\frac{\sigma^2}{\Delta^2/12} = 6n - 1.25 \tag{4.164}$$

where, σ^2 is the input signal power and n is the number of quantization. When $n \geq 8$, the noise resulting from quantization can almost be neglected. But, if the signal dynamic in the A / D converter is insufficient, then the gain loss is not neglectable. For example, if the designed signal power is σ^2, but the real input signal power is σ_0^2, and $\sigma_0^2 \leq \sigma^2$, the quantization noise is

$$SQNR|_{\text{new}} = 6n - 1.25 - 10\lg\frac{\sigma^2}{\sigma_0^2} \tag{4.165}$$

In Eq. (4.165), if $\sigma_0^2 / \sigma^2 = 0.1$, the quantization noise will increase by 10 dB and the gain loss at the interface of the hydrophone to the A / D converter will be 10 dB too.

4.9.3 *Interface Loss Due to Time Integration*

From spatial processing of beamforming to the integration of each beam represents the time accumulation procedure in signal processing. Many algorithms for special functions of sonar systems also use the beam information. All the results of these operations have to be displayed on the sonar screen. Today, the color display most frequently used is gray scale display. Theory and application shows that gray color displays satisfy the requirements of human / machine engineering. Although a multi-color display provides more options, the gray scale display still dominates the display of sonar post-processing.

In modern digital sonar, the gray scale can have up to 2^8 (256) levels. So the beamforming operation must take sufficient account of this. Suppose the sampling period of a digital sonar system is T_s and the integration time is T, then the accumulation number is

$$N_0 \approx T / T_s \tag{4.166}$$

Example

Suppose $T_s = 20$ μs, $T = 2$s, then $T / T_s = 10^5$. If the original datum is 12 bits, it follows that the accumulation results will be greater than 2^{28}. It may overflow, unless a 32-bit accumulator is used.

Multiple accumulation is a method that can avoid overflow, but guarantee accumulation number. Usually, a cascade accumulation is used. The first cascade

is a direct average and the second is a moving average. Suppose the beam data in a certain direction are $x(1)$, ..., $x(k)$, ..., then the first accumulation is

$$y_M(k) = \frac{1}{M}\sum_{l=0}^{M-1} x(k-l) \tag{4.167}$$

Generally, we can take $M = 1024$. Eq. (4.167) can be rewritten as

$$y_M(k) = x(k)/M + (1-1/M)y_{M-1}(k-1) \tag{4.168}$$

Based on this expression, in each step it is necessary only to add one new data point and to store old data. Eq. (4.168) is called the second accumulation, although it is actually the moving average. There is no overflow problem and so the possible loss of integration may be avoided.

4.9.4 *Loss at the Interface of the Signal Processor and the Display System*

As shown above, the gray scale display has 256 levels. But the output of signal processing, which is supposed to be displayed on the screen may have a much greater dynamic range than 256 levels. If we simply take away the lower bit data and convert only the higher bit data, then the performance of weak signal detection may decrease. If we take away the higher bit data, then it is possible that saturation phenomena may appear and the resolution of multiple targets will decrease. We will discuss this problem in Chapter 5. Here we only aim to express the importance of the interface between the signal processor and the display system. An incorrect interface may result in serious gain loss at the interface.

Example

Suppose the output beam datum is 16 bits and the weak signal appears in 4 – 5 bits. To display this data on screen, if we use all bits which are lower than 8 bits and convert those bits, which are higher than 8 bits, then all the data smaller or equal to 2^8 will disappear and the detection of weak signal is impossible. In contrast, if we force all data that are greater than 2^9 to be 2^8, it follows that the original two beam values of 2^{12} and 2^{14} cannot be distinguished.

4.10 Explosive Source of Underwater Sound

Explosive source is one of the most important experimental sources of underwater sound. It is extensively used in basic researches of underwater acoustics. In particular, it is used in the study of the propagation behavior of the underwater

acoustic signal in low frequency bands.

As early as the end of WW Ⅱ, researchers of underwater acoustics knew that a hand grenade with TNT explosive could be used to generate high intensity power undersea and also had a sufficiently wide frequency band. The effect of TNT explosive material is difficult to achieve by other means, for example, by transducers. It is difficult to manufacture series transducers that have high intensity and wide frequency bands. Therefore, in some cases explosive TNT becomes the only choice for underwater acoustic study of transmission behavior.

Of course, underwater explosive generates high intensity power which is harmful to sea animals, particularly to some marine mammals. This is a very important problem we have to face in using explosives as a source of underwater sound.

4.10.1 Main Characteristics of Explosive Sources of Underwater Sound

The main characteristics of explosive sources of underwater sound include the peak value of sound pressure, the impulse width of explosive sound and the frequency spectrum. Arone et al. reported their experimental results as early as the 1950s. The experimental formula has been confirmed by a large number of similar tests. Therefore, it is extensively used in the literature of underwater acoustics.

A shock wave from a TNT explosion has the form[1, 97-107]

$$p(t) = p_0 \exp(-t / t_0) \tag{4.169}$$

The peak value of sound pressure is

$$p_0 = 5.24\times10^7 \left[\frac{W^{1/3}}{R}\right]^{1.13} \ (\mathrm{N/m^2}) \tag{4.170}$$

where W is TNT mass in kg and R is the range from the explosion in m.

The wave form of TNT explosive has the form of an exponential function. t_0 can be considered as the width of the explosive impulse. It is called the "time constant" of an explosive impulse. The value of the time constant can be expressed as

$$t_0 = 92.5\times10^{-6} W^{1/3} \left[\frac{W^{1/3}}{R}\right]^{-0.22} \ (\mathrm{s}), \tag{4.171}$$

From Eq. (4.170) to Eq. (4.171) we know that 1 kg mass of TNT will generate a peak explosive impulse with broad band peak pressure at 1 m: $p_0 = 5.24\times10^{13}$ μPa, about 274.4 dB (ref. 0 db = 1 μPa).

The impulse width is about t_0 = 92.5 μs. Therefore, the spectrum width is about 10.8 kHz. That means, if the spectrum of the explosion signal is flat over the

frequency band, then the spectrum level of 1 kg of mass TNT in a frequency bin is about 234 dB at 1 m.

Figs. 4.87 and 4.88 illustrate the relationship between peak sound pressure and time constant with TNT mass. As we can see, for 1 kg TNT explosive, the peak sound pressure is as large as 270 dB (ref. 0 dB = 1 μPa). The time constant for 1 kg TNT explosive is about 90 μs. That means that TNT explosive can release a huge amount of energy in a very short time. That is why the explosive sound has a wide frequency band. The product of the time duration and the frequency band of the signal is equal to 1. Thus, the shorter the time duration is, the wider the frequency band is.

Fig. 4.89 shows the possible frequency band width of a TNT explosive source. As we can see, the frequency spectrum of an explosive source of underwater sound varies from 10 kHz to 100 kHz, when the TNT mass is from 0.1 kg to 1 kg.

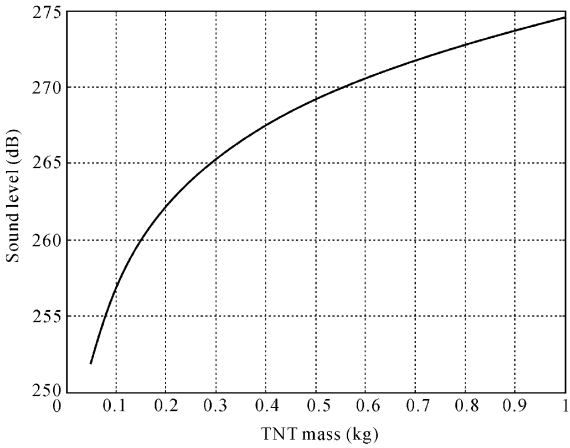


Fig. 4.87. Peak sound pressure of TNT explosive

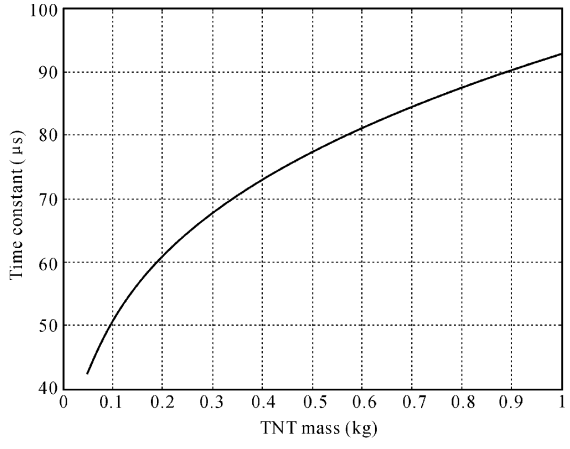


Fig. 4.88. Time constant of TNT explosive sound

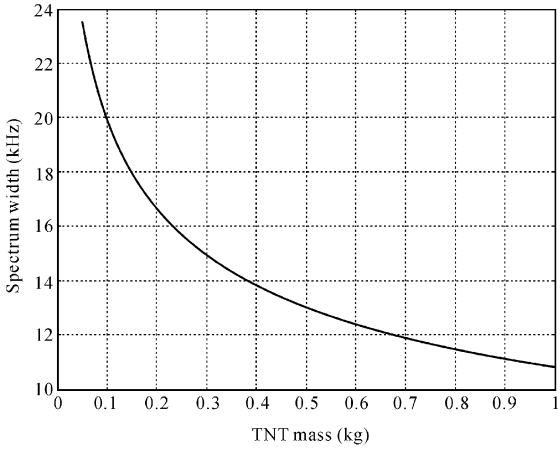


Fig. 4.89. Spectrum width of TNT explosive sound

The fine structure of the wave form of an explosive source depends on many parameters, including the depth of the explosive material, the sea surface and the sea floor. A bubble is quite often seen after the peak value of explosive sound. Arone showed that the time interval between the shock wave and the first bubble pulse[101] is

$$T = \frac{2.11 \times W^{1/3}}{(D+10)^{5/6}} \tag{4.172}$$

where W is the mass of TNT in kg and D is the depth of the explosion in m.

Fig. 4.90 shows an example of an experimental result where the TNT mass used was 50 g. The time wave form and frequency analysis results are plotted.

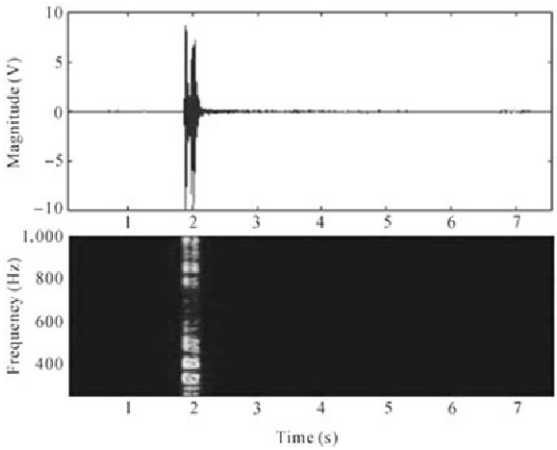


Fig. 4.90. Time / frequency records of 50 g TNT explosive (3 km, depth of receiver and explosive was 10 m)

4.10.2 Measurement of Transmission Loss by Using Explosive Source

The transmission loss of the underwater acoustic signal in an underwater acoustic channel is the basic data for sonar design. It strongly depends on the ocean environment. There is almost no theoretical method to precisely predict the law of transmission loss. The only effective method is to carry out at-sea experiments, and an explosive source of underwater sound is sometimes the only option. The hand grenade and specific signaling bomb can be used for the measurement of transmission loss. The signaling bomb can explode at a pre-set depth, making it easy to manage the relationship between the receiver and the explosive source.

Fig. 4.91 illustrates an experimental scenario for transmission loss measurement.

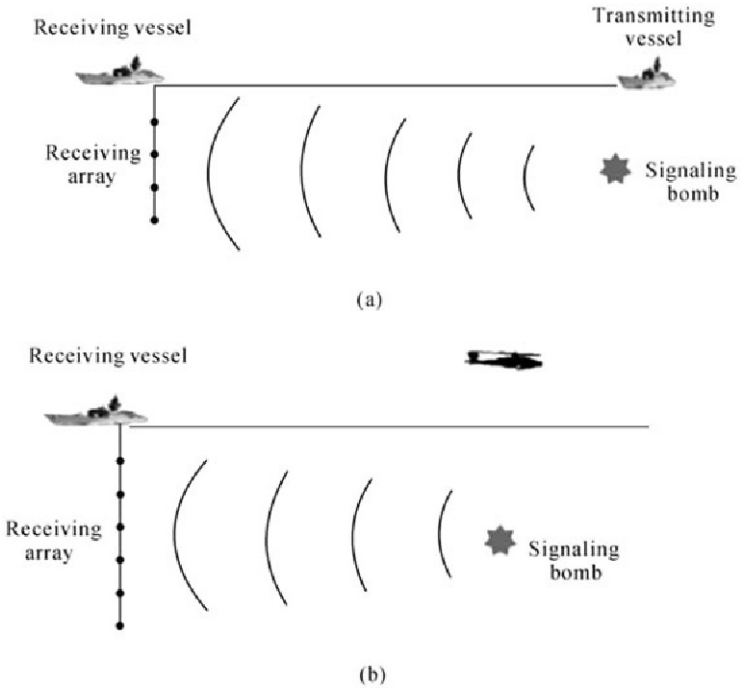


Fig. 4.91. Underwater acoustic propagation experiment using a signaling bomb. (a) Deploy signaling bomb from vessel; (b) Deploy signaling bomb from helicopter

We can use a transmitting vessel to deploy a hand grenade or signaling bomb. A helicopter used in place of a transmitting vessel can speed up the experiment procedure and give a closely synchronized transmission loss law.

The distances between the measurement points and the receiving vessel could be: 50, 100, 200, 400, 800, …, 10,000 m. They are in a straight line in logarithmic scale. Fig. 4.92 shows an example.

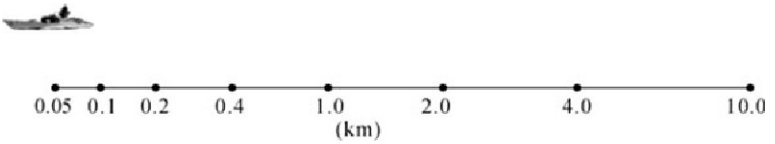


Fig. 4.92. Measurement point of transmission loss (not scaled)

The transmission loss should be calculated in different frequency bins. A rule frequently applied is to take 1/3 octave scale. If f_n is the n-th frequency bin, then

$$f_{n+1} = 2^{1/3} f_n \tag{4.173}$$

The band width of the n-th frequency bin is about

$$\Delta f_n = 0.231 \times f_n \tag{4.174}$$

Table 4.11 shows the values of the 1/3 octave frequency band. The first frequency is chosen to be 100 Hz. The frequency value will be repeated when it reaches 1,000 Hz, and the subsequent values should be 1,250, 1,600 Hz, and so on.

Table 4.11 1/3 Octave frequency band

Frequency bin (Hz)	Band width (Hz)
100	23
125	28
160	37
200	46
250	57
315	72
400	92
500	115
630	145
800	184

References

[1] Urick, R. J.: Sound Propagation in the Sea. Peninsula Pub., Los Altos (1982)
[2] Ziomek, L. J.: Underwater Acoustics: a Linear Systems Theory Approach. Academic Press, Orlando (1985)
[3] Filippi, P. D.: Acoustics: Basic Physics, Theory and Methods. Academic Press, Orlando (1998)
[4] Kuperman, W. A., Lynch, J. F.: Shallow water acoustics. Physics Today, 57-1010, pp. 55-61 (2004)
[5] Johnson, D. H., Dudgeon, D. E.: Array Signal Processing: Concepts and Techniques. Prentice Hall, Englewood (1993)

[6] Crocker, M. J.: Handbook of Acoustics. Wiley, New York (1998)
[7] Harris, C. M.: Handbook of Acoustical Measurements and Noise Control, 3rd Edn. McGraw-Hill, New York (1991)
[8] Brekhovskikh, L. M.: Fundamentals of Ocean Acoustics, 2nd Edn. Springer-Verlag, New York (1991)
[9] Coates, R. F. W.: Underwater Acoustic Systems. Halsted Press, New York (1989)
[10] Hassab, J. C.: Underwater Signal and Data Processing. CRC Press, Boca Raton (1989)
[11] Cox, A. W.: Sonar and Underwater Sound. Lexington Books, Lexington (1974)
[12] Urick, R. J.: Ambient Noise in the Sea. Penisula Pub., Wisconsin (1988)
[13] Thorp, W. H.: Analytic description of the low-frequency attenuation coefficient. J. Acoust. Soc. Amer., 42, pp. 270-275 (1967)
[14] Eckart, C.: The theory of noise in continuous medium. J. Acoust. Soc. Amer., 25, pp. 195-201 (1953)
[15] Wenz, G. M.: Acoustic ambient noise in the ocean: spectra and sources. J. Acoust.Soc.Amer., 34, pp. 1936-1956 (1962)
[16] Cron, B. F., Sherman, C. H.: Spatial correlation function for various noises models. J.Acoust.Soc.Amer. Vol. 34, pp. 1732-1738 (1962)
[17] Jacobson, M. J.: Space time correlation in spherical and circular noise fields. J. Acoust. Soc. Amer., 34, pp. 971-978 (1962)
[18] Pieng, T. S., Beng, K. T., Venugopalan, P., Chitre, M. A.: Development of a shallow water ambient noise database. In: Proc. Underwater Technology 2004 (2004)
[19] Andrew, R. K., Howe, B. M., Mercer, J. A.: A comparison of ocean ambient sound levels after 30 years for a coastal site off California. J. Acoust. Soc. Amer., 109, pp. 2386-2387 (2001)
[20] Baggeroer, A. B., Scheer, E. K., Colosi, J. A., Cornuelle, B. D., Dushaw, B. D., Dzieciuch, M. A., Howe, B. M., Mercer, J. A., Spindel, R. C., and Worcester, P. F.: Noise field statistics and coherence on the NPAL array. J. Acoust. Soc. Amer., 109, p. 2386 (2001)
[21] Fisher, F. H., Simmons, V. P.: Sound absorption in sea water. JASA, 62, pp. 558-562 (1977)
[22] Wei, R. C., Chen, C. F., Newhall, A.E., Lynch, J. F., Duda, T. F., Liu, C. S., Liu, P. C.: A preliminary examination of the low-frequency ambient noise field in the south china sea during the 2001 ASIAEX experiment. IEEE Journal of Oceanic Engr. (2001)
[23] Ross, D.: Mechanics of Underwater Noise. Peninsula Pub., Los Altos (1976)
[24] Dowet, M. T., Emling, J. W., Knudsen, V. O.: Survey of Underwater Sound Report No.4. US Navy Report, San Diego (1944)
[25] Stewart, J. L., Westerfield, E. C.: Theory of active sonar detection. In: Proc. IRE, 47, pp. 872-881 (1959)
[26] Middleton, D.: A statistical theory of reverberation and similar first order scattered fields. IEEE Trans., IT-13, pp. 372-414 (1967)

[27] Gingras, D. F.: Robust broadband matched field processing: performance in shallow water. IEEE J. of Oceanic Engr. 18-3, pp. 253-264 (1993)
[28] Poor, H. V.: Uncertainty tolerance in underwater acoustic signal processing. IEEE J. of Oceanic Engr., 12, pp. 48-65 (1987)
[29] McTaggart, B.: Thirty years of progress in sonar transducer technology. In: Proc. of UDT'91 Paris, pp. 3-11 (1991)
[30] Ma, M. T.: Theory and Application of Antenna Arrays. Wiley, New York (1974)
[31] Fisher, F. H. and Simmons, V. P.: Sound absorption in sea water. JASA Vol. 62, pp. 558-562 (1977)
[32] Van Veen, B. D., Buckley, K. M.: Beamforming: a versatile approach to spatial filtering. IEEE ASSP Magazine, 5-2, pp. 4-24 (1988)
[33] Frank, T. H., Kesner, J. W., Gruen, H. M.: Conformal array beam patterns and directivity indices. J. Acoust. Soc. Amer., 63, pp. 841-847 (1978)
[34] Steinberg, B. D.: Principles of Aperture and Array System Design. Wiley, New York (1976)
[35] Andeason, V. C.: Directivity of spherical receiving arrays. J.Acoust. Soc. Amer., 35, pp. 1162-1168 (1963)
[36] Dolph, C. L.: A current distribution for broadside arrays which optimize the relationship between beam width and side-lobe level. In: Proc. IRE, 34, pp. 335-348 (1946)
[37] Villeneuve, A. T.: Taylor patterns for discrete arrays. IEEE Trans., AP-2, pp. 1089-1093 (1984)
[38] Taylor, T. T.: Design of line source antenna for narrow beamwidth and low sidelobes. IRE Trans., AP-3, pp. 16-28 (1955)
[39] Elliott, R. S.: Antenna Theory and Design. Prentice Hall, Englewood (1981)
[40] Godara, L. C.: The effect of phase shifter errors on the performance of an antenna array beamformer. IEEE J. Oceanic Engr., 10, pp. 278-284 (1985)
[41] Tang, W. L., Overview of self-noise measurement of linear towed array, Technical Report, Harbin Engineering Univ. (1998)
[42] Lytle, D. W.: Space time correlation function for surface noise model. J. Acoust. Soc. Amer., 39, pp. 587-590 (1966)
[43] Curtis, T. E., Ward, R. J.: Digital beam forming for sonar systems. In: IEEE Proc. Part F, 127-4, pp. 257-265 (1980)
[44] Wilson, O. B.: Introduction to the Theory and Design of Sonar Transducers. Penisula Pub., Wisconsin (1988)
[45] Sullivam, E. J., Middleton, D.: Estimation and detection issues in matched field processing. IEEE J. of Oceanic Engr., 18-3, pp. 156-167 (1993)
[46] Baggerroer, A. B., Kuperman, W. A., Mikhalevsky, P. N.: An overview of matched field method in ocean acoustics. IEEE J. of Oceanic Engr., 18, pp. 379-387 (1993)
[47] Munk, W., Worcester, P., Wunsch, C.: Ocean Acoustic Tomography. Cambridge Univ. Press, New York (1995)
[48] Zhou, J. X., Zhang, X. Z.: Resonant interaction of sound wave with internal solitons in the coastal zone. J. Acoust. Soc. Amer., 90, pp. 2042-2054 (1991)

[49] Cox, H., Zeskind, R. Owen, M.: Robust adaptive beamforming. IEEE Trans., ASSP-35, pp. 1365-1376 (1987)
[50] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A.: Robust Statistics: the Approach Based on Influence Function. Wiley, New York (1986)
[51] Wilmut, M. J., Ozard, J. M., Keefe, K. O., Musil, M.: A piecewise matched field tracking algorithm. IEEE J. of Oceanic Engr., 23-3, pp.167-173 (1998)
[52] Candy, V. C.: Model-Based Signal Processing. IEEE Press, Wiley, New York (2006)
[53] Kuperman, W. A., D'Spain, G. L. (eds): Ocean Acoustic Interference Phenomena and Signal processing. In: AIP Conference Proceedings (2001)
[54] Levitus, S., Antonov, J. I., Boyer, T. P., Stephens, C.: Warming of the world ocean. Science, 287, pp. 2225-2229 (2000)
[55] Mercer, J. A.: Report on the second North Pacific acoustic laboratory (NPAL) data analysis workshop. Leavenworth, pp.16 (2000)
[56] Munk, W.: Scattering into the shadow zone. J. Acoust. Soc. Amer., 109, p. 2386 (2001)
[57] Worcester, P. F.: Report on the third North Pacific acoustic laboratory (NPAL) Data Analysis Workshop. Chicago, March, p. 18 (2001)
[58] Worcester, P. F., Howe, B. M., Mercer, J. A., Dzieciuch, M. A.: A comparison of long-range acoustic propagation at ultra-low (28 Hz) and very-low (84 Hz) frequencies. In: Proc. of the US-Russia Workshop on Experimental Underwater Acoustics, Talanov, V.I. (eds), Russian Academy of Sciences, St. Petersburg, pp. 93-104 (1999)
[59] Spindel, R. C. et al.: Results from the acoustic thermometry of ocean climate (ATOC) experiment. In: Proc. of the US-Russia Workshop on Experimental Underwater Acoustics, Talanov, V. I. (ed.), Russian Academy of Sciences, St. Petersburg, pp. 23-32 (1999)
[60] Colosi, J. A., Tappert, M. Dzieciuch, M.: Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the eastern North Pacific Ocean. J. Acoust. Soc. Amer., 110, pp. 163–169 (2001)
[61] Schmidt, H., Lee, J.: Physics of 3D scattering from rippled seabeds and buried targets in shallow water. J. Acoust. Soc. Amer., 105, pp. 1605–1617 (1999)
[62] Hinich, M. J.: Maximum-likelihood processing for a vertical array. J. Acoust. Soc. Amer., 54-2, pp. 499-503 (1972)
[63] Fink, M.: Time-reversed Acoustics. In: Scientific American, pp. 91-97 (1999)
[64] Song, H. C., Kim, S., Hodgkiss, W. S., Kuperman, W. A.: Environmentally adaptive reverberation nulling using a time reversal mirror. JASA, 116, pp. 763-768 (2004)
[65] Edelmann, G. F., Hodgkiss, W. S., Akal, T., Kuperman, W. A., Song, H. C.: An initial demonstration of underwater acoustic communication using time reversal. IEEE JOE, 27, pp. 602-609 (2002)
[66] Roux, P., Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T., and Stevenson, M.: A non-reciprocal implementation of time reversal in the

ocean. JASA, 116, pp. 1009-1015 (2004)
[67] Lingevitch, J. F., Song, H. C., Kuperman, W. A.: Time reversed reverberation focusing in a waveguide. J.Acoust. Soc. Amer., 111, pp. 2609-2614 (2002)
[68] Rouseff, D., Fox, W. L. J., Jackson, D. R., Jones, C. D., Ritcey, J. A.: Underwater acoustic communication by passive phase conjugation: theory and experimental results. IEEE JOE, 26, pp. 821-831 (2001)
[69] Baggeroer, A. B.: Acoustic telemetry: an overview. IEEE J. Oc. Acoustics, 9-4 (1984)
[70] D'Spain, G. L., Kuperman, W. A.: Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth. JASA, 106 -5, pp. 2454-2468 (1999)
[71] Yang, T. C., Yoo, K., Siderius, M.: Internal waves and its effect on signal propagation in the adventure bank. In: Proc. of 8th International Congress on Sound and Vibration, Hong Kong, pp. 3001-3008 (2001)
[72] Munk, W. H., Spindel, R. C., Baggeroer, A., Birdsall, T. G.: The heard island feasibility test. J. Acoust. Soc. Amer., 96, pp. 2330–2342 (1994)
[73] Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T., Ferla, C., Jackson, D. R.: Phase conjugation in the ocean: experimental demonstration of an acoustic time-reversal mirror. J. Acoust. Soc. Amer., 103-1, pp. 25-40 (1998)
[74] Edelmann, G. F., Hodgkiss, W. S., Akal, T., Kuperman, W. A., Song, H. C.: An initial demonstration of underwater acoustic communications using time reversal. IEEE J. Oceanic Eng., 27, pp. 602–609 (2002)
[75] Hodgkiss, W. S., Song, H. C., Kuperman, W. A., Akal, T., Ferla, C., Jackson, D. R.: A long range and variable focus phase-conjugation experiment in shallow water. J. Acoust. Soc. Amer., 105, pp. 1597–1604 (1999)
[76] Jackson, D. R., Dowling D. R.: Phase conjugation in underwater acoustics. J. Acoust. Soc. Amer., 89, pp. 171-181 (1991)
[77] Fink, M.: Time reversed acoustics. Scientific American, pp. 91-97 (1999)
[78] Widrow, B., Glover, J. R., McCool, J. M., Kaunitz, J.: Adaptive noise cancelling: principles and applications. In: Proc. IEEE, 63-12, pp. 1692-1716 (1975)
[79] Spiess, A., Kupperman, W.: The marine physical laboratory in Scripps. Oceanography, 16-3, pp. 45-54 (2003)
[80] Etter, P. C.: Underwater Acoustic Modeling: Principles, Techniques and Applications, 2nd Edn. Spon Press, New York (1996)
[81] Beranek, L. L., Ver, I. L.: Noise and Vibration Control Engineering: Principles and Applications. Wiley, New York (1992)
[82] Boyles, C. A.: Acoustic Waveguides: Applications to Oceanic Science. Wiley, New York (1984)
[83] Edgerton, H. E.: Sonar Images. Prentice-Hall, Englewood (1986)
[84] Frisk, G. V.: Ocean and Seabed Acoustics: A Theory of Wave Propagation, Facsimile Edn. Prentice Hall, Englewood (1994)
[85] Kerman, B. R.: Sea Surface Sound: Natural Mechanisms of Surface Generated Noise in the Ocean. Kluwer Academic Publishers, Dordrecht (1988)

[86] Kuperman, W. A., Jensen, F. B.: Bottom-Interacting Ocean Acoustics. Plenum Press, New York (1980)
[87] Tolstoy, A.: Matched Field Processing for Underwater Acoustics. World Scientific, Singapore (1993)
[88] Apel, J. R., Badiey, M., Chiu, C. S., Finette, S., Headrick, R.: An overview of the 1995 SWARM shallow water internal wave acoustic scattering experiment. IEEE J. Oceanic Eng., 22, pp. 465-500 (1997)
[89] Yang, T. C.: Temporal fluctuations of broadband channel impulse functions and underwater acoustic communications at 2-5 kHz. In: Proc. of MTS OCEANS 2002, Biloxi, 4, pp. 2395-2400 (2002)
[90] Rouseff, D., Turgut, A., Wolf, S. N., Finette, S., Orr, M. H., Pasewark, B. H., Apel, J. R., Badiey, M., Chiu, C. S., Headrick, R. H., Lynch, J. F., Kemp, J. N., Newhall, A. E.: Coherence of acoustic modes propagating through shallow water internal waves. J. Acoust. Soc. Amer., 111, pp. 1655-1666 (2002)
[91] Yang, T. C.: Temporal resolution of time-reversal and passive-phase conjugation for underwater acoustic communications. IEEE J Oceanic Eng., 28, pp.229-245 (2003)
[92] Knudson, V. O., Alford, R. S. and Emling, J. W.: Underwater ambient noise, J. Mar. Res., 7, pp. 410-414 (1948)
[93] National Research Council (Eds): Technology for the United States Navy and Marine Corps. In: Becoming a 21st-Century Force, NA Press, pp. 2000-2035 (1997)
[94] Miasnikov, E.: Can Russian strategic submarine survive at sea? In: The fundamental limits of passive acoustics. Science & Global Security, 4, pp. 213-251 (1994)
[95] Introduction to Naval Weapons Engineering. Military Analysis Network, http://www.fas.org/man/dod-101/navy/docs/. Accessed 1 Jan 2011
[96] Mellen, R. H.: Attenuation in surface duct. J. Acoust. Soc. of Amer., 63, pp. 1624-1628 (1978)
[97] Wang, D. Z., Shang, E. C.: Underwater Acoustics. Sciences Pub., Beijing (1984)
[98] Winder, A. A.: Sonar system technology. IEEE Trans., SU-20, pp. 291-332 (1975)
[99] Arons, A. B., Yennie, D. R., Cotter, T. P.: Long range shock propagation in underwater explosion phenomena II. US Navy Dept. Bur. Ord., NAVORD Report (1949)
[100] Arons, A. B.: Secondary pressure pulse due to glass globe oscillation underwater explosions Ⅱ. J. Acoust. Soc. Amer., 20, pp. 277 (1948)
[101] Acoustic Toolbox, URL: http://stommel.tamu.edu/-baum/linuxlist/node7.html. Accessed 1 Jan 2011
[102] Matlab Front-end for Acoustic Toolbox, URL: http://www.curtin.edu.au/curtin/centre/cmst/products/actoolbox/. Accessed 1 Jan 2011
[103] Porter, M. B., Reiss, E. L.: A numerical method for ocean-acoustic normal modes, J. Acoust. Soc. Amer., 76, pp. 244-152 (1984)

[104] Porter, M. B., Reiss, E. L.: A numerical method for bottom interacting ocean normal modes. J. Acoust. Soc. Amer., 77, pp. 1760-1767 (1985)
[105] Porter, M. B.: A KRAKEN normal mode program. Memorandum rept., SM-245, Saclantcen, Italy (1991)
[106] Porter, M. B., Bucker, H. P.: Gaussian beam tracing for computing ocean acoustic fields. J. Acoust. Soc. Amer., 82, pp. 1349-1359 (1987)
[107] Green, T. J.: Robust passive sonar. DAPAR Tech. 2000, Dallas (2000)

5

Design of Digital Sonar

The design of digital sonar is related to a wide knowledge of many subjects in science and technology. These include aspects of oceanography, platform noise, sensor technology, acoustic materials, the optimum extraction of weak signals, signal processing, the architecture of computer systems, human / machine engineering and artificial intelligence, etc. Therefore the sonar designer has to be familiar with these subjects and use them as pre-requisites in the design of good digital sonar[1-12].

On the other hand, the design philosophy is changing with time. This is because military / civilian requirements are changing; some new topics need new theories and techniques. And achieving a solution methodology to deal with these issues becomes our motivation [13-37].

The theory of digital sonar signal processing is a part of the entire theory of digital signal processing. In the development of sonar design, it is natural to absorb a lot of the achievements in related sciences and technology, for example, radar, aerospace technology, speech signal processing, digital television, telecommunications, image signal processing, medical signal processing and software engineering, etc. The development of digital sonar pushes at the frontiers of research in these technical fields[38-70].

5.1 System Architecture of Digital Sonar

From the viewpoint of the hardware structure, digital sonar is similar to a specific computer. However, from the design point of view, it is totally different, since the design of a general computer is task-oriented. The computer must complete some tasks that the user needs. The architecture of a computer system has to satisfy this requirement, i.e., the design of hardware and software must guarantee reliable completion of this task. But digital sonar is different. It is necessary to input / output signals in real time. Whatever is integrated or distributed, must work in a unified operating system. Therefore digital sonar has its specific system architecture.

5.1.1 Parallel Operation and Sequential Operation

Any algorithms adopted in digital sonar have to be completed in real time. There are two kinds of basic operation schemes that can be used in real time operation, parallel operation and sequential operation. The parallel operation is actually batch data processing, to process one batch of data and obtain a batch output. In a sequential operation, one input results in one output. Sometimes we also call multiple channels processing parallel processing. However, in this situation it is necessary to note the meaning is quite different.

Fig. 5.1 presents the block diagram of system architecture of passive sonar. We can see that most parts of operation are digital, but there are still a few analogue operations. Usually, the pre-amplifier is analogue. And the auditory system is analogue. The self failure testing can be digital and analogue. If we use the sonar simulator, it often has both analogue and digital output. The A / D conversion after S / H (sample and hold) can be parallel or sequential. It depends on the design and should affect the time delay line and beamforming.

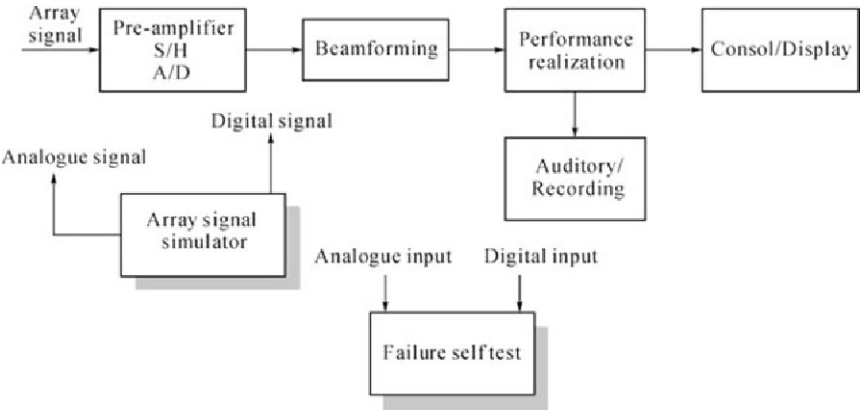


Fig. 5.1. System architecture of passive sonar

Example

If the system has 64 elements, the sampling frequency is 20 kHz and the 64-channel signals are sampled synchronously. The quantization error resulting from the sample is 50 μs / 2 = 25 μs. We can also use an asynchronous sample, and then the sampling frequency should be 64×20 kHz = 1.28 MHz. The sampling operation for each channel does not happen at the same time; the time difference of the adjacent channel is 50 μs / 64 = 0.78125 μs. When we make beamforming, this time difference should be taken into account in the time compensation.

The algorithm for beamforming can be sequential processing or batch processing; it depends on which kind of beamforming method is chosen. If the operation of beamforming is in time domain, we should adopt a sequential processing technique. If the operation is in frequency domain, we should adopt batch

processing. In order to get multiple beamforming in real time, parallel real time processing is necessary. For the realization of a specific function of digital sonar, both sequential processing and batch processing can be used. There are no serious rules for what kind of processing should be adopted for some sonar. Fig. 5.2 shows roughly the division between applications of sequential processing and batch processing in digital sonar system.

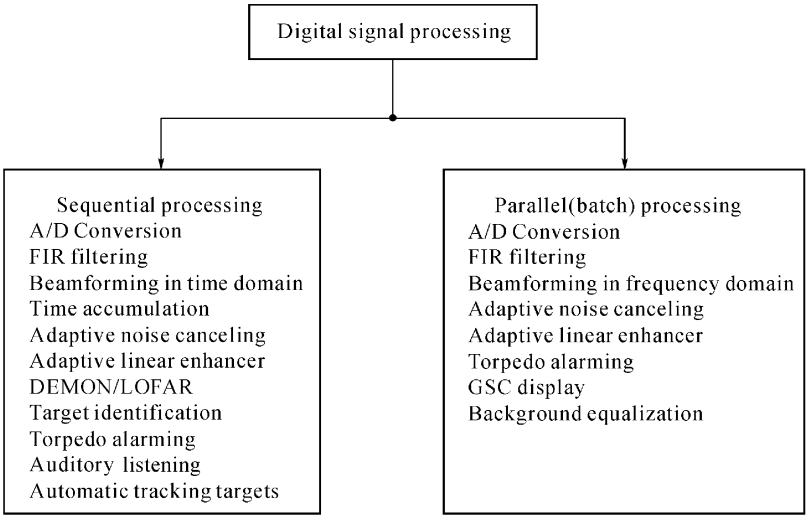


Fig. 5.2. Applications of sequential and parallel processing in sonar

Sequential processing includes A / D conversion, FIR filtering, beamforming in time domain, time accumulation, adaptive noise canceling, adaptive or linear enhancer, DEMON / LOFAR, target identification / classification, torpedo alarming, auditory listening and automatic tracking of targets, etc. Batch processing includes A / D conversion, FIR filtering, beamforming in frequency domain, adaptive noise canceling, adaptive linear enhancer, torpedo alarming, gray scale conversion (GSC) display and background equalization, etc.

5.1.2 Differences in Beamforming in Time Domain and Frequency Domain

In digital sonar, batch processing is mainly used in beamforming in frequency domain. This is because it is convenient to use the FFT algorithm in beamforming, and the FFT algorithm is typical batch processing. The advantage in calculation speed of FFT is the main reason for adopting this technique. But it also brings about some problems, which need to be taken account in sonar design. In Chapter 2, the discrete Fourier transform (DFT) is already discussed, now we are going to

show how it can be used in beamforming and can considerably speed up the calculation. Fig. 5.3 presents the block diagram of beamforming in time domain and frequency domain; it describes a sonar system with N elements and M beams.

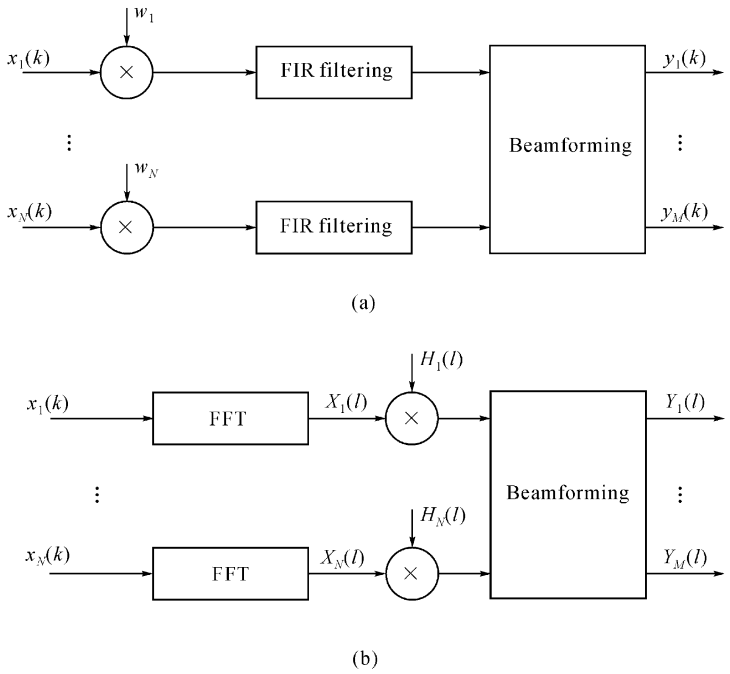


Fig. 5.3. Application of sequential and parallel processing in beamforming. (a) Block diagram of beamforming in the domain; (b) Block diagram of beamforming in frequency domain

Let us firstly consider beamforming in time domain. The i-th input signal is $x_i(k)$ $(i = 1, \ ..., \ N)$. After weighting w_i, it becomes $w_i x_i(k)$. Suppose the FIR filter is a linear phase filter of length L, the coefficients are $h(0), \ ..., \ h(L-1)$. Therefore the output of FIR filter of the i-th input signal is

$$u_i(k) = \sum_{l=0}^{L-1} w_i x_i(k-l)h(l) \quad (k \geq L-1) \tag{5.1}$$

In order to steer the beam in a certain direction, it is necessary to make a time delay, indicated by $\tau(i, \ m)$, where m is the beam number. Usually, the value of the time delay $\tau(i, \ m)$ is the integer time of sampling duration. We obtain

$$y_m(k) = \sum_{i=1}^{N} u_i[k - \tau(i, \ m)] \quad (m = 1, \ ..., \ M) \tag{5.2}$$

This is the expression in time domain of beamforming. It is worth showing that the calculation of the time delay $\tau(i, m)$ depends on the shape of the sonar array.

Now we turn to frequency domain beamforming. As we can see in Eq. (5.2), the key operation is discrete convolution. The convolution operation in time domain is equal to the multiple operation in frequency domain. There is one problem we have to resolve. That is the DFT can only realize the cycle convolution and beamforming in time domain is needed to complete the real convolution. The cycle convolution and the real convolution are not equivalent. In many sonar systems, in addition to beamforming, it is also necessary to complete other processing for beam data, therefore we need to transfer the frequency data back to the time series. That means we have to perform the IDFT (inverse DFT). In other words, whenever we make time / frequency or frequency / time transformation, it is necessary to guarantee the order of the time data series and we can use the time data series, if necessary.

Suppose the input data series is $x(0), \ldots, x(N-1)$. Assume $w(0), \ldots, w(N-1)$ is the weighting series. We want to complete the convolution operation

$$y(k) = \sum_{i=0}^{N-1} w(i)x(k-i) \ (k \geq N-1) \tag{5.3}$$

Eq. (5.3) can not be directly used to do DFT, since it is not the form of cycle convolution. Define the following cycle convolution equation

$$u(k) = \sum_{i=0}^{N-1} w(i)x(k-i) \ (k = 0, \ldots, N-1) \tag{5.4}$$

For $k - i < 0$, $x(k-i) = x(N+k-i)$. It means that the series data $x(0), \ldots, x(N-1)$ is extended to a new series with length $2N$, and the first N point values are the same as $x(0), \ldots, x(N-1)$. Denote the DFT of $x(k)$, $w(k)$, and $u(k)$ as $X(l)$, $W(l)$, and $U(l)$ respectively, i.e.

$$X(l) = \sum_{i=0}^{N-1} x(i)\exp(-2\pi \mathrm{j} il / N) \ (l = 0, \ldots, N-1)$$

$$W(l) = \sum_{i=0}^{N-1} w(i)\exp(-2\pi \mathrm{j} il / N) \ (l = 0, \ldots, N-1)$$

$$U(l) = \sum_{i=0}^{N-1} u(i)\exp(-2\pi \mathrm{j} il / N) \ (l = 0, \ldots, N-1)$$

Based on Eq. (2.68), it is easy to show that

$$U(l) = X(l)W(l) \quad (l = 0, \ldots, N-1) \tag{5.5}$$

It is obvious that $u(N-1) = y(N-1)$. That is to say that in the series data of $u(i)$ and $y(i)$ only one point has the same value. Therefore, if we perform IDFT on $U(l)$, it will be unable to get $y(i)$ $(i = 0, \ ..., \ N-1)$. In other words, the output of batch processing is not the sample of sequential processing.

In order to solve this important problem, we will use the data of extended series. Denote

$$\hat{x}(k) : x(0), \ ..., \ x(2N-1)$$

The DFT of $\hat{x}(k)$ is defined as

$$\hat{X}(l) = \sum_{k=0}^{2N-1} \hat{x}(k) \exp(-2\pi \mathrm{j} kl / 2N)$$

It is necessary to pad N zero in the series of $w(0), \ ..., \ w(N-1)$ and construct a new weighting series

$$\hat{w}(k) : w(0), \ ..., \ w(N-1), \ 0, \ ..., \ 0$$

The DFT of $\hat{w}(k)$ is defined as

$$\hat{W}(l) = \sum_{k=0}^{2N-1} \hat{w}(k) \exp(-2\pi \mathrm{j} kl / 2N)$$

The convolution of $\hat{x}(k)$ and $\hat{w}(k)$ is

$$\hat{y}(k) = \sum_{i=0}^{2N-1} \hat{w}(i)\hat{x}(k-i) \ \ (k = 0, \ ..., \ 2N-1)$$

It is easy to verify that $y(k) = \hat{y}(k)$, for $k \geq N$. That means the latter half points of $\hat{y}(k)$ are the values of $y(k)$ in $N, \ ..., \ 2N-1$. Therefore, if we perform the IDFT on $\hat{Y}(l)$ and throw away the first half points of the results, we obtain the necessary N point values of $y(k)$. Series $\hat{x}(k)$ is an extension of series $x(k)$; each batch has $2N$ points, but only N points need to be renewed. It looks like more trouble than time domain processing, but since the speed of the FFT algorithm is much faster than in time domain, therefore the advantage of frequency domain processing is obvious. Some sonar prefers to adopt 5 times FFT to complete necessary sonar signal processing in frequency domain and then switch between time / frequency domains[71-73].

5.1.3 Data Interchange among Sub-systems

There are two basic data interchange methods. The first method is based on a standard bus, such as VME (versa module Europe) bus, PCI (personal computer interface) bus, etc. The advantage of applying a standard bus is the convenience of the hardware extension and software compiler. The second method is based on a non-standard bus. The method of interchange of data depends on the architecture of the hardware of the signal processor, and the design of one's own channel to realize data interchange. The advantage is that the hardware / software design is relatively simple and the hardware cost is relatively low.

The selection of data interchange and transmission methods depends on the requirements of the sonar system. In some cases one can use network cards to complete data transmission between sub-systems; i.e., the standard bus is used inside the data transmission and the network card is used between sub-systems. This kind of data interchange and transmission is particularly suitable for data transmission between signal processors and consol / display systems and data communication between multiple displays, since in a modern large sonar system, in order to guarantee the reliable operation of the sonar system, there is often more than one display. For example, there can be one main display and one auxiliary display.

5.1.4 Real Time Data Display of the Sonar System

Fig. 5.4 shows the real time data display system for a modern sonar. As we can see, the operator and multi-processor are in a central position. We can consider that the display system of the sonar is at the interface of operator and sonar equipment. A modern sonar tends to have two massive data bases. One is the information data base, which stores the data of the ocean environment, sound propagation, image of the sea bottom and radiated noise of one's own and enemy vessels. The other one is the tactics database, which stores tactic decision rules, environmental characteristics, friendly / hostile counter measure data, etc. These two databases will play a very important role in target tracking, target identification / classification, torpedo alarming, range prediction and strategy of vessel maneuvering, etc.

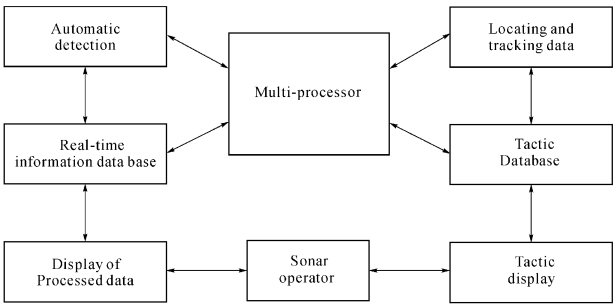


Fig. 5.4. Display of sonar real time data

5.1.5 COST Technology

COST (commercial off the shelf) technology has been extensively used in sonar design and manufacture, because of the rapid development of the micro processor industry in civil applications. As military equipment, sonar may utilize more and more COST technology and therefore one can reduce the manufacturing time and costs and increase reliability and maintainability. Some countries already propose that the utilization of COST technology is one of the conditions of equipment procurement. Fig. 5.5 presents the relationship between hardware and software of modern sonar. The central part of a sonar system is the hardware platform, and all necessary software has to be operated on this platform. Thus the desired functions of the sonar system can be realized.

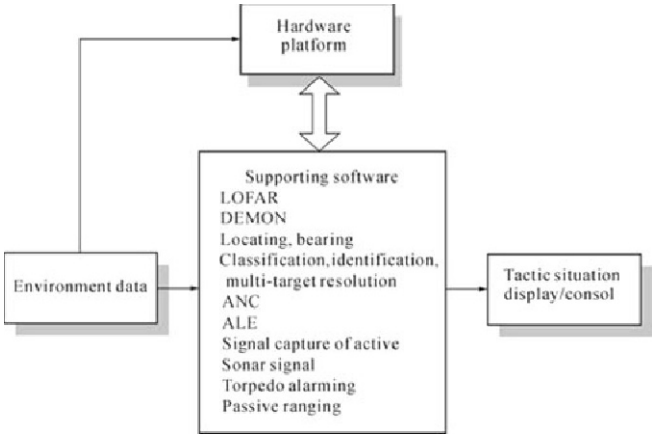


Fig. 5.5. Configuration of software and hardware for a modern sonar system

For a sonar system, whether newly designed sonar or updated old sonar, the application of COST technology is a good strategy. It is reported that to update an old sonar, 50% of the budget can be saved and 95% performance can be obtained[74], compared to buying new equipment.

5.2 Parameter Selection of Programmable Digital Sonar

Digital sonar is a complicated system. It is not an easy job to design a good sonar system; the selection of various parameters is one of the most important factors. Of course a part of the parameters can be found in the sonar equation, but some parameters do not appear in the sonar equation, but only occur in the sonar mission analysis.

As a matter of fact, many parameters of digital sonar are mutually related and

share mutual constraints. Usually, an optimum design is a trade off and is robust enough for the possible change of parameters. In this section, we will introduce the parameters which may be appear in sonar design and present the criteria for selection. Typical examples are illustrated to show how we can design a good digital sonar, but it is worth noting that these are examples just to help the discussion, and by no means have to be followed.

The main parameters for a sonar designer are briefly listed in the following.

Array Shape and Number of Elements of the Wet End

The array shape of a sonar wet end is defined by the installation platform. Any linear array, circle array, flank array, spherical array, towed array can be used in a sonar system. The number N of array elements is defined by the aperture of the array. One criterion that has to be obeyed is that the distance d of adjacent elements must be approximately equal to half of the wave length λ, i.e., $d \approx \lambda / 2$. Here λ can be considered as the wave length of the central frequency of the operation frequency band of the sonar system. Generally speaking, one prefers to adopt equally spaced array architecture, since the un-equally spaced array does not have any obvious advantage.

Baffling Form of Array

In order to reduce the interference from the platform and isolate the effect of the vibrations of a vessel's hull, it is necessary to have a baffle for the bow array and flank array. The baffle form can be categorized as being of sound absorption and sound reflection types. In most cases we would like to adopt a combination of these, i.e., to mount a layer of damping material in an air cavity. By means of this architecture we can reduce the signal distortion and suppress platform interference as much as possible. The directivity function of a single hydrophone in a baffling array must have a working sector from –60° to +60°, and the effect of interference suppression must be greater than 20 dB at 180°.

Frequency Band and Geometric Central Frequency

The frequency band of a sonar system is tightly related to the array size. Suppose the lower end frequency is f_1 and the higher end frequency is f_2. Then the working frequency band is $W = f_2 - f_1$. The definition of a central frequency has two different forms, that is an arithmetical central frequency f_{01} and geometrical central frequency f_{02}

$$\begin{aligned} f_{01} &= (f_1 + f_2)/2 \\ f_{02} &= \sqrt{f_1 f_2} \end{aligned} \tag{5.6}$$

We will adopt the geometrical central frequency f_{02} as the definition of the central frequency, if there are no specific explanations.

Example

Assume a cylindrical array with radius r = 2 m, frequency band 1 – 4 kHz; Therefore the central frequency is 2 kHz. The corresponding wave length is $\lambda = 0.75\,\text{m}$. Since the length of the circle is $L = 12.5664\,\text{m}$, so the suggested number of elements should be $N = L/(\lambda/2) \approx 33$. If the array is baffled, each beam has a working sector of 120° – 160°. Therefore we can choose the number of elements, which are taken into account in beamforming as 12 – 16.

Sampling Frequency

As we have shown in the previous section, for the N-channel signal, we can adopt sequential sampling with a much higher sampling rate or a parallel independent sampling scheme. Based on the Nyquist sample theorem, the sampling frequency f_s must at least equal two times the higher end frequency of the signal. Therefore, if the working frequency band of the sonar system is (f_1, f_2), then $f_s \geq f_2$. This is the minimum requirement. If we consider the compensation error of the time delay in beamforming, it is frequent to choose

$$f_s \geq 5 f_2 \tag{5.7}$$

The maximum compensation error of the time delay is $T_s/2$, where $T_s = 1/f_s$.

Example

Suppose the frequency band of the digital sonar is 100 – 1,200 Hz, then the sampling frequency can be chosen as $f_s = 5 \times 1,200 = 6,000$ Hz. The sampling duration is $1/6,000\ \text{s} \approx 166\ \mu\text{s}$.

A / D Conversion

The bit number of most modern sonar is 12 bits. If one needs a higher bit number, it is necessary to have perfect linearity of the pre-amplifier. From the point of view of the input dynamic range, a 2^{12}-layer architecture is acceptable. It is good enough to prevent signal distortion and gain loss.

FIR Filtering

It is reasonable to use a linear phase shift FIR filter, because this filter can guarantee there are no additional time delay compensation errors among channels. Suppose there is an N-point FIR filter (N usually is an even number). The

coefficient of filtering is $w(0), ..., w(N-1)$. For a linear phase shift filter

$$w(i) = w(N-1-i) \quad (i = 0, ..., N/2-1) \tag{5.8}$$

The relationship between input and output is

$$y(k) = \sum_{i=0}^{N-1} w(i)x(k-i) \tag{5.9}$$

Since $w(i)$ is symmetrical to $N/2$, therefore the calculation amount can be reduced to half of the original. In designing FIR filtering, it is necessary to note the relationship between passband and cut-off frequency. In digital sonar design, a suggested value for choosing cut-off frequency is 1.33 times the higher end frequency of the passband[2]. In terms of practical application, by using Parks algorithm, when $N \geq 64$, an FIR filter with desired performance can be obtained. Now there is much well established software, which can design the FIR filter easily.

Maximum Value of Taps of Time Delay Line

Maximum value of taps of time delay line is defined by the array aperture L and sampling duration T_s

$$N_0 = \mathrm{INT}\left[\frac{L}{cT_s}\right] + 1 \tag{5.10}$$

where c is the sound velocity and INT[·] indicates the arithmetic rounder operation.

Example

Assume an arc array with radius $\varphi = 3$ m. The working sector is 140° (Fig. 5.6). Then the physical aperture of this array is $L = r(1 - \cos 70°) = 0.984$ mm. If the sampling duration is $T_s = 20$ μs, then the number of taps should be $N_0 \approx 33$; i.e., the maximum time delay in beamforming is $20 \times 33 = 660$ (μs).

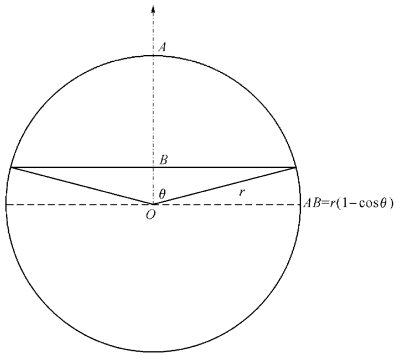


Fig. 5.6. Calculation of maximum time delay of arc array

Beamforming, Upsampling and Downsampling

Almost all modern digital sonar is a multi-beam system; the beam number depends on the subject requirement and the hardware cost. Firstly, it is necessary to calculate the ideal beam pattern, i.e., the directivity function without the error of time delay compensation. Then, to define how many beams we should complete in a beamforming system. Actually, the digital multiple beam is the discrete sampling for the beam pattern. At least we have to guarantee there are enough beams in the neighborhood of the main lobe. This is because we do not know where the target will present itself, therefore it is impossible for the target to be in the exact direction of the pre-beamforming direction. Fig. 5.7 illustrates this situation. Generally speaking, it is necessary to have 3 beams near the main lobe, i.e., $\Delta\theta_{3\,\mathrm{dB}}$. Therefore it is easy to estimate how many beams we need in 360°.

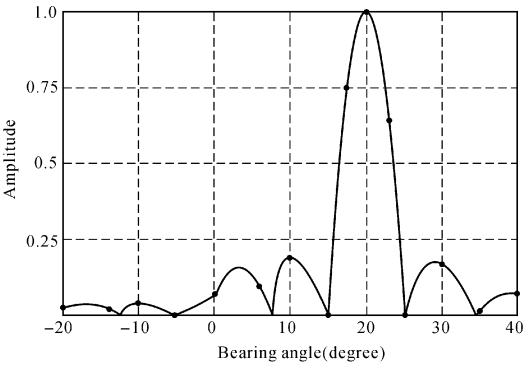


Fig. 5.7. Digital multi-beam: discrete sampling of continuous beam pattern

The selection of sampling frequency is another important problem in beamforming. Since the error in time delay compensation is equal to half of the sampling duration, therefore to increase the sampling frequency, if possible, is a method for improving the beamforming quality. Therefore, sometimes upsampling the input signal is necessary. However, of course, a high sampling frequency will result in an increase in hardware costs. In contrast to this, in the post-processing of beamforming, the data accumulation does not need a high sampling frequency, so we can decrease the sampling frequency while beamforming. This is called downsampling. The procedure of this kind of beamforming is illustrated in Fig. 5.8. We can consider it is a "fast in-low out" operation.

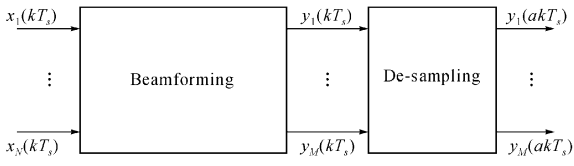


Fig. 5.8. De-sampling algorithm of digital multi-beam system

Beam Number

The number of beams is an important parameter for a sonar system. Usually one hopes the beam number will be as high as possible. But the beam number is subject to the hardware cost. A trade-off method is to interpolate the beam after beamforming. In the neighborhood of the main lobe, parabolic interpolation is a good algorithm for fining the beams, since it is proven that, in most cases, near the main lobe the shape of the directivity function can be approximated by parabolic curve. But in the side lobe, parabolic interpolation is without theoretical meaning.

For a circle array, the beam number is determined by

$$N_0 = 3 \times \frac{2\pi}{\Delta\theta_{3\,\mathrm{dB}}} \approx 53 \times \frac{r}{\lambda} \tag{5.11}$$

Similarly, for a linear array, the beam number is determined by

$$N_0 = 3 \times \frac{\pi}{\Delta\theta_{3\,\mathrm{dB}}} \approx 11 \times \frac{L}{\lambda} \tag{5.12}$$

Example

Suppose a circle array with $\lambda / r = 0.3$. Based on Eq. (5.11), the beam number for this circle array should be at least $N_0 \approx 53 / 0.3 \approx 176$.

Number of Auto-tracking Targets

The target tracking technique is used in passive sonar while a contact is occurring. In modern sonar, the most popular method to track targets is to manually input the target location and then transfer to auto-tracking; this is actually a half manual / half automatic tracking algorithm. If the target is lost in the auto-tracking procedure, it is necessary to manually input it again. The number of targets to be tracked can be up to 8 or 16, which depends on the requirements and hardware capability.

Full auto-tracking is possible for digital sonar, especially for torpedo alarming sonar. It is necessary to extract the trace of a fast-moving object and decide if it is a torpedo. This procedure combines the tracking and identification, and must be completed automatically.

Active Sonar Signal Capture

Signal capture is an important function of intercept sonar. It is different with echo signal detection. With the traditional energy detector it is impossible to complete this task. The frequency range of a captured signal must cover the possible frequency of hostile active sonar. And the entire capture procedure must be

automatic. The time / frequency domain analysis for a captured signal is necessary. The feature extraction is often based on the FFT analysis.

Target Classification and Identification

The performance of target classification / identification is one of the most important specifications of modern sonar. The correct classification / identification usually depends on a large and effective database. It is understandable that the more of target classes there are to be identified, the harder this should be. The basic requirement of target identification in anti-submarine warfare is the ablility to distinguish if a vessel is a surface ship or a submarine.

Failure Self Test

For the wet end, the failure self test must be at the hydrophone and transducer level and can also be displayed on the consol screen. For the dry end, the failure self test must at least be at printed board level, even at DSP chips level. The realization of a failure self test may depend on the hardware, software or a combination of both.

Reliability and Maintainability

The standard of reliability and maintainability is the main document which dominates the performance of the reliability and maintainability of a sonar system. Both the wet end and dry end should have the index of MTBF, MTBCF and MTTR. The quality of the specification depends on the complexity of the sonar system and environment in which it is used.

5.3 Pre-processing and FIR filtering

The signal pre-processing of a sonar system is also called signal conditioning. The goal of this process is to condition or modify the signal before input to multi-channel processing. Signal pre-processing consists of three parts, that is anti-aliazing filtering, automatic gain control (AGC) and A / D conversion (Fig. 5.9.) The implementation of signal pre-processing is not complicated, but it is located in the most forward part of a sonar system and, therefore, the importance of pre-processing is obvious.

The FIR filter is the most usable digital filter, since it satisfies the specific requirements of sonar signal processing.



Fig. 5.9. Signal Pre-processing

5.3.1 Anti-aliazing Filtering

An anti-aliazing filter is a low pass filter. The main function of this filter is to take away all frequency components in a signal which are higher than half of the sampling frequency. Therefore, there will be no frequency fold phenomena appearing in the sample process. Fig. 5.10 presents the frequency change of an anti-aliazing filter. The ideal frequency response of an anti-aliazing filter is of rectangular shape, i.e., the response from $0-f_p$ is one, otherwise zero. But, in practice, it is impossible. It usually has a transition band. In the figure it is $f_p - f_c$. The frequency f_p is called the upper limit frequency and f_c is called the cut-off frequency. There are some kind of fluctuations inside of the passband, i.e., $\pm\delta_1$. Outside of the passband there are also some fluctuations, i.e., $\pm\delta_2$.

Generally speaking, f_p, f_c and f_s satisfy the following relationship

$$f_s = f_p + f_c \tag{5.13}$$

and

$$k = \frac{f_c}{f_p} \tag{5.14}$$

is called the transition ratio. The typical value of k in underwater acoustics is 1.33[2].

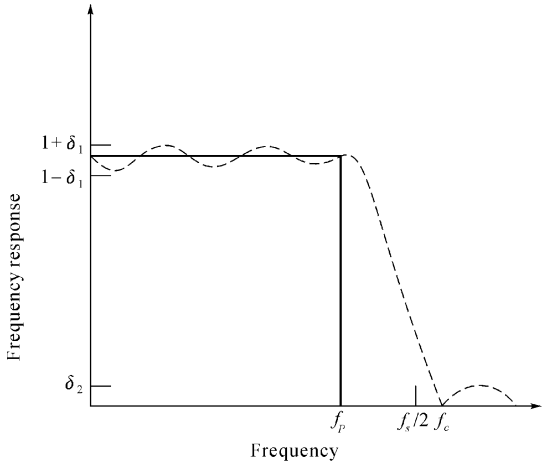


Fig. 5.10. Frequency response of anti-aliazing filter

5.3.2 A / D Conversion

For the flow chart of signal processing, the AGC is in front of A / D, but from the point of view of theoretical analysis, it is more convenient to describe A / D first. As we have shown in Chapter 2, the application of AGC has to sufficiently use the dynamic range of the A / D converter. If $V_{\max}$ is the maximum output level of the A / D converter, then

$$(AGC)_{\text{out}} = (AGC)_{\text{in}} \times k_0 = (AGC)_{\text{in}} \times \frac{V_{\max}}{k\sigma_x} \tag{5.15}$$

where k_0 is the gain of AGC, $k\sigma_x$ is the peak value of the input signal, σ_x is the mean square value of the input signal. If the input signal is Gaussian noise, then it is reasonable to take $k = 4$.

Suppose the input signal of the A / D converter is $x(nT_s)$, then the output of the A / D is

$$[x(nT_s)]_Q = x(nT_s) + e(nT_s) \tag{5.16}$$

where $[x(nT_s)]_Q$ indicates the output of the A / D converter and $e(nT_s)$ indicates the quantization error. $e(nT_s)$ is uniformly distributed in $[-\Delta/2,\ \Delta/2]$ and Δ is the quantization step, $\Delta = 2^{-n}$ (n is the bit number of quantization). Because $e(nT_s)$ is a uniformly distributed random variable, therefore its mean square value is

$$\sigma_e^2 = \Delta^2 / 12 = 2^{-2n} / 12 \tag{5.17}$$

SNR of quantization

$$SQNR = 10 \lg \frac{\sigma_x^2}{\sigma_e^2} \tag{5.18}$$

Assume the input of AGC is $x(t)$ and the output is $y(t)$, then

$$y(t) = \frac{x(t)}{k\sigma_x}$$

the variance of $y(t)$ is $\sigma_y^2 = 1/k^2$.

Replacing σ_x^2 in Eq. (5.18) by σ_y^2, it obtains

$$SQNR = 10\lg\frac{1}{k^2\sigma_e^2} \approx 6n + 10.8 - 20\lg k \tag{5.19}$$

Set $k = 4$, it obtains

$$SQNR = 6n - 1.24 \text{ (dB)} \tag{5.20}$$

This is just Eq. (2.16).

Example

Assume the bit number of the A / D converter is 8, except for the symbol sign, we have $n = 7$. Replacing it in Eq. (5.20), we obtain: $SQNR = 6\times7 - 1.24 = 40.76$ (dB).

The value of *SQNR* can be considered as the specification of the dynamic range, i.e., the minimum detectable sinusoidal signal embedded in the noise

$$DR = -10\lg\frac{\Delta^2}{12} = -10\lg\frac{2^{-2n}}{12} = 6n + 10.8 \text{ (dB)} \tag{5.21}$$

5.3.3 AGC and A / D Converter

The main goal of AGC is to adjust the amplitude of the input signal such that it satisfies the dynamic range of the A / D converter. In order to complete this function, it is necessary to choose a suitable time constant of the AGC circuit. At the same time it is necessary to choose the amplifier quantity such that the dynamic range of the A / D converter is sufficiently used. There is a simple method that can be used to calculate the amplifier times. Fig. 5.11 illustrates the AGC operation for signals.



Fig. 5.11. Principle of automatic gain control circuit

Suppose the input signal is a Gaussian random variable with probability density function $p(x) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left(-\frac{x^2}{2\sigma_x^2}\right)$. The total noise resulting from the A / D converter can be divided into two parts. One part is the quantization noise, i.e., when the input signal falls into the segment

$$n_0\Delta \le x(t) < (n_0 + 1)\Delta$$

and quantized as $n_0\Delta$, which will result in noise

$$\Delta = \frac{k\sigma_x}{2^n - 1}$$

where k is the amplifier times of AGC.

The second part of the noise is called saturation noise, when the input signal $x(t) \geq k\sigma_x$ is quantized as $(2^n - 1)\Delta$. This will result in noise, too. These two bits of noise can be calculated separately.

Quantization noise

$$QN = 2\int_0^{k\sigma_x} e^2(x)p(x)\mathrm{d}x, \quad e^2(x) = \Delta^2/12, \quad \Delta = \frac{k\sigma_x}{2^n - 1} \tag{5.22}$$

Saturation noise

$$SN = 2\int_{k\sigma_x}^{\infty} (x - k\sigma_x)^2 p(x)\mathrm{d}x \tag{5.23}$$

The straightforward calculation shows

$$QN = 2\sigma_x^2\left[\frac{k^2}{12(N-1)^2}(0.5 - F(k))\right] \tag{5.24}$$

$$SN = 2\sigma_x^2\left[(k^2+1)F(k) - \frac{k}{\sqrt{2\pi}}\exp(-k^2/2)\right] \tag{5.25}$$

where $N = 2^n$, $F(k) = \frac{1}{\sqrt{2\pi}}\int_k^{\infty}\exp(-u^2/2)\mathrm{d}u$.

The total quantization noise is

$$PN = QN + SN \tag{5.26}$$

Fig. 5.12 illustrates the relationship between *PN* and *k*. The 0 dB of the reference value of noise is σ_x^2. As we can see in Fig. 5.12, for a 12-bit quantization, when $k = 4$, the total noise level is lower than 60 dB. The amplifier times play an important role in quantization.

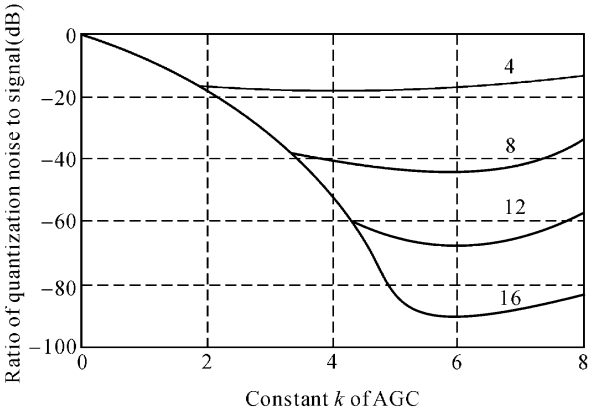


Fig. 5.12. Relationship between AGC constant k and quantization noise

5.3.4 FIR filtering

The advantage of the FIR filter is as follows:

(1) Linear phase: In digital sonar applications, the FIR filter does not bring any additional phase shift among multiple channels. It is necessary to have time delay compensation, so that any kinds of additional phase shift must be avoided.

(2) Reliability: The FIR filter is a reliable linear system; i.e., any boundary input results in boundary output, therefore no self excitation risk exists.

(3) Easy to implement: It is easy to realize FIR filtering in both hardware and software. Now on the market there are many specific FIR chips, which satisfy a wide variety of application requirements. And one can also easily compile the program of FIR in DSP ships.

With a shortage of FIR filtering it is hard to realize some special frequency responses, for example the notch filter. In this case, many more coefficients are needed than provided by the FIR filter.

In the following, we will derive the basic specification of the FIR filter, i.e., the symmetry of the coefficient.

Suppose the coefficient of an FIR filter is $h(0)$, ..., $h(N-1)$. The corresponding z transform is

$$H(z) = \sum_{n=0}^{N-1} h(n) z^{-n} \tag{5.27}$$

If the input signal is $x(nT_s)$ and the output signal is $y(nT_s)$, then the relationship between input and output is (see Fig. 5.13)

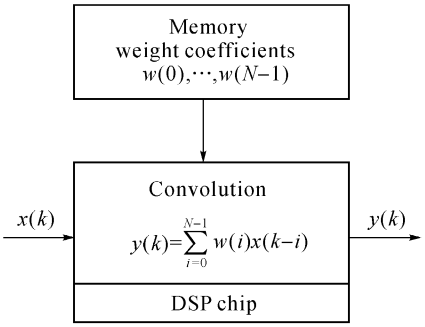


Fig. 5.13. Digital filtering of signal

$$y(kT_s) = \sum_{n=0}^{N-1} h(n)x(kT_s - nT_s) \tag{5.28}$$

Replacing z by $\exp(2\pi \mathrm{j} f T_s)$ in Eq. (5.27), it follows

$$H(\mathrm{e}^{2\pi \mathrm{j} f T_s}) = \sum_{n=0}^{N-1} h(n)\mathrm{e}^{-2\pi \mathrm{j} f n T_s} \tag{5.29}$$

Now, express $H(\mathrm{e}^{2\pi \mathrm{j} f T_s})$ by the amplitude and phase, we have

$$H(\mathrm{e}^{2\pi \mathrm{j} f T_s}) = A(f)\exp(\mathrm{j}\theta(f))$$

where $$A(f) = |H(\mathrm{e}^{2\pi \mathrm{j} f T_s})|$$

and

$$\theta(f) = \arctan\left[\frac{\mathrm{Im}(H(\mathrm{e}^{2\pi \mathrm{j} f T_s}))}{\mathrm{Re}(H(\mathrm{e}^{2\pi \mathrm{j} f T_s}))}\right]$$

The requirement of the linear phase is $\theta(f) = -2\pi f \tau \ (-1 < f < 1)$. Therefore

$$\arctan\left[-\frac{\sum_{n=0}^{N-1} h(n)\sin(2\pi f n T_s)}{\sum_{n=0}^{N-1} h(n)\cos(2\pi f n T_s)}\right] = -2\pi f \tau$$

It follows

$$\tan(2\pi f\tau) = -\frac{\sum_{n=0}^{N-1} h(n)\sin(2\pi fnT_s)}{\sum_{n=0}^{N-1} h(n)\cos(2\pi fnT_s)}$$

It is not difficult to prove that

$$\sum_{n=0}^{N-1} h(n)\sin(2\pi f\tau - 2\pi fnT_s) = 0$$

It obtains

$$\tau = (N-1)T_s / 2$$

and

$$h(n) = h(N-1-n), \quad 0 < n < N-1 \tag{5.30}$$

Eq. (5.30) shows that $h(0)$, ..., $h(N-1)$ is symmetrical and N can be an odd or even number.

Example

To design an FIR filter such that

$$H(\mathrm{e}^{2\pi \mathrm{j}fT_s}) = \begin{cases} 1 & |f| \leq f_p \\ 0 & f_p \leq |f| < f_s/2 \end{cases}$$

where f_p is the up frequency of the band and f_s is the sampling frequency. We have

$$H(\mathrm{e}^{2\pi \mathrm{j}fT_s}) = \sum_{n=-\infty}^{\infty} h(n)\mathrm{e}^{-2\pi \mathrm{j}fnT_s}$$

$$h(n) = \frac{1}{f_s}\int_{-f_s/2}^{f_s/2} H(\mathrm{e}^{2\pi \mathrm{j}fT_s})\exp(2\pi \mathrm{j}fnT_s)\mathrm{d}f$$

$$= \frac{1}{n\pi}\sin(2\pi f_p nT_s) = \frac{2f_p}{f_s}\frac{\sin(2\pi nf_p / f_s)}{2\pi nf_p / f_s}$$

There are many specific algorithms and software for FIR filter design. The Parks-McClellan algorithm is one of the famous and effective algorithms. Given passband, fluctuation and sampling frequency, it is easy to find a suitable filter. Fig. 5.14 is an example of an FIR filter with 127 orders, and is designed by using

MatLab software. We can see from this example that the frequency response is very good.

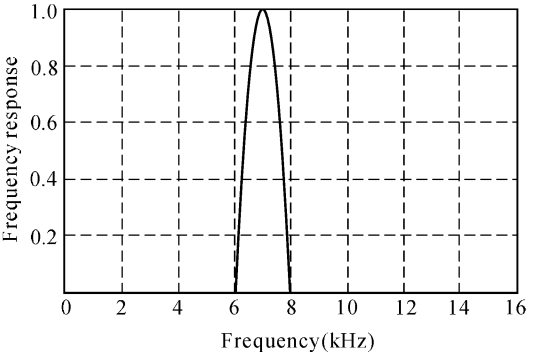


Fig. 5.14. FIR filter design (center frequency 7 kHz, band :6 – 8 kHz, 127 orders)

Now we are going to discuss the windowing problem in an FIR filter.

It is well-known that the design of a digital filter is based on the frequency response of an analogue filter. One hopes that the frequency response of a designed digital filter is close to that of an assigned analogue filter. In this sense, we can say that the design of a digital filter is actually "digital simulation of an analogue filter". For an FIR filter, since the coefficient is limited, it is a problem therefore of a cutting operation or a windowing operation for an indefinite series.

If the filter we are going to design has the frequency response $H(\mathrm{e}^{2\pi \mathrm{j}fT_s})$, in terms of the Fourier transform, the corresponding coefficients are indicated by $h_d(n)$, i.e.

$$H(\mathrm{e}^{2\pi \mathrm{j}fT_s}) = \sum_{n=-\infty}^{\infty} h_d(n)\mathrm{e}^{-2\pi \mathrm{j}fnT_s} \tag{5.31}$$

We can think that $h_d(n)$ is an ideal impulse response. The real FIR filter we want to implement has coefficient $h(n)$. Therefore, $h(n)$ is the result of windowed $h_d(n)$. The rectangle window function is

$$a(n) = \begin{cases} 1 & |n| \le (N-1)/2 \\ 0 & \text{otherwise} \end{cases} \tag{5.32}$$

Therefore

$$h(n) = h_d(n)a(n) \tag{5.33}$$

The frequency response $H_a(\mathrm{e}^{2\pi \mathrm{j}fT_s})$ of $h(n)$ is actually the convolution of the

Fourier transform $H(e^{2\pi jfT_s})$ of $h_d(n)$ and the Fourier transform $A(e^{2\pi jfT_s})$ of $a(n)$

$$H_a(e^{2\pi jfT_s}) = \frac{1}{f_s}\int_0^{f_s} H(e^{2\pi jfT_s})A(e^{2\pi j(f-u)T_s})du \tag{5.34}$$

From this equation we can calculate the real frequency response function of the digital filter.

In addition to the above rectangle window, there are also other kinds of windows. Here are some frequently used windows.

(1) Triangle window:

$$a(n) = \begin{cases} 2n/N & n = 0, \ ..., \ N/2 \\ a(N-n) & n = N/2+1, \ ..., \ N-1 \end{cases} \tag{5.35}$$

The first side lobe is about -26 dB.

(2) Hanning window:

$$a(n) = 0.5(1-\cos(2\pi n/N)), \quad n = 0, \ ..., \ N-1 \tag{5.36}$$

The first side lobe is about -32 dB.

(3) Hamming window:

$$a(n) = 0.54 - 046\cos(2\pi n/N), \quad n = 0, \ ..., \ N-1 \tag{5.37}$$

5.4 High Precision RAM Dynamic Beamforming and Optimum Spatial Filtering

Beamforming is the kernel of all sonar. The digital beamforming technique is the ingenious combination of modern digital signal processing techniques and computer technology. The post-processing, including automatic target tracking, passive ranging, auditory listening, multi-target resolution, target identification and torpedo alarming, etc., is determined by the performance of digital beamforming.

In this section, the beamforming technique in time domain is introduced and the concept of optimum spatial filtering theory is briefly discussed[75-76].

5.4.1 Digital Beamforming Using SRAM

The digital beamforming system is also called a pre-beamforming system, since

the beamforming system has to form the steering beam over a 360°range beforehand. And the digital beams are actually discrete samples of the directivity function. If there are no targets present, then the response of a beamforming system should be a horizontal line.

Fig. 5.15 shows a block diagram of a beamforming system in time domain, supposing the target signal comes from direction θ. The signal received by N elements can be expressed as

$$x(t+\tau_1(\theta)),\ ...,\ x(t+\tau_N(\theta)) \tag{5.38}$$

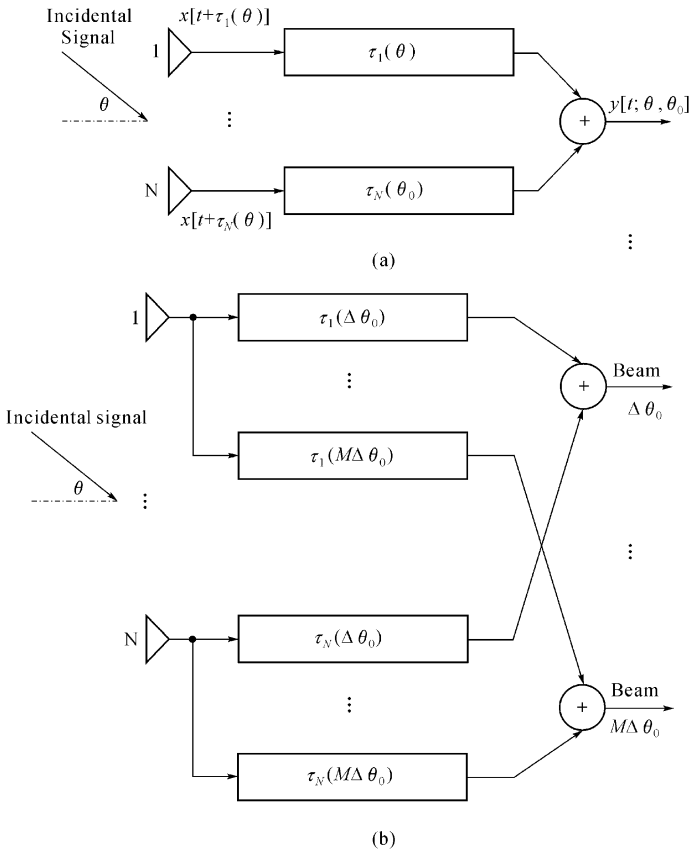


Fig. 5.15. Block diagram of digital multi-beam system. (a) Single beam system; (b) Multi-beam system

where $\tau_i(\theta)$ denotes the time delay of i-th hydrophone with some reference element. The selection of the reference element is arbitrary, and is convenient to the calculation and physical realizability. Of course, it is worth showing that in beamforming, plus or minus any given value Δ of time delay from all channel

time compensation will not affect the results of beamforming; i.e., the beamforming of $\{x(t+\tau_i(\theta))\}$ and $\{x(t+\tau_i(\theta)+\Delta)\}$ achieves the same results.

In order to form the beam steered in the direction θ_0, it is necessary to delay $\tau_i(\theta_0)$ for the signal $x(t+\tau_i(\theta))$, and we get the signal

$$x_i(t) = x(t+\tau_i(\theta)-\tau_i(\theta_0)) \tag{5.39}$$

Suming up $x_i(t)$ and taking square integration, we will get the directivity function of the sonar system, i.e., the system response that the sonar steers in direction θ_0, while the target appears in direction θ.

For a multi-beam system, the sonar system has to form M equal spaced beams, the angle of the adjacent beam is $\theta_0 = 360°/M$. In this time it is necessary to delay the signal of each channel by $\tau_i(\theta_0)$, ..., $\tau_i(M\theta_0)$, and then sum up the corresponding components in each channel to get M beams. Since the incidental angle θ of signal can be arbitrary, therefore it is hard to say that zs will be valid for some value of k. This is the bias problem of target bearing and we will discuss it later.

In order to realize a time delay for a digital signal, the traditional method is to use a digital time delay line. The digitized signal is put into a shift register; the driven period is just equal to the sampling duration of the input signal. And the maximum length of the time delay line is equal to the necessary maximum delay time. Each signal is connected to one time delay line. The output of some taps, which is calculated by the beamforming direction, is taken to the beamforming system. The time delay line is driven by the sampling impulse such that we can obtain a series of time delay signals.

This kind of traditional method of beamforming has already been replaced by the powerful SRAM (static random access memory). It allows data input sequentially, and one can take the data from any preset address. And the address table can be compiled based on the time delay value of necessary beams. This procedure is illustrated in Fig. 5.16.

For example, if the sampling duration is T_s, at the 0 moment the signal $x(iT_s)$ is put in the address 0 of SRAM. After kT_s periods, the signal $x((i+K)T_s)$ is put in address k. If we take out the signal in address 0 at the moment of kT_s, the signal $x(iT_s)$ should have a time difference with $x((i+K)T_s)$ at address kT_s. Therefore the digital time delay is realized.

The calculation of time delay compensation is based on the principle of "nearest integer". The number of taps is equal to the nearest integer of the value of time delay over sampling duration. (see Fig. 5.17.)

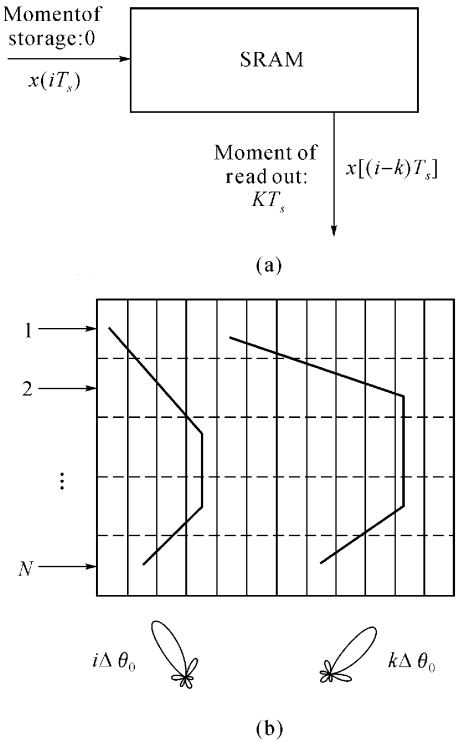


Fig. 5.16. Principle of dynamic beamforming. (a) Time delay realization by using SRAM; (b) Dynamic multi-beamforming by using SRAM

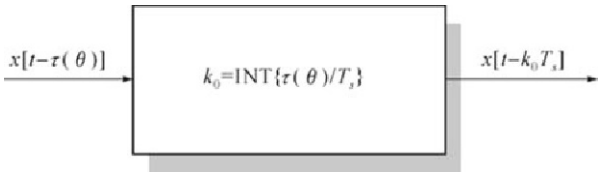


Fig. 5.17. Error of time delay compensation

If the necessary time delay is $\tau(\theta)$, then the number of taps is

$$L = \text{INT}[\tau(\theta) / T_s] \tag{5.40}$$

where function INT[ ] indicates the arithmetic round operation, i.e., INT[4.2] = 4 and INT[5.7] = 6. This algorithm will guarantee the error of time delay compensation is in $[-T_s / 2,\ T_s / 2]$. For N channels, there are N different time compensation errors

$$\Delta_i = \mathrm{INT}[\tau_i(\theta)/T_s] - \tau_i(\theta)/T_s \tag{5.41}$$

When the number of channel N is large enough, we can consider that Δ_i is uniformly distributed in $[-T_s/2,\ T_s/2]$. Because, once the Δ_i is chosen, then the accuracy of the beamforming is finalized, and therefore one should make the variance of Δ_i as small as possible. The following method can be considered as a reasonable choice in reducing the variance of Δ_i. Just as we have shown before that the beginning point of the time delay calculation is arbitrary, therefore it is always possible to find a time beginning point τ_0, such that the variance of

$$\Delta_i' = \mathrm{INT}\left[\frac{\tau_0 + \tau_i(\theta)}{T_s}\right] + \frac{\tau_0 + \tau_i(\theta)}{T_s}$$

is the minimum. So we can obtain a combination of time delay values, which has a minimum compensation time delay error.

5.4.2 Optimum Spatial Filter

The main aim of beamforming is to enhance the array sensitivity in the direction of an incidental signal and suppress the interference from the other direction. But the conventional beamforming (CBF) technique only considers the "steering direction", and neglects the other factor which will affect the signal detection. This is the serious shortage of CBF. In the application area, we should consider a more effective beamforming method. In an ideal situation, if a beam pattern which passes through the signal in one direction rejects all signals from other direction, then it is the desired "optimum" spatial filter. This kind of spatial filter is called "absolutely optimum spatial filter" by Hanna[75]. It has been proved that the optimum spatial filter can be realized using a similar technique of CBF[76].

Consider a multiple channel filtering system as in Fig. 5.18.

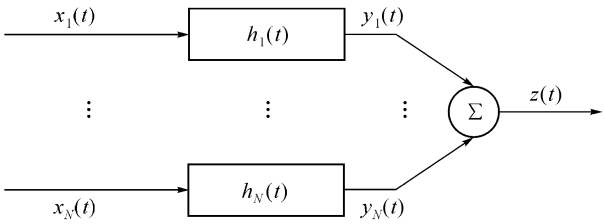


Fig. 5.18. Optimum multi-channel filter

The i-th signal received is

$$x_i(t) = p_i(t) + q_i(t) \tag{5.42}$$

where $p_i(t)$ and $q_i(t)$ are two components of $x_i(t)$, the definitions depend on the case discussed, and this will be explained later. The impulse response of each channel is $h_i(t)$ and its Fourier transform is $H_i(f)$. The output of i-th filter is

$$y_i(t) = x_i(t) * h_i(t) \tag{5.43}$$

The output of the spatial filter or array filter is

$$z(t) = \sum_{i=1}^{N} x_i(t) * h_i(t) \tag{5.44}$$

The optimum spatial filter can be simply considered as a "signals all pass and noise all reject" filter.

First of all, let us consider the simplest situation, i.e., the filter already exactly steers in the incidental target

$$x_i(t) = p(t) + q(t - \Delta_i) \tag{5.45}$$

where $p(t)$ is the signal and $q(t-\Delta_i)$ is the noise. Therefore the Fourier transform of $z(t)$ is

$$Z(f) = F[z(t)] = P(f)\sum_{i=1}^{N} H_i(f) + Q(f)\sum_{i=1}^{N} \exp(2\pi \mathrm{j} f \Delta) H_i(f) \tag{5.46}$$

Denote

$$R(f) = \sum_{i=1}^{N} \exp(2\pi \mathrm{j} f \Delta) H_i(f) \tag{5.47}$$

It is called the rejection response of the array filter. Therefore

$$Z(f) = F[z(t)] = P(f)\sum_{i=1}^{N} H_i(f) + R(f)Q(f) \tag{5.48}$$

In the condition of

$$\sum_{i=1}^{N} H_i(f) = 1 \tag{5.49}$$

one can find the minimum value of $|R(f)|^2$. This is obviously a conditional functional extreme value problem. It is proved that[76]

$$H_{i,\text{opt}}(f) = \frac{N - \exp(-2\pi \mathrm{j} f \Delta_i) \sum_{n=1}^{N} \exp(2\pi \mathrm{j} f \Delta_n)}{N^2 - |\sum_{n=1}^{N} \exp(2\pi \mathrm{j} f \Delta_n)|^2} \quad (i = 1, \ ..., \ N) \tag{5.50}$$

It is easy to verify that

$$\sum_{i=1}^{N} H_i(f) = 1 \ \text{ and } R(f) = 0 \tag{5.51}$$

unless $\Delta_n f = k_n$, and k_n is an integer. Eq. (5.50) presents the transfer function of the optimum filter in the case of a signal exactly known. If the incidental angle of the signal is not completely known, it is necessary to scan the beam to find the incidental angle.

Suppose the *i*-th signal is

$$x_i(t) = p(t - \delta_i) + q(t - \Delta_i) \quad (i = 1, \ ..., \ N)$$

Therefore

$$Z(f) = P(f) \sum_{i=1}^{N} \exp(2\pi \mathrm{j} f \delta_i) H_i(f) + Q(f) \sum_{i=1}^{N} \exp(2\pi \mathrm{j} f \Delta_i) H_i(f)$$

Denote

$$\exp(2\pi \mathrm{j} f \delta_i) H_i(f) = G_i(f)$$

we have

$$Z(f) = P(f) \sum_{i=1}^{N} G_i(f) + Q(f) \sum_{i=1}^{N} \exp(2\pi \mathrm{j} f (\Delta_i - \delta_i)) G_i(f) \tag{5.52}$$

Eq. (5.52) is similar to Eq. (5.48), but $H_i(f)$ is replaced by $G_i(f)$. Therefore we have

$$G_{i,\text{opt}}(f) = \frac{N - \exp(2\pi \mathrm{j} f (\delta_i - \Delta_i)) \sum_{n=1}^{N} \exp(2\pi \mathrm{j} f (\Delta_n - \delta_n))}{N^2 - |\sum_{n=1}^{N} \exp(2\pi \mathrm{j} f (\Delta_n - \delta_n))|^2} \quad (i = 1, \ ..., \ N) \tag{5.53}$$

or

$$H_{i,\text{opt}}(f)=\frac{N-\exp(2\pi\mathrm{j}f(\delta_i-\Delta_i))\sum_{n=1}^{N}\exp(2\pi\mathrm{j}f(\Delta_n-\delta_n))}{N^2-|\sum_{n=1}^{N}\exp(2\pi\mathrm{j}f(\Delta_n-\delta_n))|^2}\times\exp(-2\pi\mathrm{j}f\delta_i) \quad (5.54)$$

where $i=1, \ldots, N$.

Eq. (5.54) is the general expression of an optimum spatial filter. It can steer in any direction and pass through the signal and reject any noise from any other directions.

Now we will introduce two application examples.

Multiple Noise Canceling System

This system is used to suppress plane wave interference. A CBF system steers to interference, and then subtracts the interference from each channel. The structure of this system is similar to the DICANNE system, proposed by Anderson[14]. Figs. 5. 19 and 5.20 are illustrations of this spatial array filter.

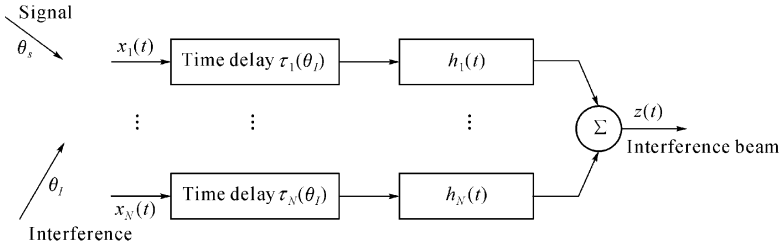


Fig. 5.19. Bearing at interference DOA

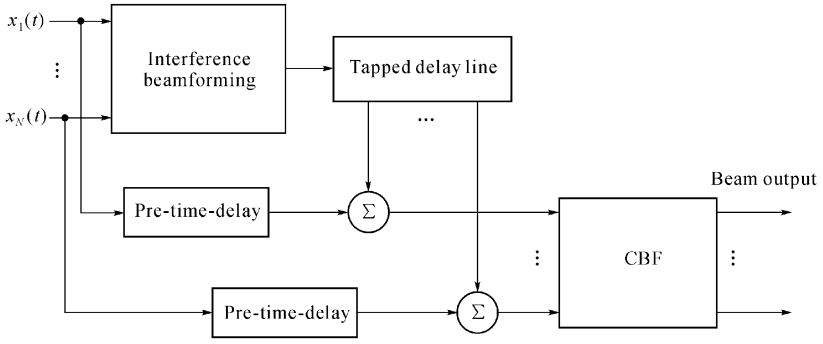


Fig. 5.20. Multi-beam interference canceling system

Suppose the signal from *i*-th channel is

$$x_i(t) = I(t) + s_i(t) = I(t) + s(t - \Delta_i)$$

where $I(t)$ is interference and $\Delta_i = \tau_i(\theta_s) - \tau_i(\hat{\theta}_I)$, where $\hat{\theta}_I$ is the estimate for the interference incidental angle θ_I, and $\hat{\theta}_I = \theta_I$.

In the DICANNE system, one uses the interference beam

$$u(t) = NI(t) + \sum_{i=1}^{N} s(t - \Delta_i) \tag{5.55}$$

There are also some residual signals or noise in the interference beam, i.e., the interference beam is not "pure". But if we use an optimum spatial filter, the situation will be totally different. Since

$$z(t) = \sum_{i=1}^{N} x_i(t) * h_i(t) = I(t) * \sum_{i=1}^{N} h_i(t) + \sum_{i=1}^{N} h_i(t) * s(t - \Delta_i)$$

Based on the Eq. (5.52), we obtain

$$H_{i,\text{opt}}(f) = \frac{N - \exp(-2\pi \mathrm{j} f \Delta_i) \sum_{n=1}^{N} \exp(2\pi \mathrm{j} f \Delta_n)}{N^2 - |\sum_{n=1}^{N} \exp(2\pi \mathrm{j} f \Delta_n)|^2} \quad (i = 1, \ ..., \ N) \tag{5.54}$$

In using this transfer function, the interference to the signal ratio in the interference beam will be infinite.

Post Interference Suppressing Technique for CBF

In the previous example, in order to suppress the interference in each channel, we first steer the beam in the direction of interference and form an interference beam. But we can also steer first in the direction of the signal and form the signal beam and then suppress the interference. The block diagram is shown in Figs. 5.21 and 5.22.

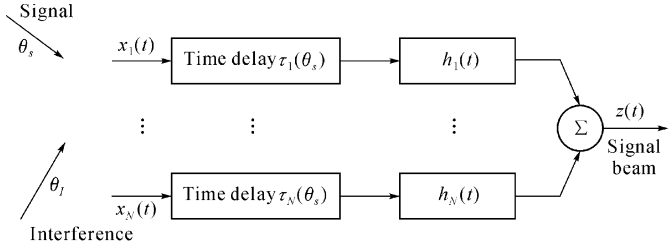


Fig. 5.21. Interference suppressing system

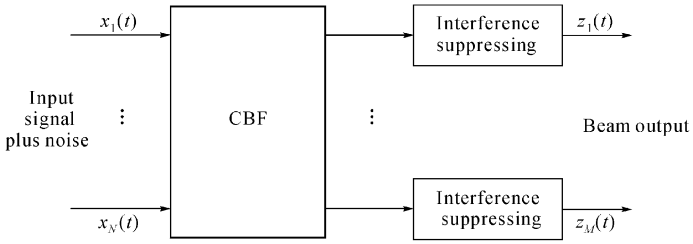


Fig. 5.22. CBF system with interference suppressing

We have $p_i(t) = s(t - \tau_i(\theta_s) + \tau_i(\hat{\theta}_s))$, $q_i(t) = I(t - \tau_i(\theta_I) + \tau_i(\hat{\theta}_s))$, and $x_i(t) = p_i(t) + q_i(t)$. The optimum transfer function is

$$H_{i,\text{opt}}(f) = \frac{N - \exp[2\pi \mathrm{j} f(\tau_i(\theta_s) - \tau_i(\theta_I))] \sum_{n=1}^{N} \exp[2\pi \mathrm{j} f(\tau_n(\theta_I) - \tau_n(\theta_s))]}{N^2 - \left| \sum_{n=1}^{N} \exp[2\pi \mathrm{j} f(\tau_n(\theta_I) - \tau_n(\theta_s))] \right|^2} \times \exp[-2\pi \mathrm{j} f(\tau_i(\hat{\theta}_s) - \tau_i(\theta_s))] \quad (i = 1, ..., N) \tag{5.55}$$

The expressions of the optimum filter introduced above look quite complicated, but it is not difficult to implement these in digital sonar with hardware / software. Some of system simulation results are illustrated in Figs. 5.23 – 5.25. Fig. 5.23 is a block diagram of a strong interference suppressing system. It is similar to the DICANNE system with a physically realizable pre-time-delay line. The adjustment of the tapped delay line can be implemented by adaptive technique.

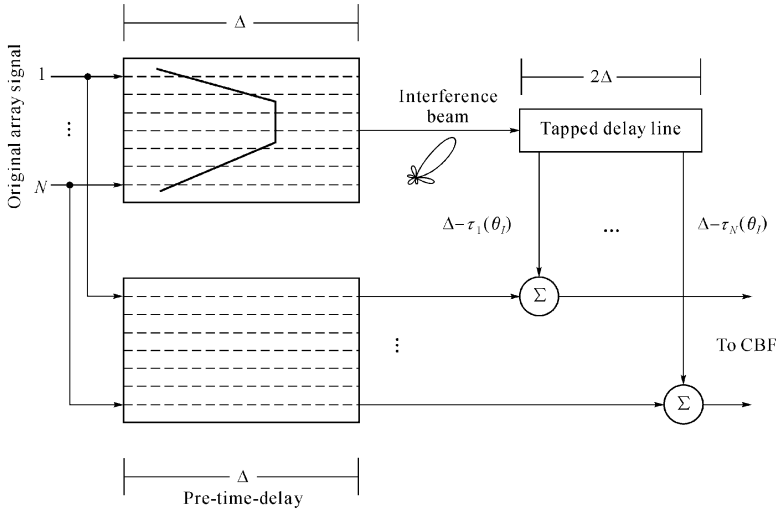


Fig. 5.23. Strong interference suppressing system

Fig. 5.24 shows the directivity function before and after interference suppressing, where $N = 48$, $d / \lambda = 0.5$. Fig. 5.24(a) is the situation of the signal only. Fig. 5.24(b) is the situation when a signal is at 0° and an interference target appears at 11°. Suppose $(SNR)_{in} = -10$ dB, and $(INR)_{in} = 7$ dB, in other words, the interference to the signal ratio is –17 dB. We almost cannot identify that there is a target at 0°. Fig. 5.25(a) is the situation after interference suppressing, and we can see the signal at 0°. Fig. 5.25(b) is the local zoom effect of the same figure.

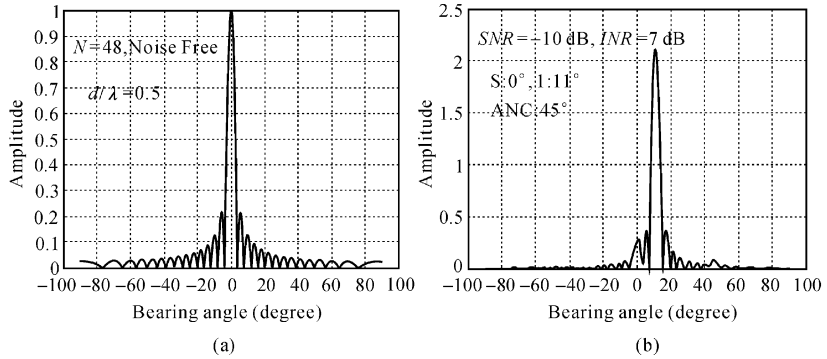


Fig. 5.24. Beam pattern before ANC. (a) Beam pattern of signal only ($N = 48$); (b) Signal plus interference

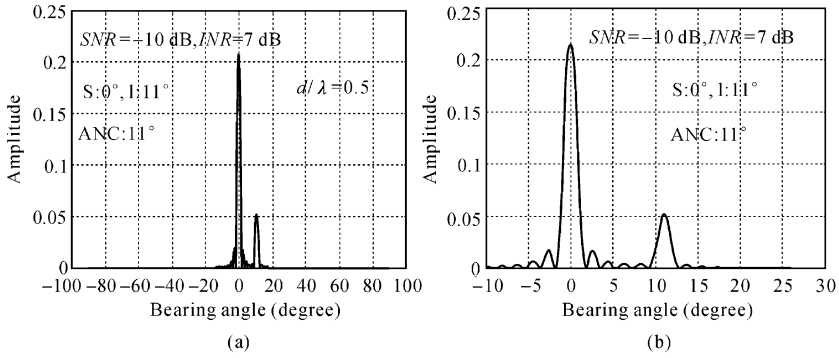


Fig. 5.25. Beam pattern after ANC. (a) Beam pattern of ANC ($N = 48$); (b) Zoom beam pattern ($N = 48$)

5.5 Target Characteristic Analysis: LOFAR, DEMON and Adaptive Line Enhancer

In modern digital sonar system, the signal detection is completed by a multiple

beam technique and the target characteristic analysis and classification / identification is completed by the single beam. Of course it might be one beam from the multiple beam system which steers the target. LOFAR and DEMON are the most effective means of analysis in digital signal processing.

In this section, the basic theory and application of the LOFAR and DEMON techniques will be introduced. The adaptive line enhancer is also discussed. From the weak signal detection point of view, the sonar designer prefers to use the signal after line enhancing instead of using the direct multi-beam signal.

5.5.1 LOFAR and DEMON Analysis

Suppose an array with N elements, the input signal series is $x_i(k)$ $(i=1, ..., N)$. The signal steered in direction $m\theta_0$ is indicated by $x_i(k, m)$ $(m=1, ..., M)$, where M is the number of beams and θ_0 is the angle of adjacent beams, $\theta_0 = 360^\circ / M$. The m-th beam signal is

$$y_m(k) = \sum_{i=1}^{N} x_i(k, m)$$

The beam output value is $D_m = E[y_m^2(k)]$. Based on the output results of multiple beams, it is easy to determine the incidental angle of the target. In order to manage the characteristics of the target signal, it is necessary to analyze the signal feature for all M beams. The analysis objective is mainly the m-th beam or the beam near it. Generally speaking, D_m changes slowly. It contains the bearing angle information, and the data $y_m(k)$ includes the necessary information of the target, noise and interference. Therefore, the signal analysis objective of LOFAR and DEMON should be $y_m(k)$. Since LOFAR and DEMON are all the frequency domain method, the FFT technique will therefore play a very important role.

5.5.1.1 Detection of Single Frequency Signal in Noise Background

Since the LOFAR and DEMON analysis will finally conclude with the detection of a weak single frequency signal, it is therefore worth studying the detection problem of a single frequency signal in the noise background. The single frequency signal is also called a sinusoidal signal. The main parameter of this signal is its amplitude and frequency. In expressing this signal, it usually has a random phase, uniformly distributed in the range of $(0, 2\pi)$.

Our signal model is

$$x(t) = A\cos(2\pi f_0 t + \varphi) + n(t) \tag{5.58}$$

where $n(t)$ is the noise, ϕ is the random phase and A is the amplitude. The SNR is $(\mathrm{SNR})_{\mathrm{in}} = A^2 / 2\sigma_n^2$.

Fig. 5.26 illustrates the results of system simulation. The signal frequency is $f_0 = 300\,\mathrm{Hz}$ and the sampling frequency is $f_s = 5,000\,\mathrm{Hz}$. From Fig. 5.26(a) to Fig. 5.26(c), we can see the sinusoidal signal is detectable, even in $(\mathrm{SNR})_{\mathrm{in}}$ = –18 dB, but below this value, the signal will not be detectable at 1,024 point FFT, Fig. 5.26(d). When we increase the data number of FFT, the sinusoidal signal becomes detectable again. Fig. 5.26(e), N = 16×1,024. Fig. 5.26(f), N = 64×1,024. We can conclude that, if the background noise is seriously stationary, then the increase in integration time will bring more system gain. But it is not always possible with large number samples.

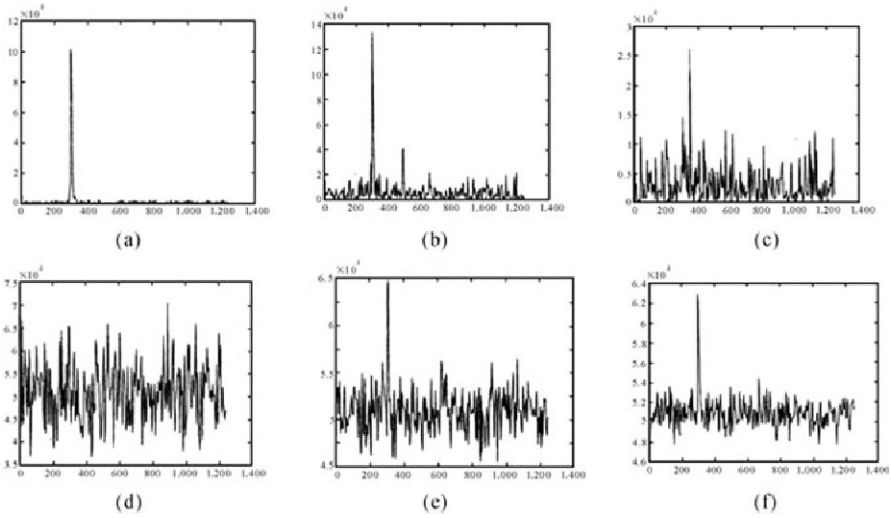


Fig. 5.26. Sinusoidal signal detection in background of white noise (f_0 = 300 Hz, f_s = 5,000 Hz). (a) SNR = 0 dB, N = 1,024; (b) SNR = –10 dB, N = 1,024; (c) SNR = –18 dB, N = 1,024; (d) SNR = –30 dB, N = 1,024; (e) SNR = –30 dB, N = 16,384; (f) SNR = –30 dB, N = 65,536

5.5.1.2 General Concept of LOFAR and DEMON

In much radiated noise, there are some kinds of single frequency components, called “tone components” or “pure tone”. It is actually the low frequency component in radiated noise, resulting from the propeller rotation of the vessels. In frequency domain, these single frequency components usually have a spectrum level higher than average, for example 10 – 25 dB. The signal analysis in frequency domain will be able to extract the single frequency characteristic. And it is helpful for weak signal detection and target identification.

Fig. 5.27 is the block diagram of LOFAR and DEMON analysis. The input

signal of LOFAR and DEMON comes from the beamforming system, therefore it usually has a SNR higher than a single hydrophone, at least 10 lg N dB, where N is the channel number.

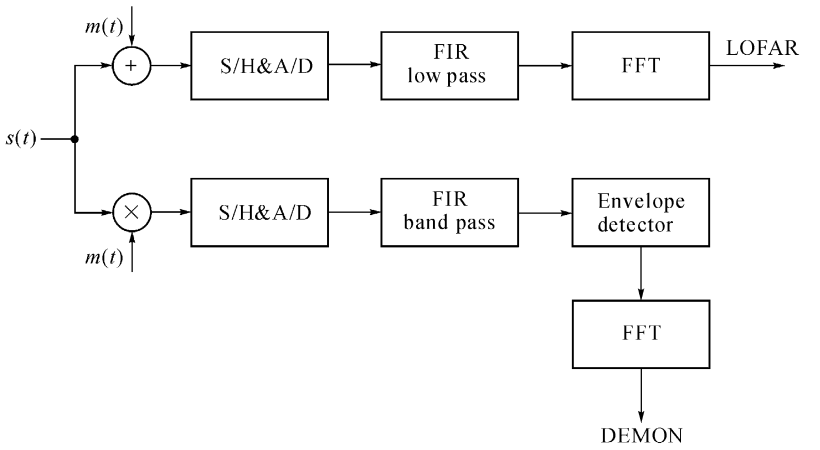


Fig. 5.27. Illustration of LOFAR and DEMON

The single frequency components in radiated noise have two different forms. One is an additive component and the other is a productive component. If we denote the component which has no single frequency signal as $m(t)$, then the additive component has the form of $x(t) = s(t) + m(t)$, and the productive component is $x(t) = m(t)s(t)$, where $s(t)$ is the single frequency component (see Fig. 5.28.)

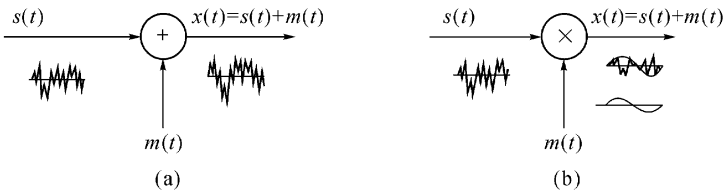


Fig. 5.28. Additive and modulated component in sonar signal. (a) Additive component; (b) Multiply component

There are different methods for extracting the signal $s(t)$. For the additive component, we can use the LOFAR technique, and for the productive component use the DEMON technique.

As a matter of fact, in the first situation

$$X(f) = S(f) + M(f) \tag{5.59}$$

And in the second situation

$$X(f) = S(f) * M(f) \tag{5.60}$$

where $X(t)$, $S(f)$, and $M(f)$ are the spectrums of $x(t)$, $s(t)$ and $m(t)$, respectively. In LOFAR analysis, the direct Fourier analysis can extract the single frequency component from the input signal. However, for DEMON analysis, it is necessary to demodulate the envelope of the input signal and then perform FFT analysis.

The extracted signal from LOFAR analysis reflects the line spectrum component in the radiated noise (usually below 100 Hz), and is related to the rotation speed of the propeller and the number of blades (see Eq. (4.45)).

Fig. 5.29 illustrates the spectrum of the amplitude modulation signal. In Fig. 5.29(a), the original signal is the broadband signal and in Fig. 5.29(b) the original signal is the sinusoidal signal. We can understand, in DEMON analysis, why we should take the envelope of radiated noise and then perform FFT analysis to extract the low frequency sinusoidal signal.

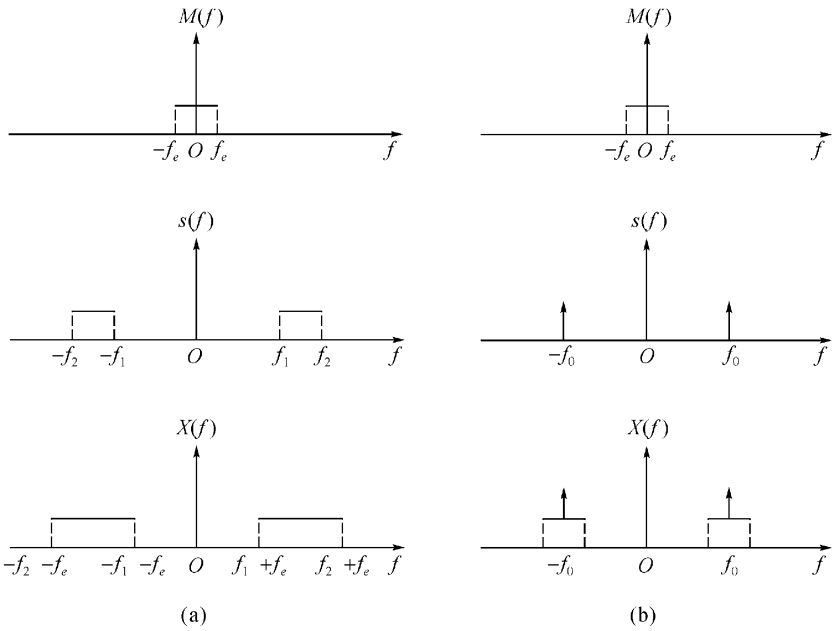


Fig. 5.29. Spectrum of amplitude modulation signal. (a) Original broad band signal; (b) Original sine wave

5.5.1.3 Examples of LOFAR and DEMON Analysis

Now two examples are presented to illustrate LOFAR and DEMON analysis. Suppose the sampling frequency is f_s = 50 kHz, and the sampling duration is T_s = 20 μs (see Fig. 5.30). Choose the beam which exactly steers to the target, or the beam which steers closest to the target. Assume the signal of this beam is $y(k)$.

Firstly, we have to filter this signal by FIR filter, the length of this filter is 64 points. The frequency range is chosen as 100 – 400 Hz. The sample number of FFT analysis is 1024 points. The time duration of each data segment should be: 1,024×20 μs = 24.5 ms.

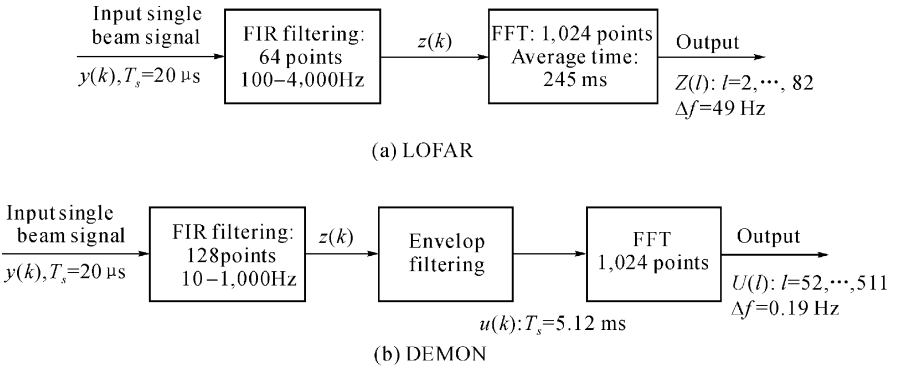


Fig. 5.30. Examples of LOFAR and DEMON realization

The DFT of $y(k)$ is

$$Y_m(l) = \sum_{i=0}^{1,023} y(m \times 1,024) + i) \exp(-2\pi \mathrm{j} il / 1,024) \quad (l = 0, \ ..., \ 1,023) \quad (5.61)$$

The average value of 10 times DFT is

$$Y(l) = \frac{1}{10} \sum_{m=0}^{9} Y_m(l) \quad (l = 0, \ ..., \ 1,023)$$

The value of $[|Y(l)|^2]^{1/2}$ is sent to display. Since the signal frequency band is limited to 100 – 4,000 Hz, therefore the spectrum lines can be selected as $l = 2, \ 3, \ ..., \ 82$. It will cover the frequency of what we are interested in. The frequency resolution of LOFAR analysis is

$$\Delta f = f_s / N = 50,000 / 1,024 \approx 49 \quad \text{(Hz)}$$

Fig. 5.31 presents the result of system simulation, where No.1 beam is supposed to steer the target. Because it is the result of a 10 times average, therefore on the display screen, each row needs 245 ms, and each beam display needs 245×64 ms ≈ 16 s. This is a typical LOFAR display scheme; it is a frequency / history display.

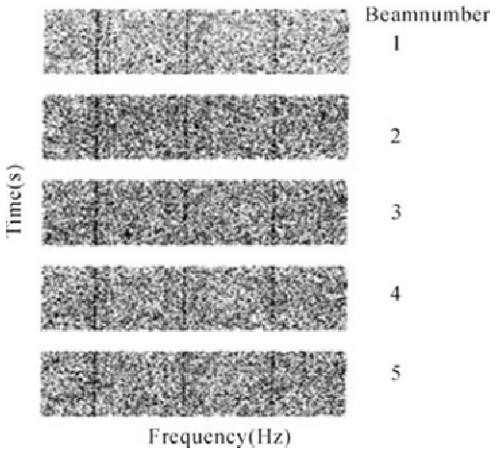


Fig. 5.31. Example of LOFAR figure

The block diagram of DEMON analysis is shown in Fig. 5.30(b). In Chapter 4, we have mentioned that the low frequency modulation signal reflects the relationship between the rotation speed of the propeller, the line spectrum frequency of radiated noise and the number of blades

$$f_m = m \times n \times s$$

where f_m is the m-th resonant wave, n is the number of blades and s is the rotation speed of the propeller in rad/s. If $n = 4$, $s = 1.5$ rad/s then $f_m = 6m$. The beam signal is filtered by FIR filter with passband 10 – 1,000 Hz. Since the lower cut-off frequency is close to 0 Hz, it is necessary to have more coefficients of the FIR filter and one chooses $N = 128$. In order to extract the envelope of the modulation signal, an absolute filtering operation is adopted

$$u(k) = \frac{1}{M} \sum_{i=1}^{M} | z(kM + i) | \tag{5.62}$$

where M is the average number, which is related to the frequency resolution of FFT in the next stage (Table 5.1).

Table 5.1 Frequency resolution of DEMON analysis

Average number M	Integration time (s)	Δf (Hz)	f_{max} (Hz)*
32	0.64	1.53	781.3
64	1.28	0.76	390.6
128	2.56	0.38	195.3
256	5.12	0.19	97.6

* The maximum signal frequency to be analyzed.

If one chooses $M = 256$, then the output rate of $u(k)$ is one sample per 5.12 ms; The frequency resolution of 1,024 points FFT is

$$\Delta f = \frac{1,000}{5.12 \times 1,024} = 0.19 \text{ (Hz)}$$

The results $|U(f)|$ of FFT for $u(k)$ are the data of the DEMON display. The time duration of each row on the screen is 5.24 s. Fig. 5.32 is the result of system simulation.

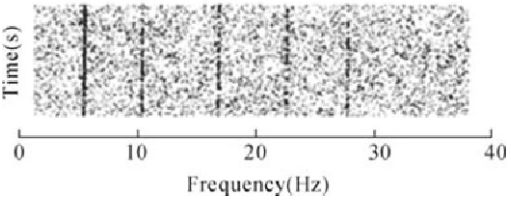


Fig. 5.32. Example of DEMON figure

5.5.2 *Adaptive Line Enhancing*

LOFAR and DEMON analysis is based on the assumption of the existence of a single frequency component in a detected signal. If there are no single frequency components, or they are very weak, then LOFAR and DEMON analysis will fail to extract the signal. That means the SNR will be essentially the key factor in using LOFAR and DEMON analysis. But how can we make the beam output of a sonar system have a higher SNR for a single frequency signal? A very effective technique to enhance the SNR for a single frequency component is to use an adaptive line enhancer (ALE). The details of ALE will be introduced as follows; the basic principles, performance in a steady state and applications in digital sonar will be presented.

Fig. 5.33 presents the principle block diagram of ALE. The primary signal consists of broadband noise $n(t)$ and sinusoidal signal $A\cos(2\pi f_0 t + \varphi)$.

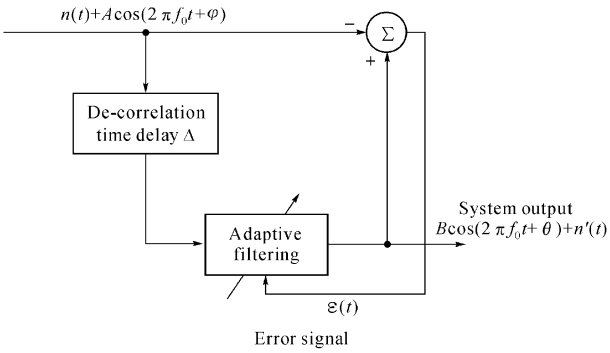


Fig. 5.33. Block diagram of adaptive line enhancer

Our main goal is to utilize the different correlation performance of the broadband noise and sinusoidal signal to enhance the signal by using an adaptive filtering technique. Since the sinusoidal signal is a periodical signal with period 2π, any kind of time delay corresponds to some phase difference. But it is still the sinusoidal signal with the same frequency; therefore the correlation characteristic is the same. But for the broadband noise, the situation is different, when the time delay is large enough the correlation between original noise and delayed noise tends to zero. Based on this fact, it is possible to enhance the sinusoidal signal and suppress noise.

First of all, it is important to show that for the sinusoidal signal, amplitude weighting can synthesize any phase. In fact, suppose we have a sinusoidal signal $\cos(2\pi f_0 t)$. After phase delay $\pi/2$, we get $\sin(2\pi f_0 t)$. Based on these two orthogonal signals we can get any phase sinusoidal signal with the same frequency

$$A\cos(2\pi f_0 t - \varphi) = w_1 \cos(2\pi f_0 t) + w_2 \sin(2\pi f_0 t) \tag{5.63}$$

where

$$A = (w_1^2 + w_2^2)^{1/2}, \quad \varphi = \arctan(w_2 / w_1) \tag{5.64}$$

Now, delay the original signal by Δ, and subtract it from the output of the adaptive filter (see Fig. 5.33)

$$\varepsilon(t) = n(t) + A\cos(2\pi f_0 t + \varphi) - [n'(t) + B\cos(2\pi f_0 t + \theta)] \tag{5.65}$$

The algorithm in ALE is to adjust the adaptive filter to make the error signal in Eq. (5.65) the minimum, since the noise is independent of the signal and the time delay is large enough such that $n(t)$ and $n(t-\Delta)$ are independent. Therefore $n(t)$ and $n'(t)$ are independent. It follows that

$$E[\varepsilon^2(t)] = E[n^2(t)] + E[n'^2(t)] + E[A\cos(2\pi f_0 t + \varphi) - B\cos(2\pi f_0 t + \theta)]^2 \tag{5.66}$$

In Eq. (5.66), $E[n^2(t)]$ is a constant, independent of the filter. Therefore the minimum value of $E[\varepsilon^2(t)]$ is equivalent to the minimum value of

$$E[n'^2(t)] + E[A\cos(2\pi f_0 t + \varphi) - B\cos(2\pi f_0 t + \theta)]^2 \tag{5.67}$$

We have shown that suitable weight coefficients A and B can make $\theta = \varphi$. That is to say, the second term of Eq. (5.67) can be zero, if necessary. At the same time, the first term might not be the minimum. The final goal of ALE is

to adjust the adaptive filter such that the variance of the noise component $n'(t)$ in output

$$y(t) = B\cos(2\pi f_0 t + \theta) + n'(t)$$

is as small as possible and the signal component $B\cos(2\pi f_0 t + \theta)$ as close as possible to $A\cos(2\pi f_0 t + \varphi)$. Therefore, the SNR in $y(t)$ reaches the maximum value. Based on this analysis, we find that ALE has very good potential ability to enhance the sinusoidal signal. The very attractive advantage is that it is not necessary to know the signal frequency.

Now we turn to analyze the ALE behavior in a steady state. The adaptive filter shown in Fig. 5.33 is implemented by the tapped delay line, just as we have illustrated in Fig. 3.14. Suppose the sampling period is T_s and the decorrelation time delay is Δ. For simplicity, assume that

$$\Delta = d_0 T_s \tag{5.68}$$

where d_0 is a positive integer.

Denote

$$x(t) = n(t) + A\cos(2\pi f_0 t + \varphi), \qquad y(t) = x(t - \Delta) \tag{5.69}$$

$$\boldsymbol{w} = (w(0), \ldots, w(L-1))^{\mathrm{T}} \tag{5.70}$$

where L is the length of the tapped delay line.

The optimum resolution is (see Eq. (3.119))

$$\boldsymbol{w}_{\text{opt}} = \boldsymbol{R}_{yy}^{-1}\boldsymbol{R}_{yx} \tag{5.71}$$

The (k, l) -th element of matrix $\boldsymbol{R}_{yy}$ is

$$R_y(k-l) = E[y(k)y(l)) = R_s(k-l) + R_n(k-l) \quad (k,\ l = 0,\ \ldots,\ L-1) \tag{5.72a}$$

The l-th element of vector $\boldsymbol{R}_{yx}$ is

$$E(y(k-l)x(k)) = R_s(l + d_0) \quad (l = 0,\ \ldots,\ L-1) \tag{5.72b}$$

Eq. (5.71) can be written as

$$\sum_{k=0}^{L-1}[R_s(k-l) + R_n(k-l)]w(k) = R_s(d_0 + l) \quad (l = 0,\ \ldots,\ L-1) \tag{5.73}$$

The solution of Eq. (5.73) is denoted by

$$\mathrm{w}_{\mathrm{opt}} = (\mu_0, \ ..., \ \mu_{L-1})^{\mathrm{T}} \tag{5.74}$$

The mean square error of ALE output is $I = E[\varepsilon^2(k)]$. When $\boldsymbol{w} = \boldsymbol{w}_{\mathrm{opt}}$, we have

$$I_{\ \min} = E[x^2(k)] - \boldsymbol{w}_{\mathrm{opt}}^{\mathrm{T}} \boldsymbol{R}_{yy} \boldsymbol{w}_{\mathrm{opt}} = E[x^2(k)] - \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} \mu_k \mu_l R_x(k-l) \tag{5.75}$$

For the ALE, Eq. (5.73) can be solved directly. Note that

$$x(k) = s(k) + n(k), \quad s(k) = A\cos(2\pi f_0 k T_s + \varphi)$$

We have

$$\begin{aligned} R_s(k) &= \frac{A^2}{2}\cos(2\pi f_0 k T_s) \\ R_n(k) &= \sigma_n^2 \delta(k) \end{aligned} \tag{5.76}$$

where $\delta(k)$ represents the Dirac function. Therefore Eq. (5.73) becomes

$$\sigma_n^2 w(l) + \sum_{k=0}^{L-1} w(k) A^2 \cos[(k-l)\alpha]/2 = A^2 \cos[(l+d_0)\alpha]/2 \ (l = 0, \ ..., \ L-1) \tag{5.77}$$

where $\alpha = 2\pi f_0 T_s$. It is proved that[31] the approximate solution of Eq. (5.77) is given by

$$\mu_k = w_{\mathrm{opt}}(k) \approx Q\cos[(k+d_0)\alpha] = Q\cos(k\alpha + \varsigma) \ \ (k = 0, \ ..., \ L-1) \tag{5.78}$$

where

$$\varsigma = d_0\alpha, \ \ Q = \frac{2(SNR)_{\mathrm{ALE,in}}}{2 + L(SNR)_{\mathrm{ALE,in}}} \tag{5.79}$$

and $(SNR)_{\mathrm{ALE,in}}$ is the input signal to noise ratio of ALE. It is obvious

$$(SNR)_{\mathrm{ALE,in}} = A^2/(2\sigma_n^2)$$

The mean square value of the output of the tapped delay line is

$$D = \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} \mu_k \mu_l \boldsymbol{R}_{xx}(k-l) = D_s + D_n$$

where D_s is the signal component and D_n is the noise component. It is easy to show that[77]

$$D_s = \sum_{k=0}^{L-1}\sum_{l=0}^{L-1}\mu_k\mu_l\cos[(k+l)\alpha]$$

$$D_n = \sigma_n^2\sum_{k=0}^{L-1}\mu_k^2$$

We obtain

$$(SNR)_{\text{ALE,out}} = \frac{D_s}{D_n} = \frac{2Lp + L^2 + q}{2(L+p)} \times (SNR)_{\text{ALE,in}} \tag{5.80}$$

where

$$p = \frac{\cos[(L-1)\alpha + 2\varsigma]\sin(L\alpha)}{\sin\alpha} \tag{5.81}$$

$$q = \left(\frac{\sin(L\alpha)}{\sin\alpha}\right)^2 \tag{5.82}$$

The system gain of ALE is

$$G_{\text{ALE}} = \frac{(SNR)_{\text{ALE,out}}}{(SNR)_{\text{ALE,in}}} = \frac{2Lp + L^2 + q}{2(L+p)} \tag{5.83}$$

The similar formula derived by Widrow[78] and Richard[79] is a special case of Eq. (5.83).

From Eq. (5.83) we can know that the system gain of ALE is about the length of the tapped delay line and to be exactly the number of taps. It is necessary to show that the theoretical study has shown that there exists so-called "iterative noise for the adaptive filtering"; the longer the tapped delay line is, the larger the iterative noise will be. Therefore the iterative noise becomes a factor which constrains the system gain. When the input SNR is very small, the system gain will decrease. This problem cannot be neglected.

5.5.3 Combined Architecture of ANC and ALE

In order to maximize the effect of ALE, it is necessary to increase the input SNR of ALE, since the system gain of ALE is the function of the input SNR. When the input SNR is lower than a certain value, the gain in ALE will decease. Therefore it

is reasonable to put an ANC before ALE.

A combined ANC and ALE system is illustrated in Fig. 5.34. The original signal is input to an ANC filter to suppress the background noise, and then is sent to the ALE system to enhance the single frequency component. The cascade system will considerably improve the LOFAR and DEMON performance.

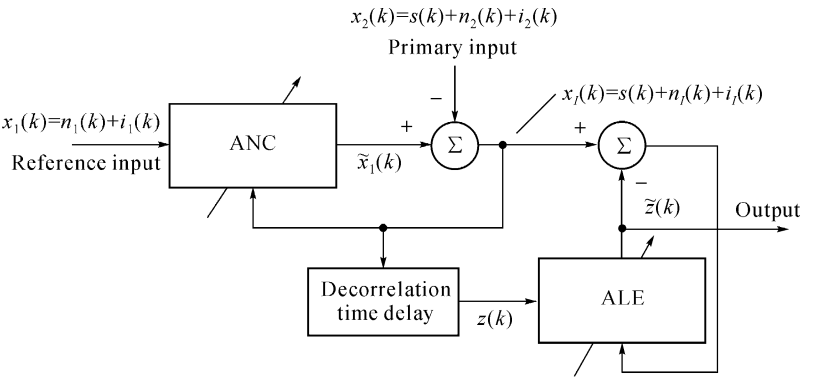


Fig. 5.34. Combined ANC and ALE

The scenario we consider here is a little more complicated than before. The beam output signal contains a weak signal to be detected, background noise and plane interference. The reference input for the ANC+ALE system is

$$x_1(k) = n_1(k) + i_1(k)$$

where $n_1(k)$ is background noise, $i_1(k)$ is directional interference. The primary input is the beam data, which steers the target

$$x_2(k) = s(k) + n_2(k) + i_2(k)$$

where $s(k)$ is the sinusoidal signal or the summation of several sinusoidal signals with different frequencies and phases, $n_2(k)$ is the background noise and $i_2(k)$ is the same interference as $i_1(k)$, but with a different time delay and amplitude. The first stage and second stage of the ANC+ALE system are illustrated in Figs. 5.35 and 5.36, respectively.

Fig. 5.35 is the block diagram of the first stage of the ANC system. Fig. 5.36 is the block diagram of the second stage of ALE. Denote the output of ANC by $x(k)$ and delay it with Δ. The output of ALE is shown in Fig. 5.36 by $\tilde{z}(k)$. Assume the system gain of ANC and ALE is G_{ANC} and G_{ALE}, respectively. Then the total gain of the cascade system is

$$G = G_{\text{ANC}} G_{\text{ALE}} \tag{5.84}$$

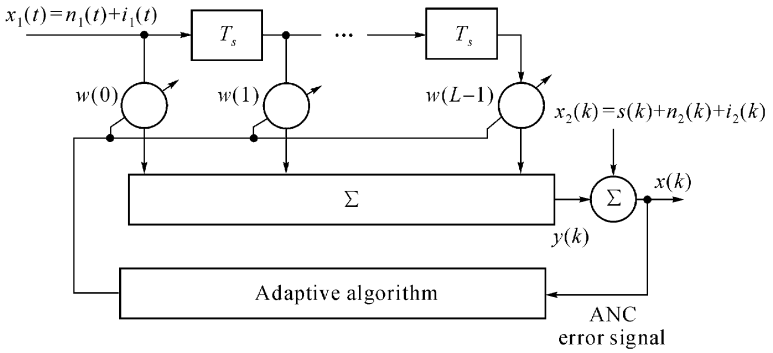


Fig. 5.35. Block diagram of adaptive noise canceling system

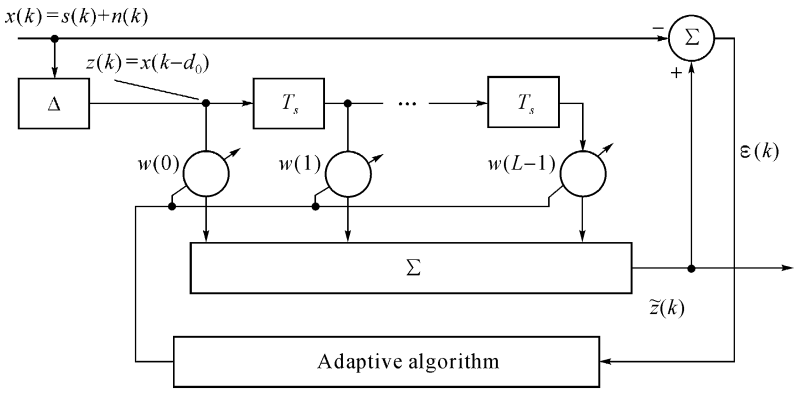


Fig. 5.36. Block diagram of implementation of ALE

It is obvious that

$$(\mathrm{SNR})_{\mathrm{ALE,in}} = (\mathrm{SNR})_{\mathrm{ANC,out}}$$

The transfer function of ANC is

$$H(f) = \sum_{l=0}^{L-1} w(l) \exp(-2\pi \mathrm{j} l T_s)$$

Suppose the tapped delay line is long enough for the difference between $H(f)$ and $H_{\mathrm{opt}}(f)$ to be neglected.

The output of ANC is

$$x(k) = x_2(k) - \tilde{x}_1(k) = s(k) + n_2(k) - \tilde{n}_1(k) + i_2(k) - \tilde{i}_1(k) = s(k) + n(k) + i(k)$$

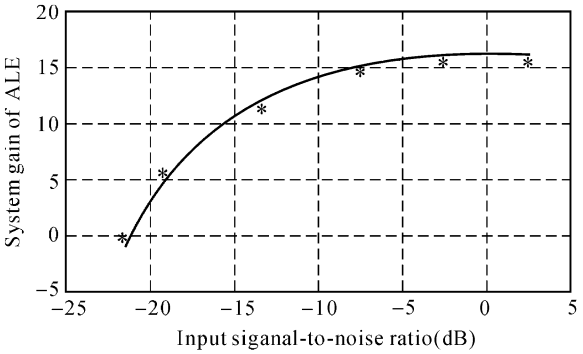
where

$$n(k) = n_2(k) - \tilde{n}_1(k), \quad i(k) = i_2(k) - \tilde{i}_1(k)$$

The mean square error of the output is

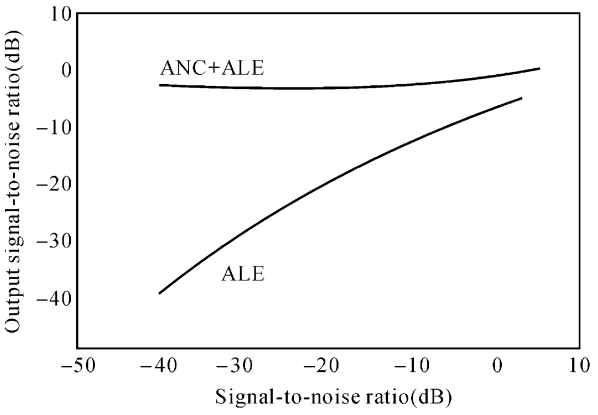
$$\begin{aligned} J &= E[x^2(t)]\,|_{H(f)=H_{\text{opt}}(f)} \\ &= E[s^2(k)] + E[(n_2(k)+i_2(k))^2] - \int_{-\infty}^{\infty} |H_{\text{opt}}(f)|^2 K_{x_1}(f)\mathrm{d}f \end{aligned} \tag{5.86}$$

Figs. 5.37 – 5.39 present the results of theoretical calculation and system simulation. In Fig. 5.37 we can see that the system gain of ALE depends on the input signal-to-noise ratio. In the case of $(\text{SNR})_{\text{ALE,in}} \geq -10$ dB, the gain will tend to be a constant. Figs.5.38 and 5.39 show that the ANC+ALE system approximately has a uniform system gain.



$x(t)=A\cos(2\pi ft)+n(t)$
$\mu A^2/2=0.0000125$ $L=64$
iteration number =1500

Fig. 5.37. System simulation results of ANC+ALE (solid line is the theoretical result)



$x_1(t)=n_1(t)+i_1(t)$
$x_2(t)=s\,(t)+n_2(t)+\,i_2(t)$
$i_2(t)=i_1(t-\tau)$
Band width index=1

Fig. 5.38. Comparison of ANC+ALE and ALE (Broadband interference)

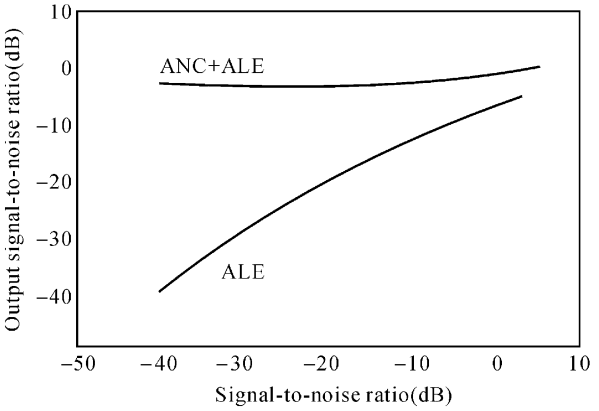


Fig. 5.39. Comparison of ANC+ALE and ALE (Narrowband interference)

5.6 Post Processing of Multibeam Data and Wild Value Filtering

After necessary post processing, the accumulated results of multibeam data will be sent to the consol / display system and displayed on the screen in one or two dimensions. If the input signal of the sonar system completely satisfies the desired model, then the system output data are the random signal and obey some kind of probability distribution. Therefore the post processing will be relatively simple. But if the input data is accidentally interfered with, then the output data is not the stationary random process. Some kinds of fluctuation or random impulse noise may appear. The so-called "wild value" problem may result in a random blank. Therefore wild value filtering is an important topic in post processing of the sonar signal. Almost the whole realization of the sonar function, including target bearing, passive ranging, active echo detection and target classification / identification, depends on the effective operation of wild value filtering.

5.6.1 Square Detection and Absolute Value Detection

The position of wild value processing just follows the post processing of beam data. Fig. 5.40 presents the exact position of wild value filtering. The unprocessed beam data, which already accumulated in post processing, is input to the wild value filter. And the processed beam data, after wild value filtering, is sent to the display system.

In a digital sonar system, the most popular algorithm for post processing is square detection, the optimum detection of a Gaussian signal in the background of Gaussian noise. Of course another detection technique can be chosen, for example

absolute value detection, since the implementation of absolute value detection is more convenient than square operation in digital form. It is proved that, theoretically, the detection performance of square operation and absolute value operation is the same (see Fig. 5.41)

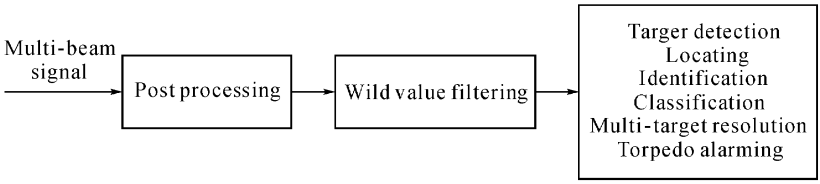


Fig. 5.40. The function of wild value filtering

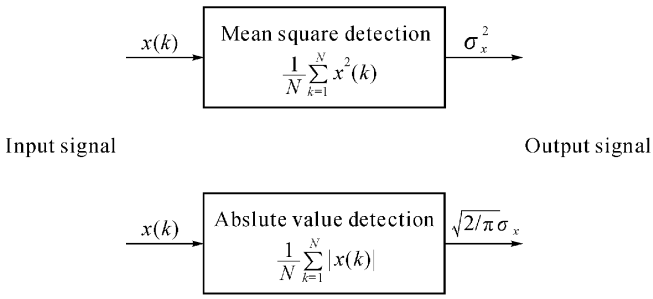


Fig. 5.41. Two different ways of post accumulation

For square detection, suppose the input data series is $x(k)$

$$\overline{x}^2 = \frac{1}{N}\sum_{k=1}^{N} x^2(k) \tag{5.87}$$

The absolute value detection completes the operation

$$\overline{y} = \frac{1}{N}\sum_{k=1}^{N} |\, x(k)\,| \tag{5.88}$$

If $x(k)$ obeys Gaussian probability distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left(-\frac{x^2}{2\sigma_x^2}\right)$$

Then $\overline{x}^2 = \int_{-\infty}^{\infty} x^2 p(x)\mathrm{d}x = \sigma_x^2$

when the signal presents $\overline{x}^2\,|_{s+n} = \sigma_s^2 + \sigma_n^2$; if the signal is absent, then $\overline{x}^2\,|_n = \sigma_n^2$.

For absolute value detection

$$\overline{y} = \int_{-\infty}^{\infty} | x | p(x) \mathrm{d}x = \sqrt{\frac{2}{\pi}} \sigma_x$$

When the signal is present

$$\overline{y} \mid_{s+n} = \sqrt{\frac{2}{\pi}} \sqrt{\sigma_s^2 + \sigma_n^2}$$

When the signal is absent

$$\overline{y} \mid_n = \sqrt{\frac{2}{\pi}} \sigma_n$$

We can see that the results of square detection and absolute value detection are the same, except as a constant factor.

5.6.2 A Method of Stationarity for Non-stationary Data

The signal after post processing may be not a stationary random signal. In this case, the theory and technique used to be applied in stationary data cannot be used. But there is one class of non-stationary signals, which can be stationarity by a relative simple method. The model we are discussing is as follows.

Assume the non-stationary signal is

$$x(k) = d(k) + s(k) \tag{5.89}$$

where $d(k)$ is a deterministic signal but $s(k)$ is a stationary signal. Therefore $x(k)$ is usually not a stationary signal anymore. We can use a difference method, step by step, to obtain a new stationary signal.

First, observe the simplest situation. Assume $x(k) = d_0 + s(k)$ where d_0 is a constant. This is a stationary random model with mean value d_0. It is easy to make this signal with zero mean by using the difference method. Denote

$$y(k) = x(k) - x(k-1) = s(k) - s(k-1) \tag{5.90}$$

This is the zero order difference.

If

$$x(k) = d_0 + d_1 k + s(k)$$

then the first order difference is

$$y(k) = x(k) - x(k-1) = s(k) - s(k-1) + d_1 \tag{5.91}$$

Differences $y(k)$ again and we can obtain the stationary signal with zero mean. Generally speaking, if $x(k)$ is a difference stationary signal of first order

$$x(k)-x(k-1)=s(k) \tag{5.92}$$

then we can know

$$x(k)=x(0)+\sum_{l=1}^{k}s(l)$$

If $x(k)$ is a difference stationary signal of second order, then

$$y(k)=x(k)-x(k-1)$$

it is not a stationary signal. But

$$z(k)=y(k)-y(k-1)=s(k) \tag{5.93}$$

is a stationary signal. And

$$x(k)-x(k-1)-x(k-1)+x(k-2)=s(k)$$

We obtain

$$x(k)=x(0)+k(x(0)-x(-1))+(\sum_{l=1}^{k}s(l))k+\sum_{l=1}^{k}(1-l)s(l) \tag{5.94}$$

By repeating use this method, it is possible to obtain stationarity of a non-stationary random signal, which is a difference stationary model. If the difference in order is no more than 2, then we can directly calculate the deterministic components in terms of the input data series.

Example

Suppose $x(k)=d(k)+s(k),\ \ d(k)=\beta_0+\beta_1 k$. Denote

$$\mu_x=\frac{1}{N}\sum_{k=1}^{N}x(k) \tag{5.95}$$

We can find β_0 and β_1 by minimal least square method

$$\begin{aligned}\beta_0&=\mu_x-\beta_1\overline{t}\\ \beta_1&=\frac{\sum_{k=1}^{N}(x(k)-\mu_x)(t(k)-\overline{t})}{\sum_{k=1}^{N}(t(k)-\overline{t})^2}\end{aligned} \tag{5.96}$$

where $t(0)=1,\ t(k)=1+k/N\ \ (k=1,\ ...,\ N)$.

$$\overline{t}=\frac{1}{N}\sum_{k=1}^{N}t(k)=\frac{1}{2}\left(3+\frac{1}{N}\right) \tag{5.97}$$

By using the similar method, we can calculate the deterministic component in non-stationary data.

5.6.3 Basic Technique of Wild Value Filtering

The wild value filtering technique in sonar design is a relative new topic. The complete theoretical framework, including signal modeling, algorithm and engineering implementation, has not been established yet. Fig. 5.42 shows the basic algorithm used in one-dimensional and two-dimensional data. There are basic assumptions for wild value filtering, such as the presence of the wild value is not continuous, but is accidental, random and discontinuous. Therefore it is separable from the signal. If the presence of the wild value is of a very high density, the question becomes complicated, since it is hard to distinguish a high density wild value from signal fluctuation resulting from signal propagation. In active sonar signal capture and echo signal detection, it is also hard to distinguish the high density wild value and the active sonar signal. Once this situation has occurred, it is necessary to adopt a new method, which is different from the wild value filtering algorithm described here.

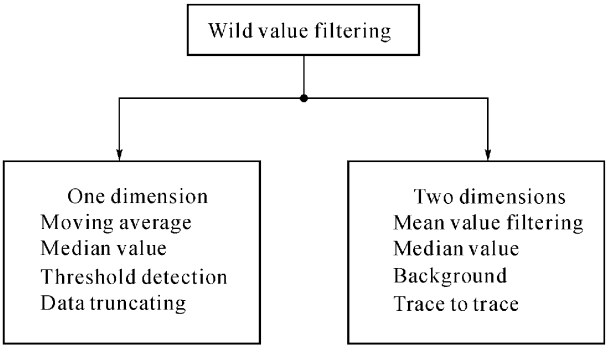


Fig. 5.42. The basic algorithm of wild value filtering

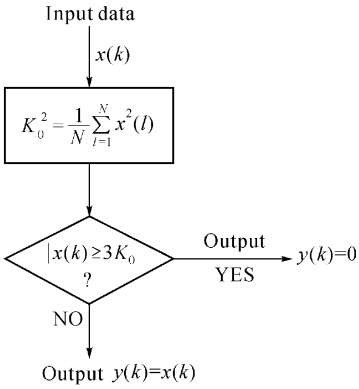


Fig. 5.43. Wild value filtering: threshold detection

For threshold detectionSee Fig. 5.43, assume the input signal $x(k)$ is Gaussian stationary data, the standard deviation is

$$K_0=\left[\frac{1}{N}\sum_{k=1}^{N}x^2(k)\right]^{1/2} \tag{5.98}$$

If the input value of this series is larger than $3K_0$, then we can think this is an abnormal value, since, for the Gaussian signal, the probability of a value larger than $3K_0$ is 0.3% and it should be taken away. This procedure is called threshold detection.

1) Moving average

The aim of a moving average is to smooth the input data, so that any accidental change in values can be suppressed (see Fig. 5.44) Suppose the input data is $x(k)$, the window length of the moving average is N=2M+1. The result of the moving average is

$$y(k)=\frac{1}{N}\sum_{l=1}^{N}x(k-l) \tag{5.99}$$

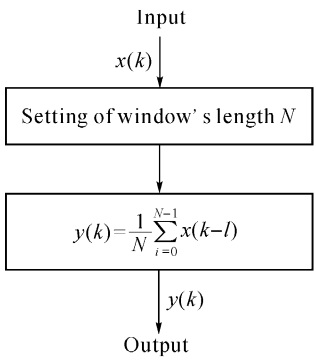


Fig. 5.44. Wild value filtering: moving average

It is necessary to show that the threshold detection and the moving average are different. In the threshold situation, most of the output data is the same as the input, but in the case of the moving average, the output data is totally different from the input data.

2) Mean value filtering

Mean value filtering is recognized as an efficient algorithm to take away accidental impulse interference. Especially in two-dimensional image processing, the random blank in the images can be removed by means of value filtering.

In Fig. 5.45, suppose the input data is $x(k)$. The window data we selected is $x(k-M)$, ..., $x(k)$, ..., $x(k+M)$, rearrange this set of data according to a magnitude from small to large: $z(1)$, ..., $z(2M+1)$. The value in the middle position $M+1$ is the output value of mean value filtering:

$$y(k) = z(M+1) \tag{5.100}$$

We can see that if the occurrence density of random impulse interference is relative low, then the wild value will be effectively removed by this method, since it is impossible that impulse interference would become a mean value.

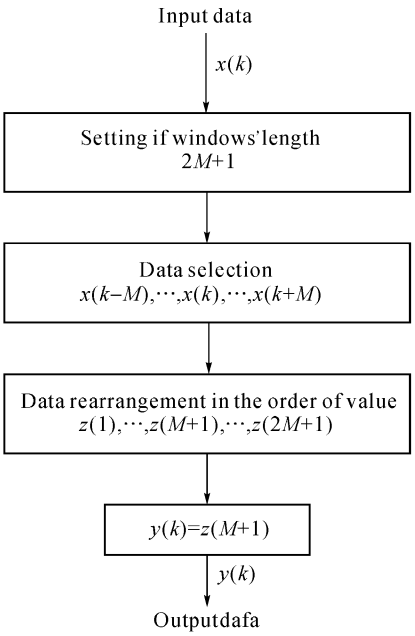


Fig. 5.45. Wild value filtering: median value filtering

5.6.4 Wild Value Filtering for Two-Dimensional Time / Bearing Data

Two-dimensional data processing in a sonar system is different from general image processing. It is necessary to keep the original time / bearing display image. Any change in the image should be local and only small. This is because the time / bearing display is the basis of localization and target tracking. The pixels on the display screen contain information of the target movement. The goal of wild value filtering is to strengthen the validity and reliability of this information.

In Fig. 5.46, two methods are discussed for image processing of a time / bearing display. One is the background equalization, which focuses on the bearing dimension, the other trace to trace correlation, which focuses on the time dimension.

The detailed theory and application of the background equalization technique will be discussed in the next section.

Fig. 5.46(a) is the block diagram of the background equalization. In any moment t, perform the equalization procedure for the data $x(t, \theta)$, where θ is in the range of 0° – 360°. In will takeaway any abnormal values and local fluctuation, etc. The point to point processing will guarantee the main lobe direction and side lobe position, and obtain a new time / bearing display image $y(t, \theta)$.

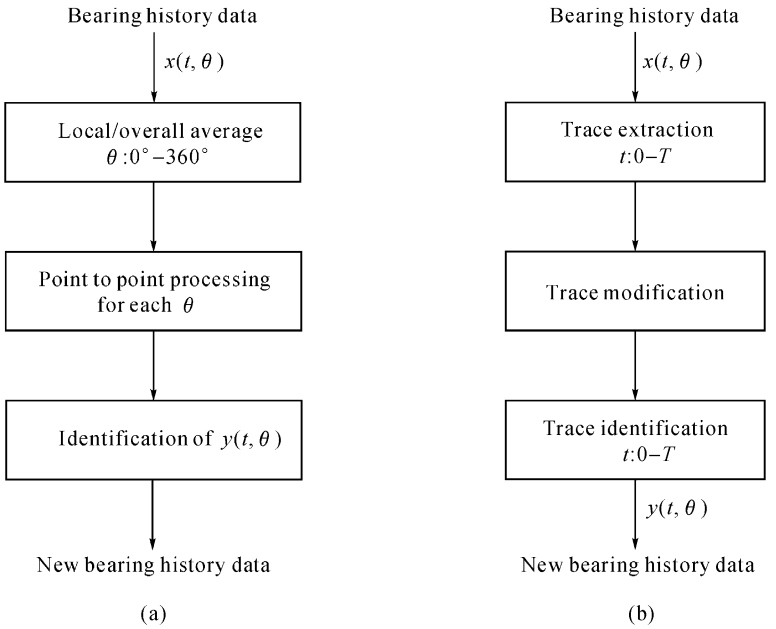


Fig. 5.46. Two dimensional wild value filtering of bearing / history data. (a) Background equalization; (b) Trace to trace correlation

Fig. 5.46(b) is the block diagram of the trace to trace correlation. The main function of this technique is to extract the trace of a moving target. In a small

range of θ, the data $x(t, \theta)$ is processed and the trace of moving target is extracted, using the trace correlation technique, removing the blank pixel and adding some necessary pixels into the trace, such that the trace of the moving target is a continuous curve. The new time / bearing display image will reflect the real situation of target movement.

Fig. 5.47 shows examples of the wild value filtering algorithm discussed in this section. The window length used in the simulation is $N = 17$.

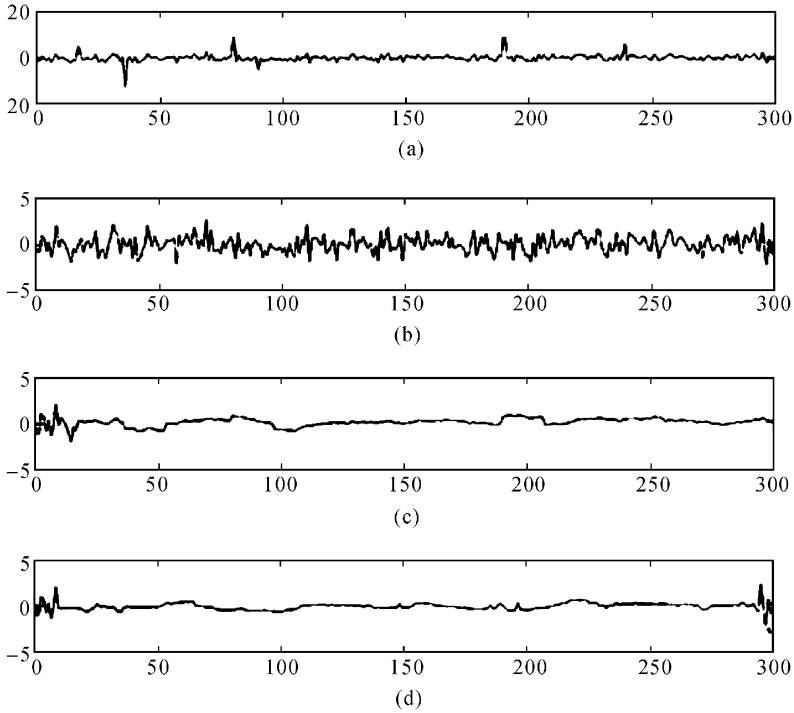


Fig. 5.47. Examples of wild value filtering. (a) Original signal; (b) Threshold detection; (c) Moving average ($W = 17$); (d) Meadian value filtering ($W = 17$)

5.7 Display and Control: High Resolution Gray Scale Display and Background Equalization

The display and control system is one important part of a sonar system. In Fig. 5.4, we have seen that from the sonar operator's point of view, the display / control system is in a central position, since, for the sonar operator, he / she can consider the signal processor in a sonar system to be a black box. They needn't know how the data is transmitted, processed and stored in a sonar system. But the display /

control system is the interface between a sonar operator and a sonar system, i.e., the human / computer interface (HCI). The design of the consol is related to human / machine engineering knowledge[18,80-82].

The display format and contents form the basis of sonar performance. A good sonar designer must be a master at configuring the display and control system.

5.7.1 Main Function of the Display / Control System

The display / control system is the basic equipment that the sonar operator needs to get information and make decisions. At the same time, it is also the equipment for information interchanges between the sonar and the sonar system. A sonar operator manages the digital sonar through a display / consol system, and also accepts feedback information from the sonar.

Therefore, the main tasks of a display / control system for sonar are as follows:

(1) To provide a fast effective operation interface for the sonar operator, decrease the probability of wrong operation, avoid a fetal wrong operation, reduce the labor costs of the operator and increase the operation speed.

(2) To clearly and simply provide necessary information to the sonar operator, increase the detection probability and decrease a false alarm probability.

The theory of human / machine engineering provides a way to satisfy the above requirements. Fig. 5.48 presents the position of a display / control system in an entire sonar.

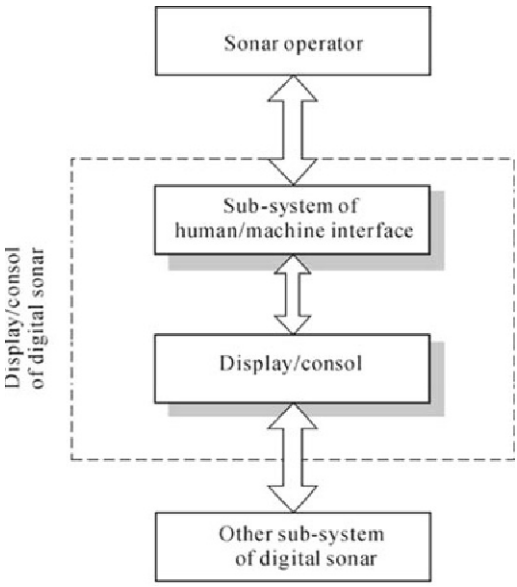


Fig. 5.48. Human / machine interface of digital sonar

The contents to be displayed in a digital sonar system are two kinds. One is functional information flow and one is management information flow.

The functional information flow includes:

(1) History of multiple transmitting of active sonar;
(2) Bearing / amplitude multi-beam display;
(3) Time / bearing multi-beam display;
(4) Target tracking information;
(5) Target parameters (series number, bearing, frequency etc.);
(6) Platform status;
(7) Result of target identification;
(8) Hostile / friend tactical status;
(9) Torpedo alarming;
(10) Results of line spectrum enhancing;
(11) LOFAR (time / frequency);
(12) DEMON (time / frequency);

The management information flow includes:

(1) Choice of operation format;
(2) Selection of various parameters (frequency band, active sonar signal, tracking method, etc.);
(3) Result of self failure test;
(4) Wet end working status;
(5) Interface working status;
(6) Monitoring results of platform noise;
(7) Monitoring results of acoustic or non acoustic sensors.

The display format can be image, figure and symbol.

5.7.2 *Applications of Human / Machine Interface Theory*

Human / machine interface theory is formed in the procedure of computer development. The user sends a command and data through the human / machine interface. After the computer processes these commands and data, the interface feeds back some information to the user. Therefore the interface actually plays the role of input / output function in a computer. Also, the user gives the control mission to the computer. Digital sonar is a special kind of computers; the design of the human / machine interface can completely utilize the knowledge stored in a computer.

Human / machine interface theory consists of many human-oriented topics, including:

(1) Knowledge psychology from a psychological point of view, studying the principles of human / machine interchange.
(2) Human / machine engineering from a system engineering and psychological point of view, studying how the sonar designer and manufacturer can meet the

requirements and extend the human ability to handle them.
(3) Software human / machine engineering. Apply and extend the software engineering.
(4) The design of the display / control system of the sonar has to take into account the following principles:
(5) According to user-based principles, it is necessary to design a user friendly interface. The sonar operator is well trained, but not an expert. Therefore it is not necessary to have many hints on the screen and many questions to be answered.
(6) Uniform principle: Any program or subroutine should have a uniform operation format. For example, the pop down menu must have the same format.
(7) Feed back information: It is necessary to have feedback information for each operation. The wrong operation should prompt hints and provide correct keystrokes.
(8) Fetal operation has to be avoided. For example, it is necessary to have a special measure for protecting the transmitting key.
(9) Use menu and Yes / No question to reduce confusion.
(10) Alarming method: sound or light.
(11) Reduce the number of keys as much as possible.

5.7.3 *Use of Color in Sonar Display*

In early sonar systems, the main method of obtaining information for the operator was auditory listening. The human hearing ability is the key method in target bearing location. Until now, the hearing ability of a well trained sonar operator is still the most important means of target detection, especially for target identification / classification.

A modern sonar system has a comprehensive display system, which provides a wide variety of information to the operator. And also, it is usually has an interface with the commander, who receives / transfers necessary information. This kind of technical innovation brings an incredible change to sonar detection. The renew frequency of a modern displayer can be 120 frames per second, and the visual residual effect of the human being is 0.1 s. Therefore the blank phenomenon does not exist any more. And the operator can do the job for a relatively long time without feeling tired. Since the improvement in storage techniques, we can obtain an image of as much as the previous 30 min. It is possible to considerably increase the detection ability.

Rasmussen et al. studied the theory of the sonar display format in detail and carried out a large amount of experiments[80-82]. The analysis of the experimental results shows that the limit of the resolution of the human eye is about 1'. Suppose the distance from eye to screen is 50 cm, then two points on the screen with a distance of 0.02 cm can be distinguished. A 17-inch screen (about 43.18 cm), when

the resolution is 1600×1280 should satisfy this requirement. Nowadays, the modern high resolution screen can provide a much better resolution than this.

Now the question is which color is the best in a sonar display. The answer, after many experiments, is a gray color. Experimental results show that the color coding could not provide a lower detectable signal-to-noise ratio. The intensity sensitivity of the human eye is optimum with a gray color. There are two advantages using a gray color. First, it provides a lower detectable signal-to-noise ratio. Second, it does not result in a tired eyes feeling in just a short time. Therefore, in modern sonar, almost all sonar screens use a gray color display. The study shows that a 2^8 level gray scale display is almost perfect.

Of course, a different color can be used to mark or indicate a symbol, if necessary.

5.7.4 Gray Scale Conversion

The signal processor in a sonar system usually has at least 16-bit effective digits, regardless of fixed or floating operation. But the gray color display has only 8-bit effective digits. Therefore it is necessary to perform digital conversion before the data is sent to the display system. That is to say, we have to convert the beam data, which contains the amplitude information, to the intensity value of the gray color. This procedure is called GSC (gray scale conversion). Fig. 5.49 illustrates this operation.

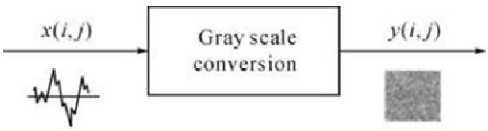


Fig. 5.49. Gray scale conversion (GSC) for digital sonar display

Suppose a system has N channels. In each channel the signal-to-noise ratio is $(\mathrm{SNR})_{\mathrm{in}}$. If the background noise in each channel is independent then in the beam which steers to the target the signal-to-noise ratio is about $N(\mathrm{SNR})_{\mathrm{in}}$, if we want to display the data of this beam without distortion. For example, $(\mathrm{SNR})_{\mathrm{in}}$ =1 (10 dB). Then for a 16-bit system, $10N$ should correspond to 2^{16}=65636 and the conversion coefficient is

$$k = \frac{2^{16}}{10N}$$

when $(\mathrm{SNR})_{\mathrm{in}}$ =0.01 (–20 dB), then the output value for the same beam is

$$N(\mathrm{SNR})_{\mathrm{in}} \times k = 0.01 \times \frac{2^{16}}{10} \approx 66$$

If we convert number 66 (in total 16-bit digit) to 8-bit gray scale value, it is obvious that it should be equal to 0. Therefore we know, if we just take care of the large signal-to-noise ratio value and neglect the small signal-to-noise ratio, the detection performance should be decreased.

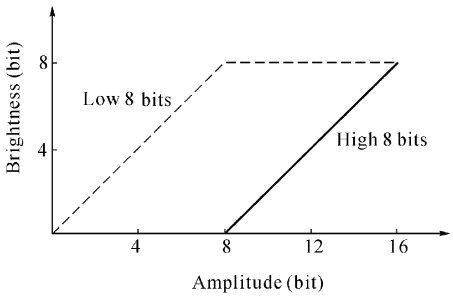


Fig. 5.50. Simplified GSC algorithm

Fig. 5.50 presents a simplified algorithm for performing GSC. In realizing 16-bit to 8-bit GSC, it takes away digits higher than 8 bits or lower than 8 bits. Although this kind of GSC works in a screen display, the performance of the sonar detection is degraded. It cannot detect the weak signal or decrease the target resolution ability.

It is our hope to find a good GSC algorithm, which can reduce the decrease in performance as much as possible.

Suppose x is a 16-bit number, i.e.,

$$x = A_0 2^{15} + A_1 2^{14} + ... + A_{14} 2 + A_{15} \tag{5.101}$$

The direct linear conversion is

$$y = \frac{x}{2^{16}} \times 2^8 = \frac{x}{2^8} \tag{5.102}$$

This is actually the algorithm of "taking the higher 8 bit". It is shown in Chapter 2 that the A-law and μ-law compending can be used in the GSC algorithm (see Eqs. (2.22) and (2.23)). The advantage of compending is that it is possible to guarantee weak signal detection and simultaneously reduce the effect of signal detection with a large signal-to-noise ratio.

Example (μ -law compending)

The relationship of the input x and output y of μ -law is

$$y = \frac{\ln(1+\mu x)}{\ln(1+\mu)} \quad (-1 \le x \le 1) \tag{5.103}$$

Suppose x is 16-bit and y is 8-bit, then

$$y = \frac{\ln(1+\mu x / 2^{16})}{\ln(1+\mu)} \times 2^8 \tag{5.104}$$

Fig. 5.51 illustrates the input and output relation for $\mu = 20$.

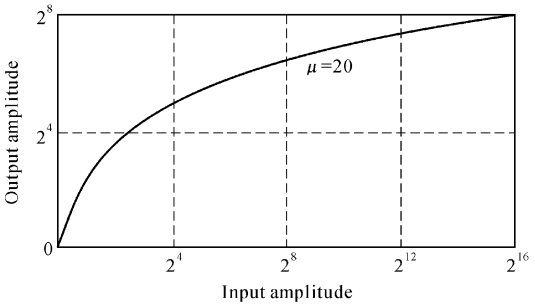


Fig. 5.51. Input and output relation

In realizing μ-law compending, broken line can be used to approximate the logarithm curve given in Eq. (5.103). Generally speaking, a three-segment line is enough. That is to say, one needs to select 4 points (x_i, y_i) $(i = 0, 1, 2, 3)$ and $x_0 = y_0 = 0$, $x_3 = y_3 = 1$.

$$y_i = \frac{\ln(1+\mu x_i)}{\ln(1+\mu)} \quad (i = 1, 2) \tag{5.104}$$

$$y = y_i + \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i) \quad (i = 0, 1, 2) \tag{5.105}$$

If the bit number of the beam data is much larger than 8 bit, for example 24 or even 32 bitss, then directly using the μ-law or A-law is not acceptable. In this situation, a trade-off treatment is recommended, that is "moving μ-law or A law" GSC. That is the beginning point of 16-bit to 8-bit GSC moving, in terms of the statistics of multi-beam results. Fig. 5.52 is an example of using the moving μ-law. Where k_0 is a factor to adjust the beginning point of moving μ-law. We can see that the result k_0=3 is the best.

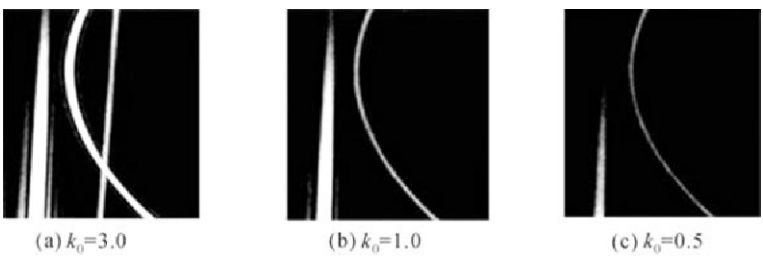


Fig. 5.52. Example of moving GSC display

5.7.5 Background Equalization

When the ambient noise field of the ocean environment is not uniform and homogeneous or even not stationary, then some kinds of fluctuautions in the time / bearing display usually appear. Pieces of strong or weak gray color will interfere with the signal detection or result in false alarming. For the towed array, the receiving acoustic module will receive self noise from its own vessel. It will result in a strong spline light area on the screen. These kinds of interference are different from the wild value discussed in the last section. It is necessary to use a special treatment for the time / bearing image, such that the image becomes "clear" and "clean". The technique is called background equalization. Fig. 5.53 shows the position of background equalization in a sonar system.

The multi-beam data is usually interpolated before being sent to the display. The reason is that the resolution of a modern display screen is very high and if the beam number is too small, one beam will occupy many pixels and the display quality will be worse. The beam interpolation will cancel the blank effect in the intensity display and improve the entire figure quality.

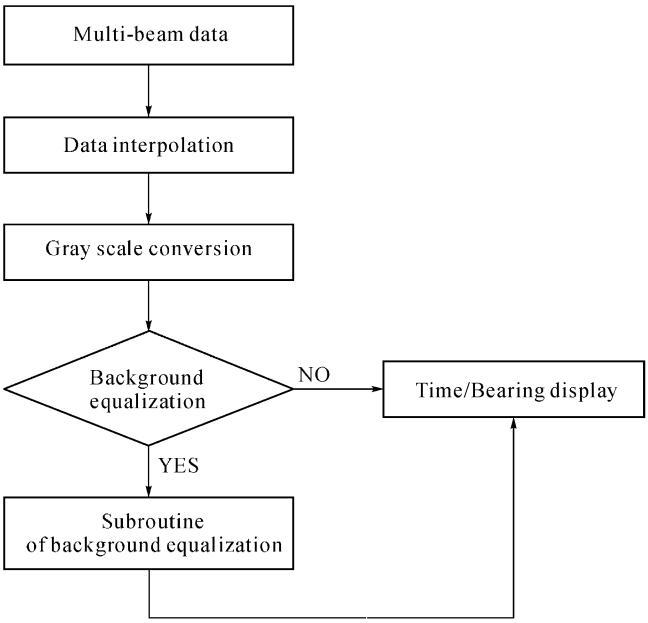


Fig. 5.53. Block diagram of background equalization in a sonar system

A useful technique, proposed by Struzinski[18], is so-called OTA (order truncate average). The method, which combines OTA and the mean value filtering technique, is more effective in background equalization. This technique is convenient to realize in real time. It is actually a row processing program.

Fig. 5.54 is the block diagram of background equalization. Suppose the multi-beam data are $x(1)$, ..., $x(N)$. It can be the direct multi-beam data, and can also be the interpolated data. It is considered as one row in a two-dimension time / bearing display. When the next data series arrives, the old row data should be shifted one row down, and the newcomer should replace the old row data. The background equalization is a point to point process program. A window for processing is selected, assuming the length of this window is $2K+1$. For each point $x(i)$, choose it as the center of the window and then perform necessary processing to get a new point, $x_{\text{new}}(i)$. According to this technique, the point in the margin must be treated specially. And we have to extend the data first:

$$x(K+1),\ ...,\ x(2),\ \ x(1),\ ...,\ x(N),x(N-1),\ ...,\ x(N-K)$$

To the left of $x(1)$ we add K points, in symmetry with $x(1)$, and to the right of $x(N)$, we add K points, in symmetry with $x(N)$. In order to simply the program compiler, we code the extended data in the following:

$$y(1),\ ...,\ y(K),\ y(K+1),\ ...,\ y(K+N),\ y(K+N+1),\ ...,\ y(N+2K)$$

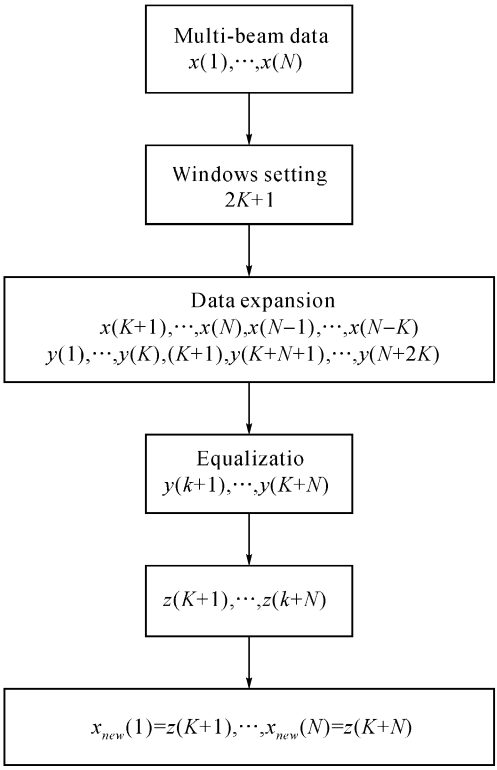


Fig. 5.54. Subroutine of background equalization

The background equalization algorithm is performed for $y(K+1)$, …, $y(K+N)$ and we obtain a new series: $z(K+1)$, …, $z(K+N)$. This new series is just what we are interested in and which should replace the old $x(1)$, …, $x(N)$: $x\mathrm{NEW}(1) = z(K+1)$, …, $x\mathrm{NEW}(N) = z(K+N)$. The algorithm from $y(K+1)$, …, $y(K+N)$ to $z(K+1)$, …, $z(K+N)$ is the mean value filtering and OTA.

Fig. 5.55 illustrates the experimental result of towed array sonar. The effect of background equalization is quite good; a spline of white color of about 40° width disappears after processing.

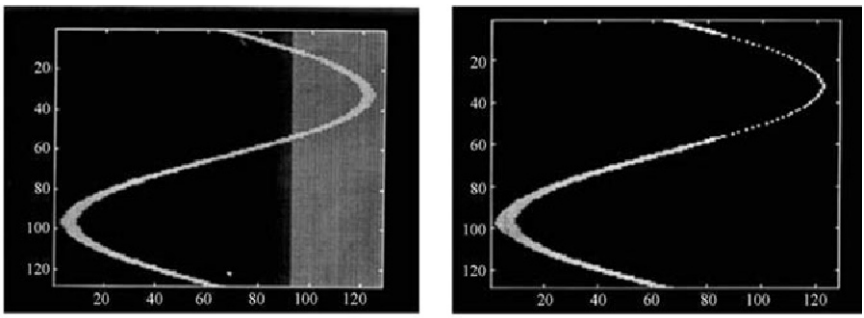


Fig. 5.55. Example of background equalization for the display of a towed array sonar. (a) Before background equalization; (b) After background equalization

5.8 Reliability Design of Digital Sonar

Sonar is an important equipment in the navy. Apart from some use in civil engineering and ocean development, the navy is the most important user of various types of sonar. Therefore sonar is a special military product.

Quality means everything, especially in the case of military sonar. Therefore, reliability design plays a very important role in the entire design procedure. The quality of the product is the summation of product performance and utilization requirements. In evaluating quality for a military product, apart from performance, it is necessary to have specifications for reliability, maintainability, security and affordability.

The goal of reliability design is to satisfy user requirements concerning reliability and maintainability. This includes designing, testing and manufacture.

In this section we will introduce reliability design, reliability prediction and reliability testing in sonar design, development and manufacture.

The literature of reliability design is very plentiful, whether it concerns hardware or software. Most countries have their own national standards or military standards for sonar systems. This usually includes design, testing, manufacture and sea trials[38,40-47].

5.8.1 The Series Standard for Reliability

The reliability standard, which is proposed by the authorities, is the rule and the law which sonar designers and users have to follow. In China, there are three levels of standards related to sonar systems. These are National (GB series), Military (GJB series) and Navy (HJB series) standards. They set the rules for sonar design in detail. And everyone who is involved in sonar design, development, testing and manufacture must follow these rules without exception.

In the USA, the MIL-STD series standard is the main law in sonar design. Serious propositions and standards should be followed in the entire procedure of sonar design, testing and manufacture.

5.8.2 Brief Introduction to Reliability Engineering

In World War II, in the development of the V-I rocket, German scientists first proposed the system reliability concept and applied it to rocket research work.

In 1949, the IEEE in the USA established a reliability technique group. This was the first professional organization which was involved in reliability studies. In the 1950's, in the Korean war, the US military found that unreliable electronic equipment not only affected their fighting ability, but also cost them large amounts. In 1952, a proposal prepared by The committee of Electronic Equipment Reliability for the Department of Defense shows that:

(1) The effective data for reliability of components and equipment can be obtained in the field.
(2) Reliable components are the basis of the reliability of equipment.
(3) It is necessary to have quantitative requirements for the reliability of components and equipment.
(4) A new product must be reliability tested before approval to manufacture is given.
(5) It is necessary to establish a permanent reliability committee.

The US Department of Defense accepted these proposals. Thereafter, a professional committee (Reliability of Electronic Equipment) was established in August 1952. In June of 1956, they published the report of the "Reliability of Military Electronic Equipment". This report became the basic document of reliability engineering. And then a series standard related to reliability was published in the name of the Department of Defense of the USA, including:

(1) MIL-STD-785: Reliability Program for Systems and Equipment Development *and Production*;
(2) MIL-STD-2173: *Reliability Centered Maintenance Requirements*;
(3) MIL-STD-690: *Program and sample of failure rate*;
(4) MIL-STD-781A: *Reliability Test Methods, Plans and Environments for Engineering Development, Qualification and Production*;

(5) MIL-HDBK-217: *Reliability Prediction of Electronic Equipment*;
(6) MIL-STD-338B: *Electronic Reliability Design Handbook*;
(7) MIL-STD-470: *Requirements of maintainability*;
(8) MIL-STD-1629A: *Procedure for Performing a Failure Model, Effects and Criticality Analysis*;
(9) IEEE 1332: *IEEE Standard Reliability Program for the Development and Production of Electronic Systems and Equipments Etc*.
In the UK they have similar documents, such as DEF STAN 00-41 to 00-49.

5.8.3 Hardware Reliability

Fig. 5.56 presents the production flow chart of a digital system, from components purchase to production delivery. First of all, it is necessary to test the components and perform sieving (filtering) and life testing to eliminate unqualified components. And then it begins the manufacture and testing of the printed board. At the same time we should design and manufacture the panel, cabinet and various interfaces. Once completing the system integration, it is necessary to perform the entire system test.

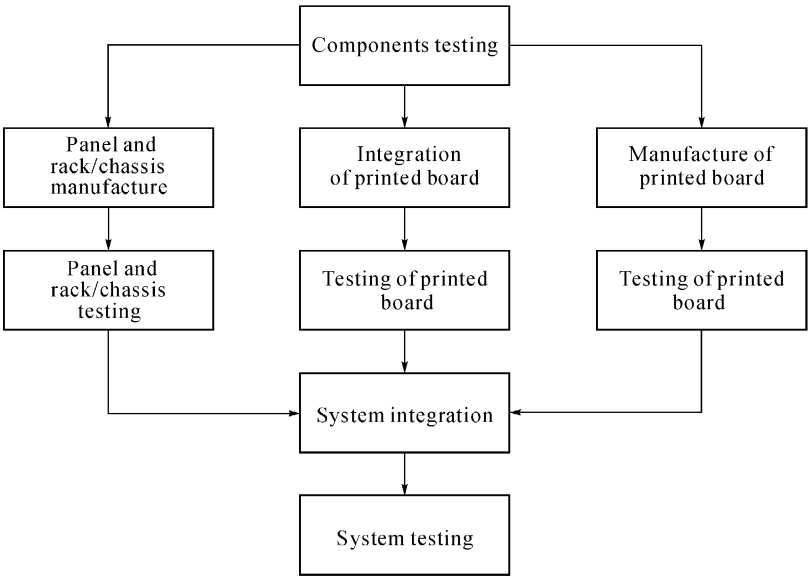


Fig. 5.56. Manufacture flow chart of digital system

A complex system consists of a large amount of components; reliability problems are related to each component. The concept of reliability is really a statistical topic. Let us consider the reliability function of a component. The reliability $R(t)$ is the function of time t. It presents the probability of the reliability

of a component, which works normally from moment 0 until moment t. Based on the statistical definition of probability, it is easy to derive the expression of $R(t)$.

Suppose there are N_0 same components. Assume there are N_s components still working normally after time t, and N_f components have failed in time [0, t]. It is obvious that at any time

$$N_0 = N_s + N_f \tag{5.107}$$

According to the definition of probability

$$R(t) \approx \frac{N_s}{N_0} \tag{5.108}$$

We know that $\mathrm{d}N_f / \mathrm{d}t$ is the number of failed components in unit time, and it can be considered as the failure rate of a component. The transient failure rate is equal to $\mathrm{d}N_f / \mathrm{d}t$ over N_s

$$\lambda(t) = \frac{\mathrm{d}N_f}{\mathrm{d}t}\frac{1}{N_s} \tag{5.109}$$

The relationship between $R(t)$ and $\lambda(t)$ can be found from Eq. (5.107) to Eq. (5.109).

$$\lambda(t) = \frac{\mathrm{d}N_f}{\mathrm{d}t}\frac{1}{N_s} = -\frac{\mathrm{d}R(t)}{\mathrm{d}t}\times\frac{N_0}{N_s} = -\frac{\mathrm{d}R(t)}{\mathrm{d}t}\times\frac{1}{R(t)} \tag{5.110}$$

$$R(t) = \exp\left[-\int_0^t \lambda(u)\mathrm{d}u\right] \tag{5.111}$$

This expression is the basic formula in reliability theory. It presents the relationship between the reliability function and the failure rate function. It also indicates that the reliability function has the form of an exponential function. Now we show the mathematical expression for MTBF, MTTR and MTTF, which we have discussed in Chapter 4. We know that

$$MTBF = MTTF + MTTR$$

It is proved that

$$MTTF = \int_0^\infty R(t)\mathrm{d}t \tag{5.112}$$

Eq. (5.12) is one of the most important formulas in reliability theory. In order to prove it, introduce a new random variable

ξ = {From moment 0, the failure time of component}

Therefore the probability distribution function of ξ is

$$F(t) = P\{\xi < t\} = 1 - P\{\xi \geq t\} = 1 - R(t)$$

MTTF is really the mathematical expectation of ξ, i.e.,

$$MTTF = E[\xi] = \int_0^\infty tF'(t)\mathrm{d}t = -tR(t)\Big|_0^\infty + \int_0^\infty R(t)\mathrm{d}t \tag{5.113}$$

Since $R(\infty) = 0$

$$MTTF = \int_0^\infty R(t)\mathrm{d}t \tag{5.114}$$

This proves Eq. (5.112).
Generaly speaking, *MTTR*<<*MTTF* and therefore

$$MTBF \approx MTTF = \int_0^\infty R(t)\mathrm{d}t$$

Based on experiments in the semiconductor industry, the failure rate function has the typical form of a bathtub (Fig. 5.57). It has an early failure period, then has a relative long normal working period, and finally, there is a wear out failure rate period. And the component cannot be used any more. One can use acceleration tests, burn-in tests and sieving (filtering) to eliminate those components, which are in the period of early failure.

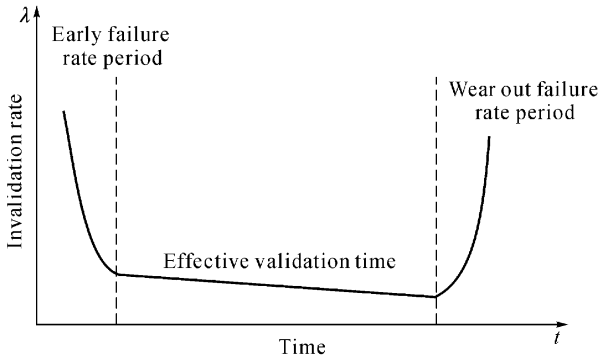


Fig. 5.57. Bathtub curve of component failure

The digital sonar system consists of many different types of components and the connection of these components is usually very complicated. The reliability of a sonar system strongly depends on the reliability of components and their forms

of connection. We will first discuss the reliability calculation in the simplest connection form. The reliability calculation of an integrated system can be simplified in some ways.

Fig. 5.58 presents two simplest connection forms, i.e., series and parallel connections. Assume there are n components (or subsystems) in a series connection. The reliability function of each component is $R_i(t)$ (i = 1, …, n). This system can work properly only in the case of each component working properly. Therefore

$$\begin{aligned} R(t) &= P\{\text{system works well in } [0, t]\} \\ &= P\{\text{All components work well in } [0, t]\} \\ &= \prod_{i=1}^{n} R_i(t) \end{aligned} \tag{5.115}$$

The last step of this equation uses the fact that the events of failure of n components are independent.

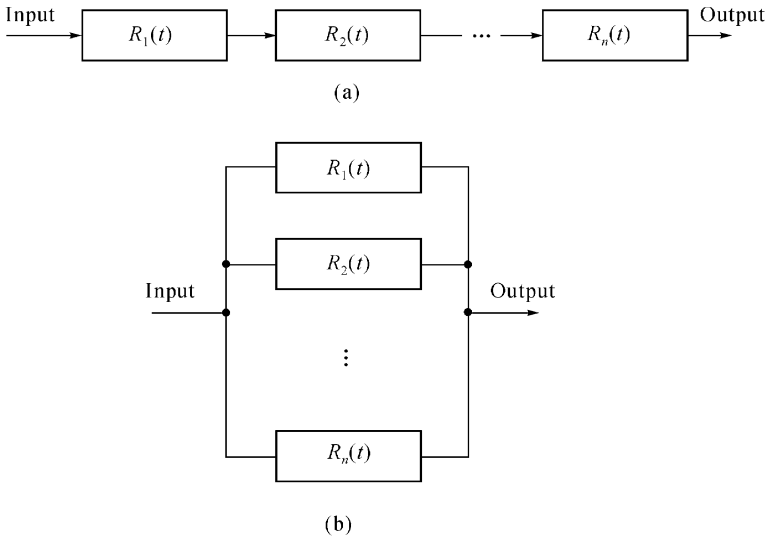


Fig. 5.58. Series and parallel connection of systems. (a) Series system; (b) Parallel system

Now turn to the situation of parallel connection. In this case, the system would work properly if at least one component works well. We have

$$\begin{aligned} R(t) &= P\{\text{system works well in } [0, t]\} \\ &= P\{\text{At least one component works well in } [0, t]\} \\ &= 1 - \prod_{i=1}^{n} (1 - R_i(t)) \end{aligned} \tag{5.116}$$

Eqs. (5.115) and (5.116) are often used to calculate the reliability of a complicated system. A special case is when the failure function does not depend on time t, i.e., it is a constant. In this special case the reliability function will have a simplified exponential form.

The following assumptions are set up:

(1) In a certain short duration Δt, the probability of failure of a component is proportional to Δt.

(2) In duration Δt, the probability of two subsequent failures can be neglected.

(3) The events of failure occurrence are independent. In these assumptions, it is proved that in duration [0, t], the probability of k times failure obeys the Poisson distribution:

$$p_k = \frac{m^k(t)}{k!}\exp(-m(t)) \tag{5.117}$$

where

$$m(t) = \int_0^t \lambda(x)\mathrm{d}x \tag{5.118}$$

In fact, the average failure time in duration [0, t] is

$$\sum_{k=0}^{\infty} k \times p_k = m(t) \tag{5.119}$$

If the failure rate is independent of t, i.e., $m(t) = \lambda t$, then

$$p_k = \frac{(\lambda t)^k}{k!}\exp(-\lambda t) \tag{5.120}$$

It is obvious that

$$R(t) = P\{\text{Component works well in } [0, t]\} = p_0 = \mathrm{e}^{-\lambda t} \tag{5.121}$$

This equation can be also derived from Eq. (5.111).

Example

Suppose $\lambda(t) = \lambda$, then

$$MTBF = \int_0^{\infty} R(t)\mathrm{d}t = 1/\lambda \tag{5.122}.$$

Eq. (5.122) shows that when the failure function is independent of time, the MTBF is equal to the inverse of the failure rate. Fig. 5.59 is an example of the reliability function.

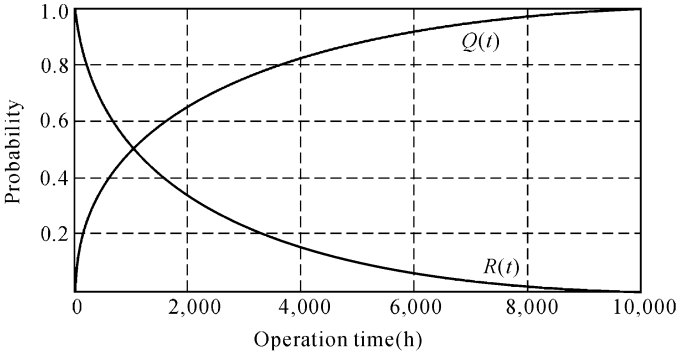


Fig. 5.59. Reliability function and failure function

In the case of $\lambda(t) = \lambda$, the calculation of the reliability of a series or parallel connection system becomes simple. We will list some important results in the following.

(1) Series connection:

Suppose there are n components in a series connection. The reliability function of the i-th component is $R_i(t) = \exp(-\lambda t)$. Therefore Eq. (5.115) becomes

$$R(t) = \exp\left[-\sum_{i=1}^{n} \lambda_i t\right] \tag{5.123}$$

The mean time between failures is

$$MTBF = \frac{1}{\sum_{i=1}^{n} \lambda_i} \tag{5.124}$$

In particular, if $\lambda_i = \lambda$ $(i = 1, \ldots, n)$

$$MTBF = \frac{1}{n\lambda} \tag{5.125}$$

So we know the MTBF of a series connection system with n same components will decrease to $1/n$ of MTBF for a single component.

(2) Parallel connection:

Suppose there are n components in a parallel connection. The reliability function of the i-th component is $R_i(t) = \exp(-\lambda t)$ $(i = 1, \ldots, n)$, then we have

$$MTBF = \frac{1}{\lambda} + \frac{1}{2\lambda} + \ldots + \frac{1}{n\lambda} \tag{5.126}$$

Note that the mean time between failures of two same components is not two times that of a single component, but 1.5 times.

• A parallel connection system only allows one component failure:

Suppose there are n components in a parallel connection, but only one component failure is allowed. The reliability function of the i-th component is $R_i(t) = \exp(-\lambda t)$ ($i = 1, \ldots, n$). Then

$$MTBF = \frac{n}{(n-1)\lambda} - \frac{n-1}{n\lambda} \tag{5.127}$$

• A parallel connection system only allows r components failure:

Suppose there are n components in a parallel connection, but only r components failure is allowed. The reliability function of the i-th component is $R_i(t) = \exp(-\lambda t)$ ($i = 1, \ldots, n$). Then

$$MTBF = \frac{1}{n\lambda} + \frac{1}{(n-1)\lambda} + \ldots + \frac{1}{(n-r)\lambda} \tag{5.128}$$

Note that the Eqs. (5.126) and (5.127) are the special case of Eq. (5.128). Since for $r = n-1$, Eq. (5.128) becomes Eq. (5.126), and for $r = 1$, Eq. (5.128) becomes Eq. (5.127).

In reliability design theory, there is one important concept that is the concept of a spare system. If one subsystem is in failure, then a similar spare subsystem can replace it and the entire system can work well. This is actually the redundancy concept in reliability design. In designing a large complicated system, one should prepare some spare subsystem, in a key position of course. Once a subsystem fails to work, the spare system can replace it. In a certain sense, the spare system is similar to a parallel connection. The difference is that the components in the parallel connection are in working status, but the spare subsystem is not usually in working status.

It can be proved that the reliability function of a system with n same spare components is

$$R(t) = \exp(-\lambda t)\left[1 + \lambda t + \frac{(\lambda t)^2}{2} + \ldots + \frac{(\lambda t)^{n-1}}{(n-1)!}\right] \tag{5.129}$$

$$MTBF = \frac{n}{\lambda} \tag{5.130}$$

The mean time between failures of a system with n spare components, which have a different failure rate $\lambda_1, \ldots, \lambda_n$, is

$$MTBF = \frac{1}{\lambda_1} + \ldots + \frac{1}{\lambda_n} \tag{5.131}$$

5.8.4 Software Reliability

The concept of software reliability was proposed in the late 1970's and attracted much attention in application areas. In some fields, both hardware failure and software failure will result in disaster, for example in the space shuttle mission, the operation of a nuclear power station, etc. In some fields, although the software failure does not result in a disaster, the results are serious, for example, in banking, ticketing systems, etc.

Since software applications grew so fast, so reliability experts began to be concerned with the reliability of software. Nowadays, the definition of software reliability, which is accepted by most researchers, is the probability that the software will run properly during an indicated period in specific circumstances.

Assume n is the total number for running a certain program, n_f is the failure number, then the reliability of this program is defined as

$$R = 1 - \lim_{n \to \infty} \frac{n_f}{n} \tag{5.132}$$

In designing a sonar system, in order to guarantee the correctness of the software, it is necessary to repeatedly test it.

There are three kinds of models for software reliability studies: estimation model, prediction model and measurement model[44].

5.8.5 Technical Means of Hardware Reliability Design

Digital sonar design has to pay attention to both the dry end and the wet end. Since the wet end usually works in very harassing conditions, it necessarily has higher requirements[40]. If we use underwater cable or optical fiber cable, it should have armor to prevent shark bites[45].

In the reliability design of the dry end, several means can be adopted to increase reliability.

(1) Redundancy design: Since the sonar is usually a multiple channel system, a small part failure of channels should not affect the entire system performance seriously. Therefore, if we can afford to increase the number of channels, it will effectively increase the reliability, and the mission reliability specification is easier to attain.

(2) Reduce rated design: In conditions in which a component can operate properly, reducing the operation rate is an important means. For example, if a component can work properly at a voltage of 9 – 12 V, then it is recommended that we choose a power supply of 9 V.

(3) Simplified design: In conditions where there is no loss of system performance, it is necessary to simplify design, adopt a qualified electronic module as far as possible, and increase the size of the same printed board.

(4) The sieving and burn-in testing of components: The sieving and burn-in testing of components is to eliminate those components which show early period failure, as shown in Fig. 5.57.
(5) Rigid design: A rigid design of printed boards, racks and cabinets is necessary; this can prevent damage from vibration, shock and erosion.
(6) Built in failure self test: A real time hardware and software self test is necessary. Real time monitoring can guarantee that a sonar system properly works. Most of modern digital sonar has the ability to real time monitor the failure and the results of monitoring are directly sent to the display / consol system. Therefore the operator knows the running status of the entire sonar. It is very useful that the failure self test for hardware can locate the failure down to a printed board level, even in DSP chips or CPU chips.

As an example, we will show the calculation of the reliability of the submarine sonar. Suppose the channel number is 64, the number of failure channels is less or equal to 5, the sonar system still works well. The failure rate of the hydrophone is

$$\lambda = 2.0\times10^{-5}\ (\mathrm{h}^{-1})$$

Then the failure rate of the array is

$$\lambda_1 = \left[\left(\frac{1}{64}+\frac{1}{63}+\frac{1}{62}+\frac{1}{61}+\frac{1}{60}\right)\times\frac{1}{\lambda}\right] = 2.5\times10^{-4}\ (\mathrm{h}^{-1})$$

The failure rate of cable and penetrator is

$$\lambda_2 = 0.5\times10^{-5}\ (\mathrm{h}^{-1})$$

The total failure rate of the wet end is

$$\lambda\,|_{\text{wet end}} = \lambda_1 + \lambda_2 = 2.55\times10^{-4}\ (\mathrm{h}^{-1})$$

It follows that

$$MTBF\,|_{\text{wet end}} = 1/\lambda\,|_{\text{wet end}} = 3,921\ (\mathrm{h})$$

The calculation of reliability is based on the series connection model. The failure rate of components is listed in Table 5.2.

Total failure rate is

$$\lambda_3 = 777.1\times10^{-6}\ (\mathrm{h}^{-1})$$

$$MTBF\,|_{\text{dry end}} = 1297\ \ (\mathrm{h})$$

The mean time between failures of the entire system is

$$MTBF = \left(\frac{1}{3921}+\frac{1}{1287}\right)^{-1} = 975\ (\mathrm{h}).$$

Table 5.2 Failure rate calculation for the signal processor

Item	Failure rate $\times 10^{-6}h^{-1}$	Quantity	Total failure rate$\times 10^{-6}h^{-1}$
ROM	0.07	150	10.5
RAM	0.60	150	90
DSP chips	0.02	110	2.2
Diode	0.03	10	0.3
Metal resistor	0.015	600	9
Synthesis resistor	0.015	200	3
Isinglass resistor	0.016	200	3.2
Ceramic capacitor	0.007	1,000	7
Tantalum capacitor	0.03	120	3.6
Fans	0.52	4	2.1
Crystal oscillator	0.52	4	2.1
Junction spline	0.26	300	78
Solder hole	0.006	30,000	180
Solder point	0.005	30,000	150
Plug in connector	0.005	4,000	2.0
Plug of cabinet	0.005	3,600	18
MS integrator	0.08	2,400	192

5.8.6 *Failure Self Test*

The failure self test does not belong to the category of reliability design, but it is an important specification in modern digital sonar. One hopes that the probability of failure will be as small as possible, but failure occurrence cannot be avoided. The problem facing a sonar designer is how to monitor the operational situation in real time. Once the failure has happened, the alarm signal immediately appears and the system can be fixed as soon as possible.

There are two kinds of failure self test: the off-line test and the on-line test. In the off-line form, the sonar system is disconnected from other signals. Based on the simulated signal, which is generated by the sonar system itself, one can test the sonar system. For example, beamforming, LOFAR and DEMON analysis, etc. Any standard input and expected output can be precisely tested before, so the system failure can be easily found and located by this method. But the off-line test should be completed in a short time, otherwise it is unacceptable.

The on-line failure test is a real time self test and can be completed by hardware or software. The printed boards of a sonar system are real time tested. Once there is a failure occurrence, the test system will send an alarm signal, either sound / optical or a symbol to the display / control system. The problem is that the test system itself must be tested. Therefore it is impossible to have a perfect system, which is absolutely running normally all the time.

5.9 Integrated Design of Digital Sonar: Multi-sensor, Multi-layer Data Fusion

Modern vessels usually have several sonar systems installed. The sources of target detection come in different ways. In order to maximize the utilization of information from multiple sensors, it is necessary to analyze the data and information from different signal processors. A part of the information synthesis can be carried out by the operator, but most of it has to be done by the sonar system. This function is called data fusion. Data fusion is one of the hot topics in digital sonar research[48-53]. Sometimes it is called "information fusion". we prefer to use the term data fusion since the concept of information not only includes data, but also covers sound, images and speech, etc. In the ocean sciences, data fusion is called "data assimilation".

The data fusion of multiple sensors can be traced back to the early work in radar research. For example, in 1964, Sittler proposed the concept of data association when he studied multiple targets tracking with radar[83]. With the development of radar and sonar signal processing, the data fusion problem attracted more and more interest from researchers.

In this section we will introduce the basic theory of data fusion and its application in digital sonar. And we will also briefly show some results of joint detection and estimation studies.

5.9.1 Basic Concept of Data Fusion

The definition of data fusion given by Hall[49] is: "The integration of information from multiple sources to produce specific and comprehensive unified data about an entity."

Data fusion can be categorized as three levels: the data level, feature level and decision level (Fig. 5.60).

In data level of data fusion, the data from N different information sources are processed to extract the feature of a target and then make a decision. This is the most fundamental level of data fusion. In feature level of data fusion, each channel has its own feature extraction unit. N features are fused to make the final decision. Finally, in decision level of data fusion, each channel independently extracts features of the target and makes a decision. The goal of data fusion is to synthesize N decisions.

The algorithms of data fusion are various, and can mainly be categorized into three topological structures (Fig. 5.61).

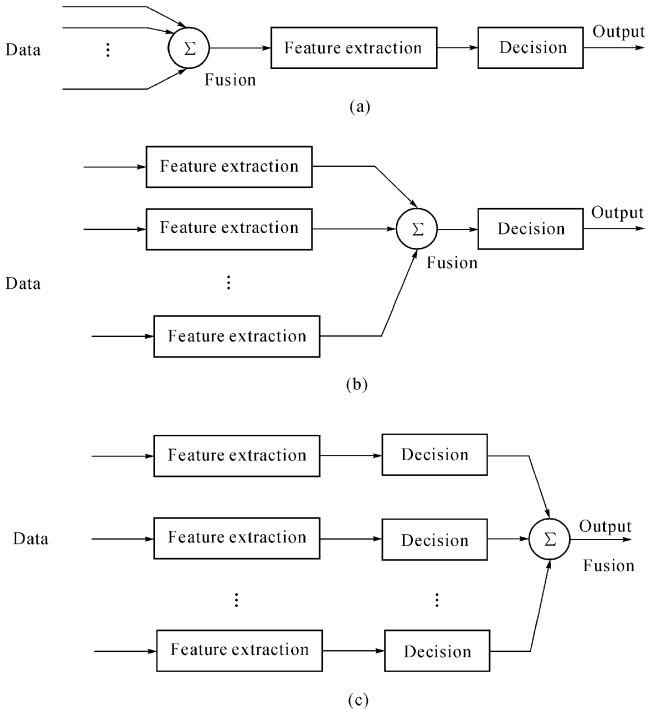


Fig. 5.60. Three categories of data fusion. (a) Data level; (b) Feature level; (c) Decision level

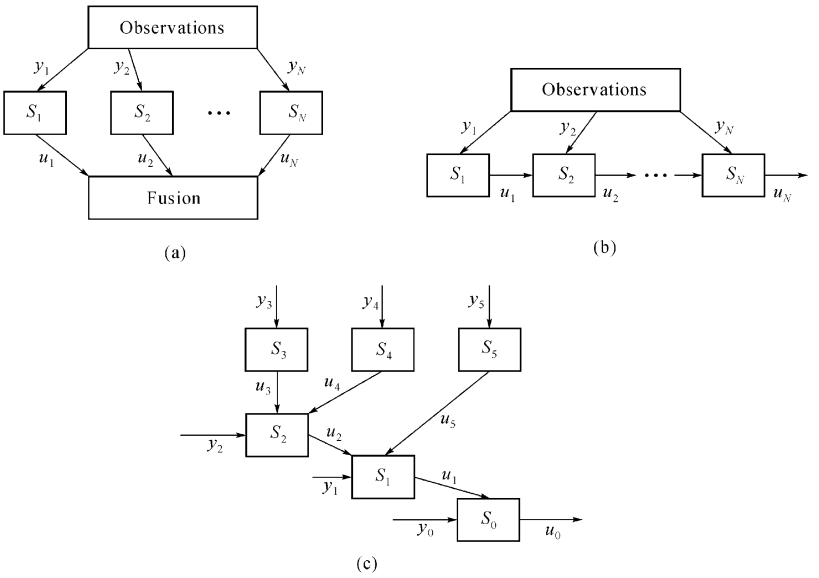


Fig. 5.61. Topologic architecture of data fusion. (a) Parallel topology; (b) Series topology; (c) Tree type topology

5.9.2 Optimum Linear Data Fusion

The basic idea of data fusion is to collect multiple observation results and derive a new result. It is obvious that the results of data fusion must be better, or at least no worse than single observation since the data fusion procedure is meaningless in theory and applications if the result of data fusion is worse than before. We want to prove the basic theorem in data fusion. That means, from the statistical point of view, the result of data fusion should be no worse than before.

First of all, we consider two independent observations. Suppose x, y are the two independent observation results for some parameter θ. Assume

$$E(x) = \theta, \quad E(y) = \theta; \quad \mathrm{Var}(x) = \sigma_x^2, \quad \mathrm{Var}(y) = \sigma_y^2 \tag{5.133}$$

This means the two observations are unbiased and with different variance. The goal of data fusion is to find a function $Z_{xy} = f(x, y)$ such that it is an unbiased estimate of θ and the variance is no greater than σ_x^2 and σ_y^2. The simplest situation is linear weighting of x and y, i.e.,

$$Z_{xy} = w_x x + w_y y \tag{5.134}$$

to find w_x and w_y such that

$$\mathrm{Var}(Z_{xy}) = E[(Z_{xy} - \theta)^2] \tag{5.135}$$

is minimal. The constraint condition is $w_x + w_y = 1$. This is a natural requirement, since we hope that Z_{xy} is an unbiased estimate for θ.

Based on the independence assumption of x and y, from Eq. (5.135) we have

$$\mathrm{Var}(Z_{xy}) = w_x^2\sigma_x^2 + w_y^2\sigma_y^2 \tag{5.136}$$

By using the Lagrange multiplier method to find the conditional extremal:

$$w_x = \frac{\sigma_y^2}{\sigma_x^2 + \sigma_y^2}$$

$$w_y = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2} \tag{5.137}$$

The meaning of Eq. (5.137) is easy to understand. The optimum weighting coefficients are inversely proportional to the variance. The greater the variance is, the smaller the weighting coefficient. It obtains that

$$Z_{xy} = \frac{1}{\sigma_x^2 + \sigma_y^2}[\sigma_y^2 x + \sigma_x^2 y] \tag{5.138}$$

The straightforward calculation follows

$$\mathrm{Var}[Z_{xy}] = \left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} \right)^{-1} \tag{5.139}$$

It is easy to prove that

$$\mathrm{Var}[Z_{xy}] \le \left(\frac{1}{\sigma_x^2} \right)^{-1} = \sigma_x^2 = \mathrm{Var}[x] \tag{5.140}$$

$$\mathrm{Var}[Z_{xy}] \le \left(\frac{1}{\sigma_y^2} \right)^{-1} = \sigma_y^2 = \mathrm{Var}[y] \tag{5.141}$$

This result has a very important meaning. That is to say that the variance of the optimum linear data fusion of two independent observations is less, or equal to, the variance of the original components. Therefore the application of optimum linear data fusion is reasonable.

The above result for two independent observations can be easily generated in the case of N independent observations. Let us rewrite Eq. (5.138) as

$$Z_{xy} = \left[\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} \right]^{-1} \times \frac{1}{\sigma_x^2} x + \left[\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} \right]^{-1} \times \frac{1}{\sigma_y^2} y \tag{5.142}$$

This equation is the same as Eq. (5.138), but is the basis for generating the case of N independent observations.

If $x_1, \ldots, x_N$ are N independent unbiased observations for parameter θ, then the optimum linear data fusion is

$$Z_N = \left[\sum_{i=1}^{N} \frac{1}{\mathrm{Var}(x_i)} \right]^{-1} \times \sum_{i=1}^{N} \frac{1}{\mathrm{Var}(x_i)} x_i \tag{5.143}$$

The variance of Z_N is generated from Eq. (5.139), i.e.,

$$\mathrm{Var}[Z_N] = \left[\sum_{i=1}^{N} \frac{1}{\mathrm{Var}(x_i)} \right]^{-1} \tag{5.144}$$

It is more convenient to use the vector notation to prove Eq. (5.144). Note that $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ is the observation vector and $\boldsymbol{w} = (w_1, \ldots, w_N)^{\mathrm{T}}$ is the weight coefficient vector. The linear data fusion is

$$y = w^{\mathrm{T}} x \tag{5.145}$$

to find $\boldsymbol{w}$, minimized as

$$I = E\{[w^{\mathrm{T}}(x - E(x))]^2\}$$

subject to the condition $w^{\mathrm{T}}u = 1$ where the vector u represents the vector where all components are equal to 1.

Suppose $R_{xx} = E[(x - E(x))(x - E(x))^{\mathrm{T}}]$ is the auto-correlation matrix of x. By using the Lagrange multiplier method, we can find the minimum value of

$$z = I + \lambda w^{\mathrm{T}}u = w^{\mathrm{T}}R_{\mathrm{xx}}w + \lambda w^{\mathrm{T}}u$$

It is proved that the optimum weight vector is

$$w_{\mathrm{opt}} = \frac{R_{xx}^{-1}u}{u^{\mathrm{T}}R_{xx}^{-1}u} \tag{5.146}$$

and

$$I_{\min} = w_{\mathrm{opt}}^{\mathrm{T}}R_{xx}w_{\mathrm{opt}} = \frac{1}{u^{\mathrm{T}}R_{xx}^{-1}u} \tag{5.147}$$

Since x_i and x_j are independent for $i \neq j$, therefore R_{xx} is the diagonal matrix. It follows that

$$w_{\mathrm{opt}} = \frac{1}{\left(\frac{1}{\sigma_1^2} + \ldots + \frac{1}{\sigma_N^2}\right)} \times \left(\frac{1}{\sigma_1^2}, \ \ldots, \ \frac{1}{\sigma_N^2}\right)^{\mathrm{T}} \tag{5.148}$$

and

$$I_{\min} = \left(\frac{1}{\sigma_1^2} + \ldots + \frac{1}{\sigma_N^2}\right)^{-1} \tag{5.149}$$

These are Eqs. (5.143) and (5.144).

It is easy to prove that

$$I_{\min} \leq \sigma_i^2 \quad (i = 1, \ \ldots, \ N) \tag{5.150}$$

The results discussed above are in the case of independent observations. Now we are going to generate the conclusion in the case of dependent observations. Let us begin our discussion from two dependent observations. Suppose there are two unbiased observations x and y for parameter θ, $E(x) = E(y) = \theta$. The variance in observation is $\mathrm{Var}(x) = \sigma_x^2$, $\mathrm{Var}(y) = \sigma_y^2$. The correlation function of x and y is $R_{xy} = E[(x - \theta)(y - \theta)]$. The optimum linear data fusion we are finding is

$Z_{xy} = w_x x + w_y y$, subject the constraint condition of $w_x + w_y = 1$.

It is easy to calculate that

$$\mathrm{Var}[Z_{xy}] = w_x^2\sigma_x^2 + w_y^2\sigma_y^2 + 2w_x w_y R_{xy}$$

Let

$$u = w_x^2\sigma_x^2 + w_y^2\sigma_y^2 + 2w_x w_y R_{xy} + \lambda(w_x + w_y - 1)$$

It can be proved that if we follow the same procedure as in independent observations, we can get the optimum weighting coefficients

$$w_x = \frac{\sigma_y^2 - R_{xy}}{\sigma_x^2 + \sigma_y^2 - 2R_{xy}}, \quad w_y = \frac{\sigma_x^2 - R_{xy}}{\sigma_x^2 + \sigma_y^2 - 2R_{xy}}$$

Therefore

$$\begin{aligned} Z_{xy} &= \frac{\sigma_y^2 - R_{xy}}{\sigma_x^2 + \sigma_y^2 - 2R_{xy}}\, x + \frac{\sigma_x^2 - R_{xy}}{\sigma_x^2 + \sigma_y^2 - 2R_{xy}}\, y \\ &= \left(\frac{1}{\sigma_x^2 - R_{xy}} + \frac{1}{\sigma_y^2 - R_{xy}}\right)^{-1} \times \frac{x}{\sigma_x^2 - R_{xy}} \\ &\quad + \left(\frac{1}{\sigma_x^2 - R_{xy}} + \frac{1}{\sigma_y^2 - R_{xy}}\right)^{-1} \times \frac{y}{\sigma_y^2 - R_{xy}} \end{aligned} \tag{5.151}$$

$$\mathrm{Var}[Z_{xy}] = \frac{\sigma_x^2\sigma_y^2 - R_{xy}^2}{(\sigma_x^2 - R_{xy})(\sigma_y^2 - R_{xy})} \tag{5.152}$$

It is proved that[53] $\mathrm{Var}[Z_{xy}] \le \mathrm{Var}[x]$, $\mathrm{Var}[Z_{xy}] \le \mathrm{Var}[y]$.

That is to say the variance of optimum linear data fusion of two dependent observations is no more than the variance of the original components.

As an example, we prove $I_{\min} = \mathrm{Var}[Z_{xy}] \le \sigma_x^2$. It is necessary to prove

$$\sigma_x^2 - I_{\min} \ge 0$$

In fact

$$\sigma_x^2 - I_{\min} = \sigma_x^2 - \frac{\sigma_x^2\sigma_y^2 - R_{xy}^2}{\sigma_x^2 + \sigma_y^2 - 2R_{xy}} = \frac{(\sigma_x^2 - R_{xy})^2}{\sigma_x^2 + \sigma_y^2 - 2R_{xy}} \tag{5.153}$$

Based on the Schwartz inequality, we have

$$2R_{xy} = 2E[(x-\theta)(y-\theta)] \le 2\sigma_x\sigma_y \le \sigma_x^2 + \sigma_y^2$$

Therefore the denominator on the right side of Eq. (5.153) is non-negative. It follows that $\sigma_x^2 - I_{\min} \ge 0$.

For the general situation of N dependent observations, the derivation of optimum linear data fusion is much more complicated than in the case of two observations. Since the auto-correlation matrix of the data vector is not diagonal, the inverse of this matrix is quite complicated. But it is proved that the basic theorem discussed before is still valid and[53]

$$\begin{aligned} \boldsymbol{w}_{\text{opt}} &= \frac{\boldsymbol{R}_{xx}^{-1}\boldsymbol{u}}{\boldsymbol{u}^T\boldsymbol{R}_{xx}^{-1}\boldsymbol{u}} \\ I_{\min} &= \boldsymbol{w}_{\text{opt}}^{\mathrm{T}}\boldsymbol{R}_{xx}\boldsymbol{w}_{\text{opt}} = \frac{1}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{R}_{xx}^{-1}\boldsymbol{u}} \end{aligned} \tag{5.154}$$

Fig. 5.62 illustrates the result of optimum linear data fusion of two independent observations. Fig. 5.63 illustrates the result of optimum data fusion of two dependent observations. As we expected, for $R_{xy} = -0.5\sigma_x\sigma_y$, the value of Var[$Z_{xy}$] becomes much smaller. We can conclude that data fusion for dependent observations is sometimes very helpful.

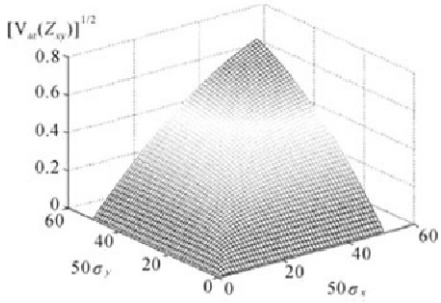


Fig. 5.62. Data fusion: independent observations

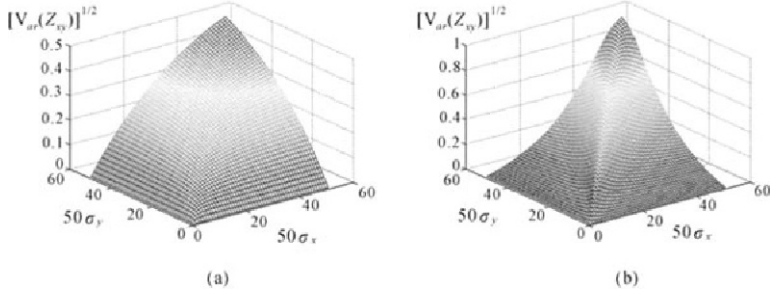


Fig. 5.63. Data fusion: two dependent observations. (a) $P_{xy} = -0.5$; (b) $P_{xy} = -0.95$

Fig. 5.54 illustrates the weight coefficients of optimum data fusion. As shown in Eq. (5.137), the value of the weight coefficient is related to the variance in observation. The magnitude is inversely proportional to the variance, i.e., the observation error.

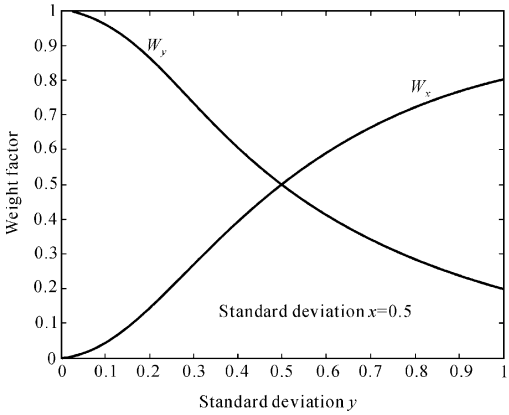


Fig. 5.64. Data fusion: weight coefficients of two independent observations

Example

It is well known that the bearing error of a circular array or cylindrical array is basically uniform, i.e., it does not depend on the incidental angle of the target. But for a linear array, the bearing error strongly depends on the incidental angle of the target. Therefore, if a bow sonar and a towed array sonar are steered at the same target, they usually have different bearing errors. In this case, data fusion should be very useful to balance these errors and results in much better bearing accuracy. Fig. 5.65 is an example of data fusion for a circular array and linear array. We can see from this figure that the combined estimation of the target incidental angle will have approximately the same bearing error.

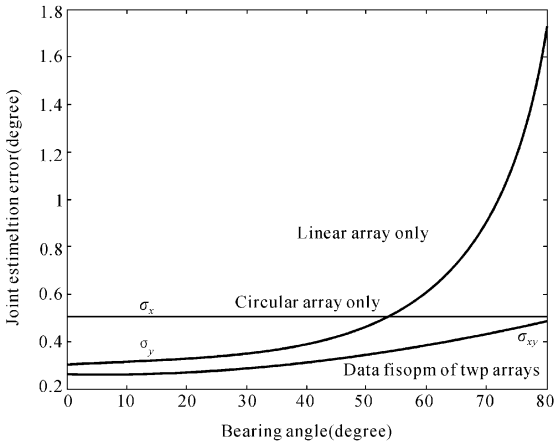


Fig. 5.65. Example of data fusion application

5.9.3 Relationship of Weight Coefficients with SNR

In the above discussion, the observation error is considered as the parameter of data fusion. However, sometimes if we use the SNR it will be more comprehensive. As an example, let us consider the case of two observations. Suppose $Z_{xy} = w_x x + w_y y$ and $E(x) = \theta$, $E(y) = \theta$; $Var(x) = \sigma_x^2$, $Var(y) = \sigma_y^2$. When the signal power is fixed, then σ_x^2 is inversely proportional to the input SNR $(SNR)_{\text{in}}$, i.e.,

$$\sigma_x^2 = \frac{k_0}{(SNR)_{\text{in}}}$$

where k_0 is a constant. The same conclusion is correct for σ_y^2. Therefore, based on Eq. (5.138) we have

$$\begin{aligned} Z_{xy} &= \frac{1}{\sigma_x^2 + \sigma_y^2}[\sigma_y^2 x + \sigma_x^2 y] \\ &= \frac{(SNR)_x}{(SNR)_x + (SNR)_y} x + \frac{(SNR)_y}{(SNR)_x + (SNR)_y} y \end{aligned} \tag{5.155}$$

5.9.4 Joint Detection and Estimation Problems

In discussing sonar design problems, we focus on how to meet the user requirements and how to realize the various functions. For example, when we ask if a target is present, this is a detection problem; when we talk about bearing accuracy, it is a problem of parameter estimation. That is to say, when we discuss the estimation problem, we inherently suppose that the signal is detected, i.e., the estimation procedure should be placed after detection. However, from the statistical point of view, the detection and estimation issues can be simultaneously considered. This is the so-called "joint detection and estimation" problem. Middleton proposed this topic in the late 1960's[84-85]. Recently, some authors expressed their interest again[66, 67], even suggesting that estimation be classified before detection. We will briefly introduce this issue.

Fig. 5.66 presents the block diagram of detection before the estimation model. Based on this model, if the observed data x belongs to the accepted region X_1, then we make the decision "signal present", and the subsequent estimation procedure can be carried out.

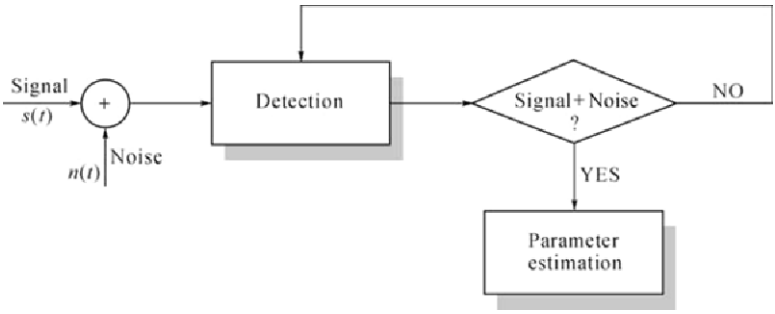


Fig. 5.66. Model of detection before estimation

Of course, any decision is statistical, and error probability exists, including false alarm probability α and missing probability β. Let us review the basic detection model discussed before.

H_0: noise only, $x(t) = n(t)$;

H_1: signal presents, $x(t) = s(t) + n(t)$.

In the sample space: $X = X_0 + X_1$. When $x \in X_0$, make decision D_0: H_0 true; when $x \in X_1$, make decision D_1: H_1 true.

The false alarm probability

$$\alpha = P(D_1 \mid H_0) = \int_{X_1} p_0(x)\mathrm{d}x$$

The missing probability

$$\beta = P(D_0 \mid H_1) = \int_{X_0} p_1(x)\mathrm{d}x$$

The detection probability $p_d = 1 - \beta$.

The joint detection and estimation model is shown in Fig. 5.67. Middleton suggested that it is necessary to use the average risk function $R(p, r)$ to treat the joint detection and estimation problem. Where p is the prior probability density function, r is some kind of rule. The density function p can be expressed as

$$p(s) = P(H_1)p(s \mid H_1) + P(H_0)\delta(s) \tag{5.156}$$

where $\delta(\cdot)$ is the Dirac function. Similarly, for the parameter θ to be estimated, we have

$$p(\theta) = P(H_1)p(\theta \mid H_1) + P(H_0)\delta(\theta) \tag{5.157}$$

From Eq.(5.156) we found that only in the case of $P(H_1) = 1$ did we have

$p(s) = p(s \mid H_1)$. In the other case, $P(H_1) < 1$, therefore $P(H_0) > 0$. The calculation of the risk function depends on $p(x \mid H_0)$ and $p(x \mid H_1)$. Therefore, the important thing is to calculate the false alarm probability and detection probability, regardless of which kind of models is applied.

Let us consider the deterministic signal detection in the Gaussian noise background. The model is as follows:

$$H_0 : n(t), \quad H_1 : n(t) + \mu .$$

$$p_0(x) = p(x \mid H_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2} \right) \tag{5.158}$$

$$p_1(x) = p(x \mid H_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \tag{5.159}$$

The likelihood ratio function

$$\Lambda(x) = \frac{p_1(x)}{p_0(x)} = \exp[-\frac{\mu}{\sigma^2}(\frac{\mu}{2} - x)] \tag{5.160}$$

If $\mu \geq 0$, then

$$X_0 : x < \mu / 2; \quad X_1 : x \geq \mu / 2$$

If $\mu \leq 0$, then

$$X_0 : x > \mu / 2; \quad X_1 : x \leq \mu / 2$$

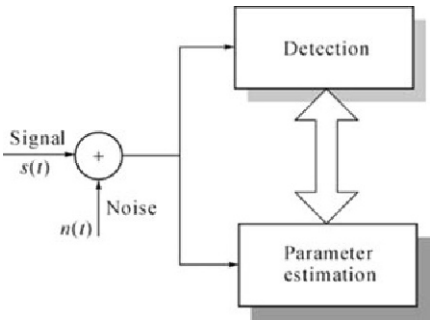


Fig. 5.67. Model of joint detec and estimate

Fig. 5.68 illustrates the figure of $p_0(x)$ and $p_1(x)$, in the case of $\mu \geq 0$ and $\mu \leq 0$.

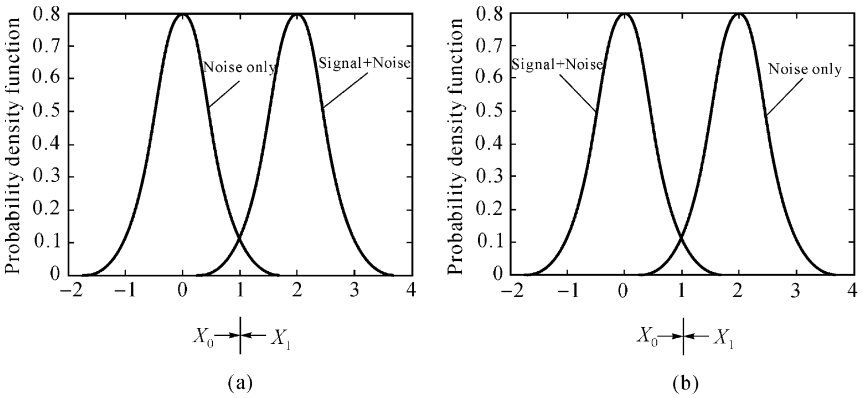


Fig. 5.68. Optimum detection of deterministic signal in the background of white noise. (a) μ>0; (b) μ<0

It is proved that the false alarm probability

$$\alpha = \int_{X_1} p_0(x)\mathrm{d}x = \Phi\left(\frac{|\mu|}{2\sigma}\right) \tag{5.161}$$

The detection probability

$$p_d = \int_{X_1} p_1(x)\mathrm{d}x = 1 - \Phi\left(\frac{|\mu|}{2\sigma}\right) \tag{5.162}$$

where $\Phi(\cdot)$ indicates the probability integration

$$\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)\mathrm{d}t \tag{5.163}$$

The model of Gaussian signal detection in the Gaussian noise background is s follows: $H_0 : n(t)$, $H_1 : s(t) + n(t)$.

$$p_0(x) = p(x \mid H_0) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{x^2}{2\sigma_n^2}\right) \tag{5.164}$$

$$p_1(x) = p(x \mid H_1) = \frac{1}{\sqrt{2\pi}(\sigma_n^2 + \sigma_s^2)^{1/2}} \exp\left[-\frac{x^2}{2(\sigma_n^2 + \sigma_s^2)}\right] \tag{5.165}$$

The likelihood ratio function

$$\Lambda(x) = \frac{p_1(x)}{p_0(x)} = \frac{\sigma_n}{\sqrt{\sigma_n^2 + \sigma_s^2}} \exp\left[\frac{\sigma_s^2 x^2}{2\sigma_n^2(\sigma_s^2 + \sigma_n^2)}\right]$$

and $X_0: \ |x|<a; \quad X_1: \ |x|\geq a$.

where

$$a=\left[\frac{2\ln\sqrt{1+SNR}}{SNR}(\sigma_s^2+\sigma_n^2)\right]^{1/2}, \quad SNR=\sigma_s^2/\sigma_n^2 \tag{5.166}$$

The false alarm probability

$$\alpha=\int_{|x|\geq a} p_0(x)\mathrm{d}x=2\Phi\left(\frac{a}{\sigma_n}\right)=2\Phi\left(\sqrt{\frac{2\ln\sqrt{1+SNR}}{SNR}}\sqrt{1+SNR}\right) \tag{5.167}$$

The detection probability

$$\begin{aligned} p_d &= \int_{|x|\geq a} p_1(x)\mathrm{d}x \\ &= \int_{|x|\geq a}\frac{1}{\sqrt{2\pi(\sigma_n^2+\sigma_s^2)}}\exp\left[-\frac{x^2}{2(\sigma_n^2+\sigma_s^2)}\right]\mathrm{d}x \\ &= 2\Phi\left[\sqrt{\frac{2\ln\sqrt{1+SNR}}{SNR}}\right] \end{aligned} \tag{5.168}$$

Fig. 5.69 shows the optimum detection of the Gaussian signal in the Gaussian noise background.

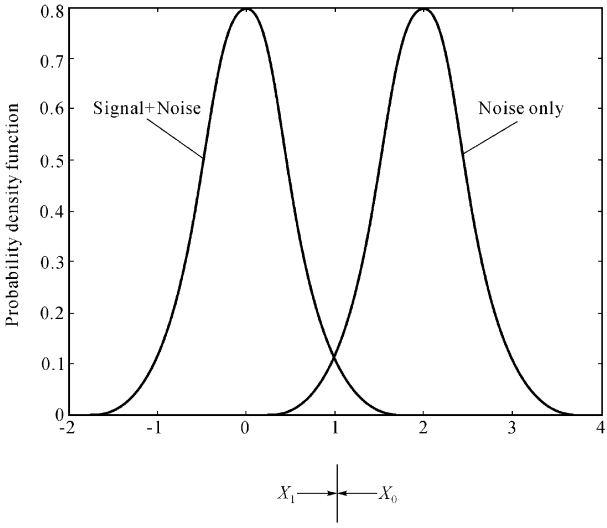


Fig. 5.69. Optimum detection of Gaussian signal in the background of Gaussian noise

We found that the detection performance of the optimum detector depends on the SNR, whether it is a deterministic signal or Gaussian signal. Until recently, the study of joint detection and estimation has not achieved any encouraging results. At least we can confirm that the detection performance of the detection before estimation model is no worse than the joint detection and estimation model.

As for the parameter estimation, the Cramer-Rao lower bound already sets the error bound that we can reach. So we have to compare the estimation error of the joint detection and estimation model to the Cramer-Rao lower bound. There is no closed form of expression for that in terms of the risk function of Eq. (5.156).

References

[1] Wiener, N.: Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications. Wiley, New York (1949)
[2] Knight, W. C., Pridham, R. G., Kay, S. M.: Digital signal processing for sonar. In: Proc. IEEE, 69, pp. 1451-1506 (1981)
[3] Urick, R. J.: Multipath propagation and its effects on sonar design and performance in the real ocean. In: Proc. of NATO ASI on Underwater Acoustics (1976)
[4] Ol'shevskii, V. V.: Statistical Method in Sonar. Consultants Bureau, New York (1978)
[5] Horton, C. W.: Signal Processing of Underwater Acoustic Waves. US Government Printing Office (1969)
[6] Hamblen, W.: Next generation of stealth submarines. Sea Technology, 39-11, pp. 59-62 (1998)
[7] Candy, J. V.: Signal Processing: the Model-Based Approach. McGraw-Hill, New York (1986)
[8] Papoulis, A.: Maximum entropy and spectral estimation : a review. IEEE Trans., ASSP-29., pp. 1176-1186 (1981)
[9] Kassam, S. A., Poor, H. V.: Robust techniques for signal processing: a survey. In: Proc. IEEE, 73, pp. 433-481 (1985)
[10] Johnson, D., Dungeon, D.: Array Signal Processing: Concepts and Techniques. Prentice Hall, New Jersey (1993)
[11] Hart, J. E.: A look at DSP chips. BYTE, pp. 250-255 (1989)
[12] Gupta, A., Toong, H.: Microprocessors : the first twelve years. In: Proc. Of IEEE, 71-11, pp. 1236-1256 (1983)
[13] Anderson, V. C.: Side lobe interference suppression with an adaptive null processor. J. Acoust. Soc. Amer., 69, pp. 185-190 (1981)
[14] Anderson, V. C.: DICANNE, a realizable adaptive process. J. Acoust. Soc. Amer., 45, pp. 398-405 (1969)
[15] Widrow, B., Glover, J., McCool, J. M., et al.: Adaptive noise canceling: principles and applications. In: Proc. IEEE, 63, pp. 1692-1716 (1975)

[16] Li, Q. H., Li, S. Q.: The theoretical consideration of adaptive noise canceling filter in flank array sonar. In: Proc. of UDT'97, Hamburg (1997)
[17] Gallagher, N. C. Jr.: Median filter : a tutorial. In: IEEE Proc. Of Intl. Symp. Circuits and Systems, pp. 1737-1744 (1988)
[18] Struzinski, W. A., Lowe, E. D.: A performance comparison of four noise background normalization schemes proposed for signal detection. J. Acoust. Soc. Amer., 76-6, pp. 1738-1742 (1984)
[19] Lindley, C. A.: Practical Image Processing in C. Wiley, New York (1991)
[20] Wickerhauser, M.: High Resolution Still Picture Compression. Digital Signal Proc., 2-4, pp. 204-226 (1992)
[21] Lim, J. S.: Two-Dimensional Signal and Image Processing. Prentice Hall, New Jersey (1990)
[22] Struzinski, W. A., Lowe, E. D.: The effect of improper normalization on the performance of an automated energy detector. J. Acoust. Soc. Amer., 78-3, pp. 936-941 (1985)
[23] Pratt, W. K.: Digital Image Processing. Wiley, New York (1978)
[24] Hinich, M. J.: Frequency wave-number array processing. J. Acoust. Soc. Amer., 69, pp. 732-737 (1981)
[25] Wang, H. S. C.: Amplitude shading of sonar transducer arrays. J. Acoust. Soc. Amer., 57, pp. 1076-1084 (1975)
[26] Kuo, S. M., Morgem, D. R.: Active Noise Control Systems. Wiley, New York (1996)
[27] Vural, A. M.: Effects of perturbation on the performance of optimum / adaptive arrays. IEEE Trans., AES-15-1, pp. 76-87 (1979)
[28] Farrier, D. R., Durrani, T. S., Nightingale, J. M.: Fast beamforming techniques for circular arrays. J. Acoust. Soc. Amer., 58, pp. 920-922 (1975)
[29] Gray, D. A.: Effect of time delay errors on the beam pattern of a linear array. IEEE J. of Oceanic Engr., 10, pp. 241-247 (1985)
[30] Pridham, P. G., Mucci, R. A.: A novel approach to digital beamforming. J. Acoust. Soc. Amer., 63, pp. 425-434 (1979)
[31] Fisher, B., Bershad, N. J.: ALE behavior for two sinusoidal signal models. IEEE Trans., ASSP-33, pp. 658-665 (1985)
[32] Frost, O. L. III: An algorithm for linearly constrained adaptive array processing. In: Proc. IEEE, 60-8, pp. 926-935 (1972)
[33] Tuteur，F. B., Presley, J. A. Jr.: Spectral estimation of space-time signals with DIMUS array. J. Acoust. Soc. Amer., 70, pp. 74-79 (1981)
[34] Hero, A. O., Schwartz, S.: A new generalized cross correlator. IEEE Trans., ASSP-33, pp. 38-45 (1985)
[35] Messer, H.: The use of spectral information in optimal detection of a source in the presence of a directional interference. IEEE J. of Oceanic Engr., 19-3, pp. 416-421 (1994)
[36] Pridham, R. G., Mucci, R. A.: Digital interpolation beamforming for low-pass and bandpass signals. In: Proc. IEEE, 67-6, pp. 904-919 (1979)
[37] Cantoni, A., Godara, L. C.: Performance of a postbeamformer interference

canceller in the presence of broadband directional signals. J. Acoust. Soc. Amer., 76, pp. 128-138 (1984)
[38] Siewiorek, D. P., Swarz, R. S.: The Theory and Practice of Reliable System Design. Digital Press, Daytona Beach (1982)
[39] Smith, D. J., Harris, J. I.: Line tracking using artificial neural networks and fuzzy inference. In: Proc. of UDT'91, pp. 148-152 (1991)
[40] Bray, A. V.: Underwater hardware life testing. Sea Technology, 33-12, pp. 56-61 (1992)
[41] Pecht, M. G., Nash, F. R.: Predicting the reliability of electronic equipment. In: Proc. of IEEE, 82-7, pp. 992-1004 (1994)
[42] Amstadter, B. L.: Reliability Mathematics. McGraw Hill, New York (1978)
[43] Lapp, S. A.: Derivation of an exact expression for mean time to repair. IEEE Trans. On Reliability, R-35-3, pp. 336-337 (1986)
[44] Lyu, M. R. (Eds): Handbook of Software Reliability Engineering. McGraw Hill, New York (1996)
[45] Marra, L. J.: Sharkbite of the submarine lightwave cable system: history, causes and resolution. IEEE J. of Oceanic Eng., 14-3, pp. 230-237 (1989)
[46] McCluskey, E. J.: Built-in self test technologies. IEEE Design and test of Computers, 2, pp. 21-28 (1985)
[47] Ralston, A.: Encyclopedia of computer science engineering, 2nd Edn. Van Nostrand Reinhold, New York (1985)
[48] Varshney, P. K.: Distributed Detection and Data Fusion. Springer, New York (1996)
[49] Hall, D. L., Llinas, J.: An introduction to multisensor data fusion. In: Proc. IEEE, 85-1, pp. 6-23 (1997)
[50] Viswannathan, R., Varshney, P. K.: Distributed detection with multiple sensors: Part I – fundamentals. In: Proc. IEEE, 85-1, pp. 54-63 (1997)
[51] Tenny, P. R., Sandel, N. R.: Detection with distributed sensors. IEEE Trans. AES-17-4, pp. 501-510 (1981)
[52] Li, Q. H.: The optimum linear data fusion of independent observation data. Chinese J. of Acoustics, 20-1, pp. 18-24 (2001)
[53] Li, Q. H.: The optimum linear data fusion of dependent observation data. Chinese J. of Acoustics, 20-2, pp. 97-102 (2001)
[54] Hamson, R. M.: The theoretical gain limitation of a passive vertical line array in shallow water. J. Acoust. Soc. Amer., 68-1, pp. 156-164 (1980)
[55] Yang, T. C.: A method of range and depth estimation by modal decomposition. J. Acoust. Soc. Amer., 82-5, pp. 1736-1745 (1987)
[56] Hawkes, M., Nehorai, A.: Acoustic vector sensor correlations in ambient noise. IEEE J. of Oceanic Eng., 26-3, pp. 337-347 (2001)
[57] Nehorai, A., Paldi, E.: Acoustic vector-sensor array processing. IEEE Trans. Signal Processing, 42, pp. 2481-2491 (1994)
[58] Horton, C. W.: Signal Processing of Underwater Acoustic Waves. Supt of Docs., U.S. Govt. Print. Off. Washington (1969)
[59] Jensen, F. B.: Computational Ocean Acoustics. American Institute of Physics, New York (1993)

[60] Wilson, O. B.: Introduction to The Theory and Design of Sonar Transducers. Peninsula Publishing, Newport Beach (1985)
[61] Beranek, L. L., Ver, I. L.: Noise and Vibration Control Engineering: Principles and Applications. Wiley, New York (1992)
[62] Jonson, A. J. L.: Address at US naval institute Annapolis seminar and 123rd annual meeting. Annapolis (1997)
[63] Moffat, J.: Complexity Theory and Network Centric Warfare. CCRP, Washington (2003)
[64] Albert, D. S., Garstka, J. Stein, F. P.: Network Centric Warfare: Developing and Leveraging Information Superiority, 2nd Edn. CCRP, Washington (2002)
[65] Hosking, R. H.: Choosing the right DSP for real time embedded systems. Electronic Design, 48-24, pp. 137-146 (2000)
[66] Shin, F. B., Kil, D. H.: Full spectrum signal processing using a classify-before-detection paradigm. J. Acoust. Soc. Amer., 99-4 (1996)
[67] Olmo, G., Magli, E., Presti, L. L.: Joint statistical signal detection and estimation Part I: theoretical aspects of the problem. Signal Processing, 80, pp. 57-73 (2000)
[68] Walrod, J.: Sensor networks for network centric warfare. In: Proc. of NCW Conference, Fall Church (2000)
[69] Bonito, G.: Sensor Array Projects and Networks & Other Useful Links, www.lternet.edu (2003)
[70] National Research Council of USA (Eds): Network Centric Naval Forces: a Transition Strategy for Enhancing Operational capabilities. NA Press, Washington (2000)
[71] Trider, R. C.: A fast Fourier transform (FFT) based on sonar signal processor. IEEE Trans., ASSP-26, pp. 15-20 (1978)
[72] Rudnick, P.: Digital beamforming in the frequency domain. J. Acoust. Soc. of Amer., 46, pp. 1089-1090 (1969)
[73] Bartram,J. F., Ramseyer, R., Heines, J.: Fifth generation digital sonar processing. IEEE trans., OE-2, pp. 337-343 (1977)
[74] Feeley, M.: ARCI Heralds Acoustic Revolution. UDT Forum, pp. 10-14, Lockheed Martin Comp., Bethesda, http://www.udt-europe.com (2005)
[75] Hanna, M. T., Simaan, M: Absolutely optimum array filters for sensor arrays. IEEE Trans. ASSP-33, pp. 1380-1386 (1985)
[76] Li, Q. H: The performance of the optimum array filter for sensor arrays. In: Proceedings of ICASSP 87, pp. 2324-2327, Dallas (1987)
[77] Li, Q. H.: A Combined Architecture of Adaptive Line Enhancer and Adaptive Noise Canceller. Technical Report, Princeton Univ. (1985)
[78] Widrow, B., Steams, S. D.: Adaptive Signal Processing. Prentice Hall, New Jersey (1985)
[79] Rickard, J. T., Zeidler, J. R.: Second order output statistics of adaptive line enhancer. IEEE Trans., ASSP-27, pp. 31-39 (1979)
[80] Rasmussen, R. A.: Studies related to the design and use of time / bearing sonar display. AD report 690540 (1968)

[81] Danid, E., Neri, D. F., Kobus, D. A., et al.: Effect of background and foreground color coding on detection: acoustic data analysis. AD-A report 154105 (1985)
[82] Butler, W. B.: Engineering guidelines for the use of color on the sonar display. AD report 782597 (1974)
[83] Sittler, R. W.: An optimal data association problem in surveillence theory. IEEE Trans. Military Electronics, 8-2, pp. 125-139 (1964)
[84] Middleton, D., Esposito, R.: Simultaneous optimum detection and estimation of signal in noise. IEEE Trans., IT-14-3, pp. 434-444 (1968)
[85] Jaffer, A. and Gupta, S. C.: Coupled detection-estimation of Gaussian processes in Gaussian noise. IEEE Trans., IT-18-1, pp. 106-110 (1972)

6

Implementation Methods of Various Functions of Digital Sonar

The combination of digital signal processing theory and techniques with underwater acoustics and oceanography establishes the foundation of digital sonar design[1-6], and also provides the possibility to complete the tactical and technical specifications. But this is just one problem. In order to implement various functions of a digital sonar system, it is necessary to carefully manage the hardware and software, choose the right DSP chips and the algorithms, which can perfectly run the necessary compiler. The philosophy of a sonar designer should be: based on the user's requirements, find the optimum, but the simplest design. It is necessary to use the simplest, most reliable means to implement our design. The system that we designed must have a friendly interface, extendable space[7-13].

In this chapter, we will introduce a method for performing various functions of digital sonar, including target bearing / tracking, passive ranging, target classification / identification, active sonar signal capture and underwater communication[14-30].

6.1 Precise Bearing for Target

If the sonar system confirms a contact showing that a target is present, it is necessary to show the exact incidental angle of the target, since the multi-beam system has discrete beams, which are equally distributed through 360°. Therefore it is necessary to have a specific algorithm to show the precise value of the incidental angle.

The time difference of the received signal in the hydrophone is the basic data for estimating the incidental angle of the target. Therefore, the time delay estimation of the direction of arrival (DOA) is the most important issue in achieving a precise bearing. The problem of time delay estimation has attracted much research since the 1960's.

A split beam algorithm is also an effective method for estimating time delay.

The pioneer study of Carter and Piersol is extensively used in array signal processing[19-21, 24-28].

6.1.1 Concept of the Bearing Accuracy of Sonar

Bearing accuracy and multiple target resolution problems are the important issues in modern sonar. But in the earlier literature this topic is seldom mentioned. The most serious issue in traditional sonar is the directivity function. One believes that the main lobe behavior of the directivity function is related to the bearing accuracy and multiple target resolution.

Actually, the bearing accuracy is tightly related to the time delay problem. The Cramer-Rao lower bound already shows the optimum error bound of time delay estimation. Therefore the bearing accuracy can be given by the time delay estimation based on the array shape.

Fig. 6.1 shows the directivity function of a single target and two equal strength targets. We will prove that the bearing accuracy of the target is determined by the main lobe behavior of the directivity function. From Fig. 6.1(a) we can image that if the width of the main lobe is narrow, then the bearing accuracy should be better. For multiple targets resolution, with a narrow main lobe it is obviously easy to distinguish two targets (see Fig. 6.1(b)). We will show that the bearing accuracy of a hydrophone array can actually be confirmed in the case of two receivers, since the final calculation of the exact value of the incidental angle of the target is defined by the two virtual receivers, derived from the array system. Fig. 6.2(a) shows the case of two receiving hydrophones. Suppose the angle between the incidental direction and the normal of the two receivers is θ. If the signal from the left hydrophone is $s(t)$, then the signal received in the right hydrophone is $s(t-\tau(\theta))$, where $\tau(\theta) = d\sin\theta / c$ and c is the sound speed.

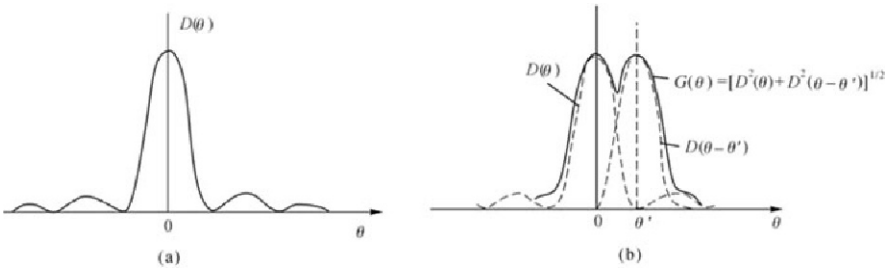


Fig. 6.1. Beam pattern of one target and two equal-strength targets. (a) Beam pattern of one target; (b) Beam pattern of two equal-strength targets

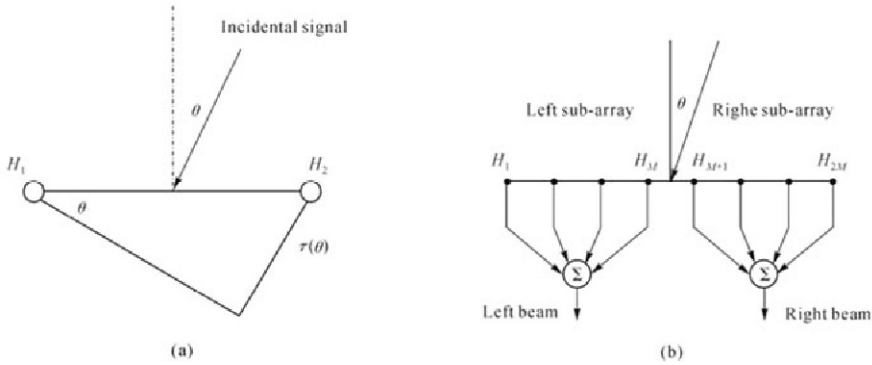


Fig. 6.2. Calculation of time delay and target bearing. (a) Calculation of time delay; (b) Split beam bearing

If the receiver array consists of $2M$ elements, then the split beam system is similar to the case of two elements (Fig. 6.2(b)). The left sub-array is H_1, ..., H_M, and the right sub-array is H_{M+1}, ..., H_{2M}. The left beam and right beam are indicated by $x_l(t)$ and $x_r(t)$, respectively. The time delay between two beams can also be calculated as follows.

As an example, suppose the incidental signal is sinusoidal with frequency f. The left beam signal is

$$\begin{aligned} x_l(t) &= \sum_{i=1}^{M} A_i \cos[2\pi f(t-\tau_i(\theta))] \\ &= \left[\sum_{i=1}^{M} A_i \cos(2\pi f \tau_i(\theta))\right]\cos(2\pi ft) + \left[\sum_{i=1}^{M} A_i \sin(2\pi f \tau_i(\theta))\right]\sin(2\pi ft) \\ &= A\cos(2\pi f(t-\tau(\theta))) \end{aligned} \tag{6.1}$$

where

$$A = \left[\left(\sum_{i=1}^{M} A_i \cos(2\pi f \tau_i(\theta))\right)^2 + \left(\sum_{i=1}^{M} A_i \sin(2\pi f \tau_i(\theta))\right)^2\right]^{1/2} \tag{6.2}$$

$$\tau(\theta) = \frac{1}{2\pi f} \times \arctan\left[\frac{\sum_{i=1}^{M} A_i \sin(2\pi f \tau_i(\theta))}{\sum_{i=1}^{M} A_i \cos(2\pi f \tau_i(\theta))}\right] \tag{6.3}$$

Similarly, we can obtain the expression of the right beam $x_r(t)$. If the

amplitude of the left beam and the right beam are the same, then we may write

$$\begin{aligned} x_l(t) &= s(t) + n_l(t) \\ x_r(t) &= s(t - \tau(\theta)) + n_r(t) \end{aligned} \tag{6.4}$$

where $n_l(t)$ and $n_r(t)$ represent the background noise, which is independent of the signal. It is worth showing that the value of the time delay between $n_l(t)$ and $n_r(t)$ is related to the frequency. Therefore, for a broadband signal, it is necessary to synthesize the estimated results of the time delay for different frequency bins. We will show the method later.

There are many methods that can be used to estimate $\tau(\theta)$. For example, the maximum value method, sum-difference beam method, etc. Whichever method is used the estimation accuracy is limited by the Cramer-Rao lower bound.

Now we are going to introduce an algorithm which is based on the cross power spectrum calculation, for time delay estimation.

The cross power spectrum of $n_l(t)$ and $n_r(t)$ is

$$K_{lr}(f) = K_s(f)\exp(2\pi f\tau) \tag{6.5}$$

where $K_s(f)$ represents the power spectrum of signal $s(t)$. From Eq. (6.5) we know that the information of the time delay is contained in $K_{lr}(f)$ since $K_s(f)$ is real, so that

$$2\pi f\tau = \arctan\left[\frac{\mathrm{Im}(K_{lr}(f))}{\mathrm{Re}(K_{lr}(f))}\right] \tag{6.6}$$

It follows that

$$\hat{\tau} = \frac{1}{2\pi f} \times \arctan\left[\frac{\mathrm{Im}(K_{lr}(f))}{\mathrm{Re}(K_{lr}(f))}\right] \tag{6.7}$$

Knapp shows that the Cramer-Rao lower bound of this method is [117]

$$\min[\mathrm{Var}(\hat{\tau})] = \left\{2T\int_{-\infty}^{\infty}(2\pi f)^2\frac{|u(f)|^2}{1-|u(f)|^2}\mathrm{d}f\right\}^{-1} \tag{6.8}$$

where T is the observation time. $u(f)$ is the coherence spectrum between $n_l(t)$ and $n_r(t)$, which is defined as

$$|u(f)|^2 = \frac{K_s^2(f)}{[K_s(f)+K_n(f)]^2}$$

In the case of a small SNR

$$|u(f)|^2 \approx \frac{K_s^2(f)}{K_n^2(f)} \tag{6.9}$$

Therefore

$$\min[\mathrm{Var}(\hat{\tau})] \approx \left\{ 2T \int_{-\infty}^{\infty} (2\pi f)^2 \frac{K_s^2(f)}{K_n^2(f)} \mathrm{d}f \right\}^{-1} \tag{6.10}$$

In the simplest situation, if the spectrum of $K_s(f)$ and $K_n(f)$ has a rectangular shape, i.e.,

$$K_s(f) = S_0, \text{ and } K_n(f) = N_0, \ f \in (f_1, f_2)$$

then we have

$$\min[\mathrm{Var}(\hat{\tau})] \approx \frac{3}{8\pi^2 T} \times \frac{1}{(S_0/N_0)^2} \times \frac{1}{f_2^3 - f_1^3}$$

Denoting $[\min(\mathrm{Var}(\hat{\tau}))]^{1/2}$ by $\Delta\tau_{\min}$ obtains

$$\Delta\tau_{\min} \approx \left(\frac{3}{8\pi^2 T} \right)^{1/2} \times \frac{1}{SNR} \times \frac{1}{\sqrt{f_2^3 - f_1^3}} \tag{6.10}$$

In order to calculate the bearing accuracy, it is necessary to transfer the time delay τ to the incidental angle θ. Of course, it depends on the array shape.

6.1.2 Optimum Precise Bearing of Line Array and Circle Array

We will show how to calculate the bearing angle based on the time delay value. Let us first consider an equal spaced line array with $2M$ elements. The i-th received signal is expressed as

$$s_i(t) = \cos[2\pi f(t + \tau_i(\theta))]$$

where

$$\tau_i(\theta) = (i-1)d \sin\theta / c,$$

and d is the distance of adjacent elements. The left beam signal is

$$x_l(t) = \left(\sum_{i=1}^{M} u_i\right)\cos(2\pi ft) - \left(\sum_{i=1}^{M} v_i\right)\sin(2\pi ft)$$

where $u_i = \cos(2\pi f \tau_i(\theta))$, $v_i = \sin(2\pi f \tau_i(\theta))$.

It is easy to rewrite $x_l(t)$ in the sinusoidal form

$$x_l(t) = A\cos(2\pi ft + \alpha)$$

where $A = \left[\left(\sum_{i=1}^{M} u_i\right)^2 + \left(\sum_{i=1}^{M} v_i\right)^2\right]^{1/2}$, $\alpha = \arctan\left[\left(\sum_{i=1}^{M} v_i\right) / \left(\sum_{i=1}^{M} u_i\right)\right]$.

Similarly, we can obtain the right beam signal

$$x_r(t) = B\cos(2\pi ft + \beta)$$

where $B = \left[\left(\sum_{i=M+1}^{2M} u_i\right)^2 + \left(\sum_{i=M+1}^{2M} v_i\right)^2\right]^{1/2}$, $\beta = \arctan\left[\left(\sum_{i=M+1}^{2M} v_i\right) / \left(\sum_{i=M+1}^{2M} u_i\right)\right]$.

Let $\psi = 2\pi d \sin\theta / \lambda$, we obtain

$$\sum_{i=1}^{M} u_i = \sum_{i=1}^{M} \cos(i-1)\psi, \quad \sum_{i=1}^{M} v_i = \sum_{i=1}^{M} \sin(i-1)\psi$$

therefore

$$\alpha = \arctan[\sin(M\psi/2) / \cos(M\psi/2)] = M\pi d \sin\theta / \lambda \tag{6.11}$$

Similarly, we can get the phase difference in the right beam

$$\beta = 3M\pi d \sin\theta / \lambda \tag{6.12}$$

Therefore the phase difference of the left beam and the right beam for the line array is

$$\phi = \beta - \alpha = 2M\pi d \sin\theta / \lambda \tag{6.13}$$

and the corresponding time delay is

$$\tau = \phi / (2\pi f) = Md \sin\theta / c \tag{6.14}$$

From the above discussion we know that once we obtain the estimation value of the time delay, it is not difficult to resolve the incidental angle of the target. In the case of a continuous line array, the result is easy to find from Eq. (6.14). Suppose $L = 2Md$, we obtain

$$\sin\theta = c\tau / (Md) = 2c\tau / L \tag{6.15}$$

since $c\tau / L << 1$, so usually we have

$$\theta \approx 2c\tau / L \tag{6.16}$$

In Eq. (6.10), indicating $f_{\text{rms}} = \sqrt{f_1 f_2}$, $W = f_2 - f_1$ and we have

$$\Delta\tau_{\min} = \frac{1}{2\pi} \times \frac{1}{\sqrt{2TW}} f_{\text{rms}}^{-1} \times (SNR)^{-1}$$

Substituting it into Eq. (6.16), we obtain the bearing accuracy

$$\theta_{\text{opt}} = \frac{2\Delta\tau_{\min} c}{L} = \frac{c}{L\pi} \times \frac{1}{\sqrt{2TW}} \times \frac{1}{f_{\text{rms}}} \times \frac{1}{SNR} \tag{6.17}$$

Example

The 3 dB beam width of the line array is (Eq. (4.104))

$$\Delta\theta_{3\text{dB}} = 0.88 \times \frac{c}{Lf_{\text{rms}}}$$

If $TW = 500$, $SNR = 1$, it follows from Eq. (6.17)

$$\theta_{\text{opt}} \approx \Delta\theta_{3\text{dB}} / 80 \tag{6.18}$$

In this example, the bearing accuracy of the line array is about 1/80 of the 3 dB beam width.

Now we turn to discuss the arc array. Suppose the total elements of the circle array are *N*, and the working sector consists of $2M = N / 2$ elements. The radius of the circle array is *r*. Assume $\alpha_0 = 2\pi / N$. It is straightforward to calculate the phase difference for the left beam and right beam of the arc array

$$\phi = \beta - \alpha = \frac{4rf\theta N}{cM} \times \sin^2(M\alpha_0 / 2) \approx 8rf\theta / c \tag{6.19}$$

and

$$\tau = \phi / (2\pi f) = 4r\theta / (c\pi) \tag{6.20}$$

therefore

$$\theta = \tau\pi c / (4r) \tag{6.21}$$

We have

$$\theta_{\text{opt}} = \frac{\pi \Delta \tau_{\min} c}{4r} = \frac{c}{8r} \times \frac{1}{\sqrt{2TW}} \times \frac{1}{f_{\text{rms}}} \times \frac{1}{SNR} \tag{6.22}$$

Example

For the circle array based on Eq. (4.114), we have

$$\Delta\theta_{3\text{ dB}} \approx 1.12 \times \frac{c}{\pi r f_{\text{rms}}}$$

If $SNR = 1$, $TW = 500$, we obtain

$$\theta_{\text{opt}} \approx \Delta\theta_{3\text{ dB}} / 80 \tag{6.23}$$

It is interesting that Eq. (6.23) has the same form as Eq. (6.18). Fig. 6.3 illustrates the relationship between the accuracy of time delay estimation and bearing accuracy.

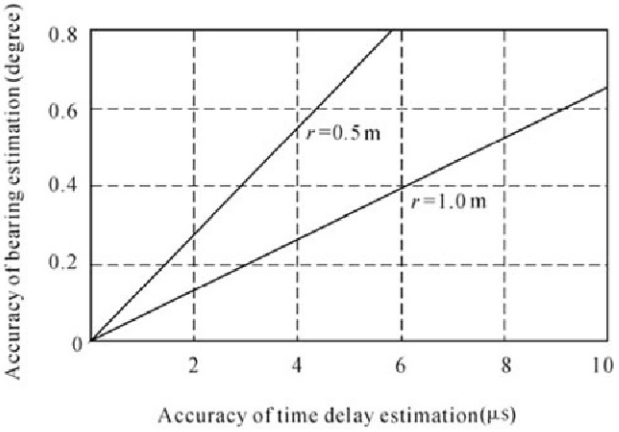


Fig. 6.3. Bearing accuracy of circle array (N=64)

6.1.3 Implementation of Precise Bearing

The accuracy of a bearing is found in the previous section. We are going to show how to realize the sonar bearing. In Fig. 6.4, the different target tracking functions are illustrated. The tracking form can be roughly categorized as automatic, semi-automatic and manual, etc. In addition to these methods, there is also a method called pre-setting tracking, which is a tracking method artificially interposed.

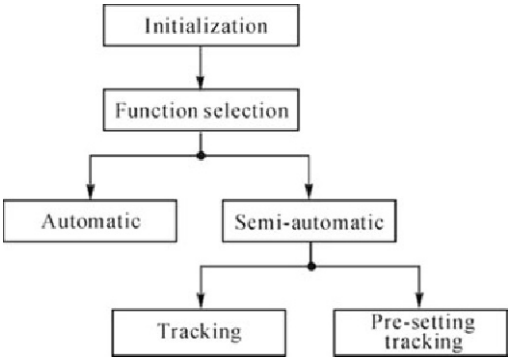


Fig. 6.4. Different tracking methods in passive sonar

There are two methods that can be used to realize a precise bearing: super fine beamforming, and multibeam interpolation (see Fig. 6.5). Let us review the procedure for beamforming. Suppose the true value of the target incidental angle is θ. After calculating the time delay $\tau_i(\theta)$, it is necessary to compensate this value according to half of the sampling period $T_s/2$. Therefore, we obtain a set of time delay compensation $\{INT[\tau_i(\theta)/(T_s/2)]\}$. In terms of this set, the sonar forms the beam, steered in direction θ. It is easy to understand that if two directions θ and θ' are very close, then the set of time delay compensation may be the same. In this case, two beams will have no differences. We conclude that, even for the single beam system, the beam distribution is still discrete.

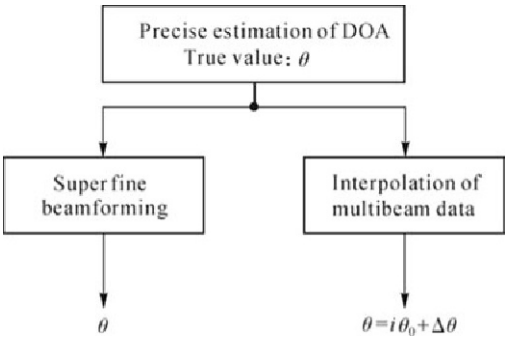


Fig. 6.5. Two methods of precise bearing in digital sonar

For the multi-beam system, suppose the angle between adjacent beams is θ_0. In order to steer the target incidental angle θ, it is necessary to find a beam number, for example $i\theta_0$, which is closest to θ. And then we can adopt the beam interpolation algorithm to find the true value of θ

$$\hat{\theta} = i\theta_0 + \Delta\theta$$

Generally speaking, before calculating the exact bearing angle, beam interpolation is necessary. The aim of beam interpolation is to increase the density of the beams, so we can more precisely estimate the DOA.

There are several interpolation methods, which can refine the beam structure in space. Here we introduce a simple algorithm of interpolation, which focuses on the main lobe of the directivity function. Therefore it is helpful to increase the accuracy of target bearing. Suppose the amplitude response of a sonar system is $D(i\theta_0)$ $(i = 1, \ ..., \ N_0)$, where N_0 is the beam number and $\theta_0 = 2\pi / N_0$. The question is as follows: how can we get $\hat{\theta}$ from $D(i\theta_0)$ $(i = 1, \ ..., \ N_0)$?

We will show that in the neighborhood of the main lobe the directivity function has a parabolic shape. In fact, for a line array with length L, the directivity function has the form of

$$R(\phi) = \frac{\sin\phi}{\phi} \tag{6.24}$$

where

$$\phi = L\pi(\sin\theta - \sin\theta_0) / \lambda$$

where θ is the incidental angle of the target and θ_0 is the bearing angle. In the neighborhood of the main lobe we have $\theta \approx \theta_0$. Based on Taylor's expansion of function $\sin x / x$ in the neighborhood of the zero point, we obtain

$$\sin x / x \approx 1 - x^2 / 3 \tag{6.25}$$

This means that the directivity function of the line array has a parabolic shape in the neighborhood of the main lobe. For the circle array, the conclusion is similar as we know the directivity function of the circle array is a zero order Bessel function

$$D(\theta) = J_0\left[\frac{4\pi r}{\lambda}\sin((\theta - \theta_0)/2)\right]$$

The behavior of $J_0(x)$ at $x = 0$ is

$$J_0(x) = 1 - x^2 / 4 \tag{6.26}$$

It is also has a parabolic form.

In terms of Eqs. (6.25) and (6.26), we know that the directivity functions of the line array and circle array have a parabolic shape in the neighborhood of $\theta \approx \theta_0$.

This is a very important fact for beam interpolation. We can think that the discrete beam response is actually the sample of this parabolic function; therefore it is possible to find the peak value of this parabolic curve in terms of discrete data. And it also provides a theoretical basis for beam refining, which allows us to double the beam number, if necessary.

Fig. 6.6 is a principal description of parabolic interpolation. Fig. 6.6(a) is the back diagram of beam fining and Fig. 6.6(b) is the parabolic interpolation technique used for precise target bearing.

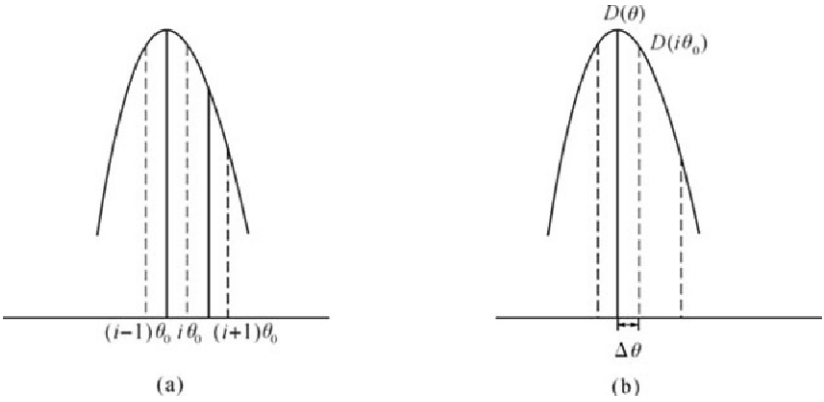


Fig. 6.6. Beam interpolation. (a) Interpolation of parabolic curve; (b) Precise bearing: beam interpolation

In Fig. 6.6(a), there are three discrete beam values $D((i-1)\theta_0)$, $D(i\theta_0)$, $D((i+1)\theta_0)$. The goal of interpolation is to double the beam number, i.e., to find the value of $D((i-0.5)\theta_0)$ and $D((i+0.5)\theta_0)$. For simplicity, suppose

$$y = f(x) = ax^2 + bx + c$$

In terms of the assumption, it is known that $y_i = f(x_i)$ $(i = -1,\ 0,\ 1)$. We want to find the unknown value of $f(x_{-1/2})$ and $f(x_{1/2})$ where

$$x_{-1/2} = \frac{x_{-1} + x_0}{2}, \quad x_{1/2} = \frac{x_0 + x_1}{2}$$

It is straightforward to calculate the parabolic interpolation results

$$f(x_{-1/2}) = \frac{3y_{-1} - y_1 + 6y_0}{8}, \quad f(x_{1/2}) = \frac{3y_1 - y_{-1} + 6y_0}{8} \tag{6.27}$$

Eq. (6.27) is called three-point parabolic interpolations. By using this expression,

it is easy to double the beam number if $x_{-1} = (i-1)\theta_0$, $x_0 = i\theta_0$, $x_1 = (i+1)\theta_0$; $y_{-1} = D((i-1)\theta_0)$.

As we can see, based on the parabolic interpolation algorithm, the beam number is doubled. There is a similar algorithm, which can interpolate three new beam values between two beams, i.e., between $[x_{-1}, x_0]$ interpolate $f(x_{-3/4})$, $f(x_{-1/2})$ and $f(x_{-1/4})$; between $[x_0, x_1]$ interpolate $f(x_{1/4})$, $f(x_{1/2})$ and $f(x_{3/4})$. It is proved that

$$f(x_{-3/4}) = \frac{21y_{-1} + 14y_0 - 3y_1}{32}, \qquad f(x_{-1/2}) = \frac{12y_{-1} + 24y_0 - 6y_1}{32};$$

$$f(x_{-1/4}) = \frac{5y_{-1} + 30y_0 - 3y_1}{32}, \qquad f(x_{1/4}) = \frac{-3y_{-1} + 30y_0 + 5y_1}{32};$$

$$f(x_{1/2}) = \frac{-4y_{-1} + 24y_0 + 12y_1}{32}, \qquad f(x_{3/4}) = \frac{-3y_{-1} + 14y_0 + 21y_1}{32}$$

Example

This is the result of system simulation. The beam duration is 4°, the basic data to be used in the interpolation are $D(0°)$, $D(4°)$, and $D(8°)$. If the DOA of the target is $\theta = 0°$, this coincides with the maximum value $D(0°)$ of the directivity function. But when the DOA of the target is $\theta = 3°$, the maximum value does not appear in $D(0°)$, $D(4°)$, and $D(8°)$. But, after interpolating, $D(3°) = 1.00$ becomes the new maximum value and steers in DOA. Therefore the effect of interpolation is obvious. The calculation results are listed in Table 6.1.

Table 6.1 Simulation results of five-point parabolic interpolation ($d / \lambda = 0.2$)

Target DOA $\theta = 0°$			Target DOA $\theta = 0°$		
θ(°)	D (θ)	Interpolation	θ(°)	D (θ)	Interpolation
0	1.000	–	0	0.974	–
1	0.997	0.997	1	0.988	0.988
2	0.998	0.988	2	0.997	0.997
3	0.974	0.974	3	1.000	1.000
4	0.955	–	4	0.997	–
5	0.930	0.931	5	0.988	0.989
6	0.900	0.901	6	0.974	0.975
7	0.866	0.867	7	0.955	0.956
8	0.827	–	8	0.931	–

Since the multi-beam system is a set of discrete responses to the environment, even the beams are fine enough, therefore it is necessary to have a special algorithm to show the exact target DOA. As we can see in Fig. 6.6(b), the interpolation algorithm is a good choice for a precise bearing. But this time we will interpolate the bearing angle instead of the beam value.

Suppose $i\theta_0$ is the nearest beam to θ. One can select tow beams to the left and right of $i\theta_0$. A virtual parabolic is simulated by $D((i–1)\theta_0)$, $D(i\theta_0)$, and $D((i+1)\theta_0)$. The maximum value $D(\hat{\theta})$ of this parabolic is just where the target DOA is located, i.e.

$$\hat{\theta} = i\theta_0 + \Delta\theta \tag{6.28}$$

The value of $\Delta\theta$ can be calculated as follows: Suppose $y_0 = D(i\theta_0)$, $y_{-1} = D((i-1)\theta_0)$, $y_1 = D((i+1)\theta_0)$ and y_0, y_{-1}, y_1 are located in the parabolic function $y = ax^2 + bx + c$. The maximum value of the parabolic should be $D(\hat{\theta}) = D(i\theta_0 + \Delta\theta)$. It is proved that

$$\Delta\theta = \frac{y_1 - y_{-1}}{4y_0 - 2y_1 - 2y_{-1}} \times \theta_0 \tag{6.29}$$

From Eqs. (6.28) and (6.29) we have

$$\hat{\theta} = i\theta_0 + \frac{y_1 - y_{-1}}{4y_0 - 2y_1 - 2y_{-1}} \times \theta_0 \tag{6.30}$$

From Eq. (6.29) we can see that $\Delta\theta \geq 0$ if $y_1 \geq y_{-1}$; otherwise $\Delta\theta \leq 0$.

6.1.4 *Example of Realization of Precise Bearing Function in Digital Sonar*

We will present a practical example for realizing the precise bearing function in digital sonar. Suppose the sampling frequency of digital sonar is f_s = 10 kHz, the sampling duration T_s = 100 μs. There are 128 beams equally distributed over 0° – 360°. Fig. 6.7 illustrates the post processing of this sonar. The first and second accumulations are carried out for 128 beams, and then double the beam number. Therefore we have a total of 256 beams. The data of 256 beams are used for target tracking and precise bearing. The principle block diagram of the first accumulation and second accumulation is shown in Fig. 6.8.

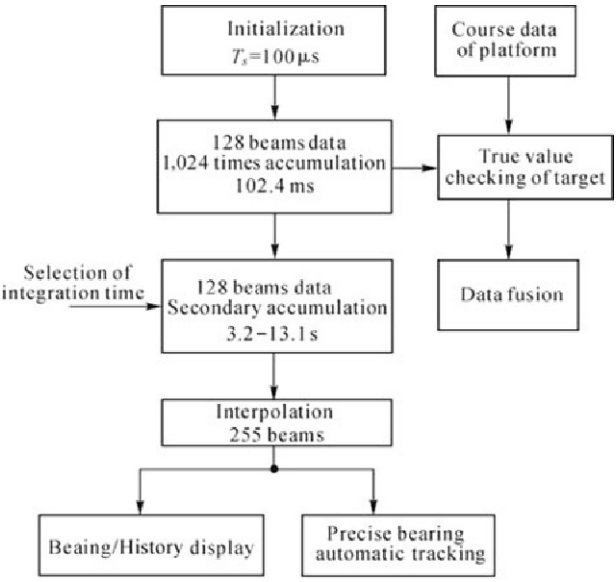


Fig. 6.7. An example of post processing of multibeam data

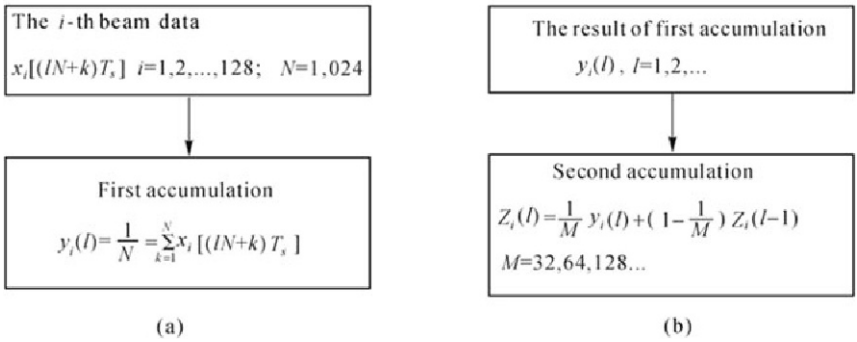


Fig. 6.8. An example of the first and second accumulations. (a) First accumulation; (b) Second accumulation

The time sample of the i-th accumulation number is $x_i[(lN+k)T_s]$ (i = 1, 2, …, 128; N = 1,024). It corresponds to the integration time of 1,024×100μs = 102.4 ms.

l is the series number of the segment. As we can see, the data is not overlaid in the accumulation procedure. The accumulation result of the i-th beam is

$$y_i(l) = \frac{1}{N}\sum_{k=1}^{N} x_i[(lN+k)T_s] \tag{6.31}$$

Obviously, the output rate of $y_i(l)$ decreases to 1 / 1,024 of the original data $x_i(kT_s)$.

The second accumulation is a moving average, the integration time is determined by M. Usually we choose $M = 32,64,128,\ldots$

$$z_i(l) = \frac{1}{M} y_i(l) + \left(1 - \frac{1}{M}\right) z_i(l-1) \tag{6.32}$$

The second accumulation further decreases the output data rate to 1 / M. If $M = 32$, then the output data rate is 3.2 s for one sample. Denote 128 samples by $z(1)$, …, $z(128)$. In order to double the beam numbers, we can use parabolic interpolation, i.e., in the neighborhood of the main lobe we use three-point parabolic interpolation, and otherwise we use a simple average algorithm, see Fig. 6.9. Suppose the maximum value of $z(l)$ is at i_0: $A_0 = \max[(z(i)] = z(i_0)$.

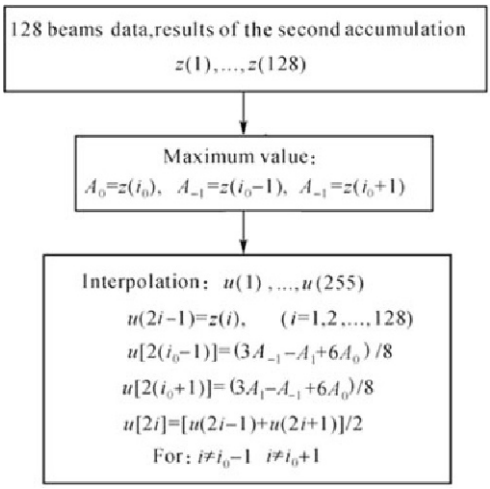


Fig. 6.9. Subroutine of beam interpolation

Denote $A_{-1} = z(i_0 - 1)$, $A_1 = z(i_0 + 1)$, based on the rule listed below, we can get new beam data $u(1)$, ..., $u(255)$: $u(2i-1) = z(i)$ $(i = 1, \ldots, 128)$; $u(2(i_0 - 1)) = (3A_{-1} - A_1 + 6A_0)/8$, $u(2(i_0 + 1)) = (3A_1 - A_{-1} + 6A_0)/8$; for $i \neq i_0 - 1$ and $i \neq i_0 + 1$, $u(2i) = [u(2i-1) + u(2i+1)]/2$.

After refining the beams, the beam duration becomes $\theta_0 = 360°/255$. Based on the data $u(1)$, …, $u(255)$, the precise bearing problem can be solved (see Fig. (6.10)). Suppose the maximum value of the main lobe is $u(i_0)$. Denote $Y_0 = u(i_0)$, $Y_1 = u(i_0+1)$ and $Y_{-1} = u(i_0-1)$, we obtain

$$\Delta\theta = \frac{Y_1 - Y_{-1}}{4Y_0 - 2Y_1 - 2Y_{-1}} \times \theta_0$$

The exact target DOA is given by

$$\hat{\theta} = i\theta_0 + \frac{Y_1 - Y_{-1}}{4Y_0 - 2Y_1 - 2Y_{-1}} \times \theta_0$$

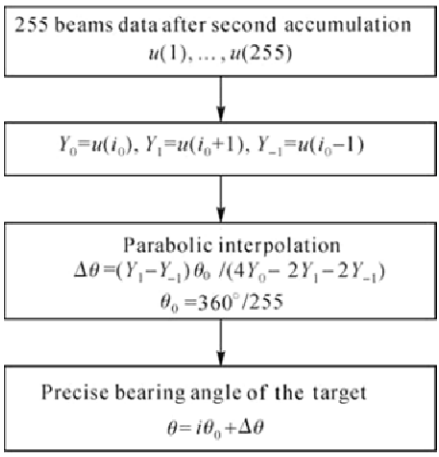


Fig. 6.10. Subroutine of precise bearing of target

There is another problem in target bearing. If the sonar system is installed in a vessel, then the target DOA estimated by the sonar system is related to the moving status of this platform. Sometimes it is important to convert the value of the bearing to the true value, which is related to due north. This procedure is called true bearing angle calibration. The true value of the target DOA is useful sometimes for the commander of a vessel. The course angle α of the vessel is used to calibrate the bearing angle, and then one can obtain the true target DOA. The subroutine of this algorithm is shown in Fig. 6.11.

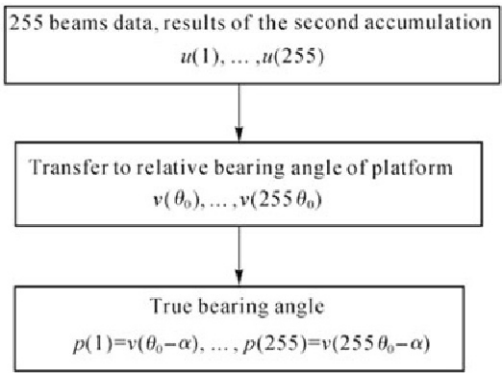


Fig. 6.11. Subroutine of true bearing angle validation

6.1.5 Method for Improving Bearing Accuracy

Based on previous discussions, we know that beamforming data is used to estimate the target DOA, especially the beam which is nearest to the target DOA. Except for the interpolation algorithm described previously, there are also some algorithms that can be used to improve the bearing accuracy. A useful algorithm is time delay estimation for split beams. Although the performance of this method is limited by the sampling frequency, the direct time delay estimation often provides a more precise target DOA estimation. We will introduce the principles of this method of direct time delay estimation, and then show the implementation algorithm.

Suppose the beam, which steered to the target or close to the target, has been found, the left and right beam signals are

$$\begin{aligned} x_1(t) &= s(t) + n_1(t) \\ x_2(t) &= s(t-\tau) + n_2(t) \end{aligned} \tag{6.33}$$

The summation channel signal is: $y(t) = x_1(t)+x_2(t)$..

Suppose the auto-correlation coefficient of the signal $s(t)$ is $\rho_s(\tau)$, then the variance in $y(t)$ is

$$\sigma_y^2 = E[y^2(t)] = 2\sigma_s^2[1+\rho_s(\tau)] + 2\sigma_n^2 = Y(\tau) \tag{6.34}$$

The difference channel signal is

$$z(t) = x_1(t) - x_2(t),$$

$$\sigma_z^2 = E[z^2(t)] = 2\sigma_s^2[1-\rho_s(\tau)] + 2\sigma_n^2 = Z(\tau) \tag{6.35}$$

From Eqs. (6.34) and (6.35), we can see that the differences between $Y(\tau)$ and $Z(\tau)$ are $1+\rho_s(\tau)$ and $1-\rho_s(\tau)$. Our goal is to increase the bearing accuracy. Therefore, it is hopeful that a small change in τ results in a big change in output. It is not that easy, since the variation of $\rho_s(\tau)$ in the neighborhood of $\tau = 0$ is very slow. For example, as we have shown before, if $\rho_s(\tau)$ has the form

$$\rho_s(\tau) = \exp(-\alpha \mid \tau \mid)\cos(2\pi f \tau)\ (\alpha > 0) \tag{6.36}$$

the variation of $\rho_s(\tau)$ at the zero point is very slow.

But if we consider another kind of quantity, such as

$$G(\tau) = \frac{1}{Z(\tau) - 2\sigma_n^2} = \frac{1}{2\sigma_n^2(1-\rho_s(\tau))} \tag{6.37}$$

we found that a small change of $\rho_s(\tau)$ will result in a big variation of $G(\tau)$. In other words, the split beam method has the potential ability to increase bearing accuracy. Of course, there are two preliminary conditions: the precise estimation of the time

delay value τ and the noise power σ_n^2.

A method described here can partially solve these problems, i.e., the method of a time delay estimation algorithm using the cross power spectrum of a split beam.

The model given in Eq. (6.33) is used again. For simplicity, neglect the noise term since, in calculating the cross power spectrum, the cross power of independent noise will be zero.

Suppose we have two signals: $x(t)$ and $y(t) = x(t+\tau)$. The information of time delay τ is in the cross power spectrum of $x(t)$ and $y(t)$.

Denote the Fourier transform of $x(t)$ and $y(t)$ by $X(f)$ and $Y(f)$ respectively

$$X(f) = \int x(t)\exp(-2\pi ft)\mathrm{d}t$$

$$Y(f) = \int y(t)\exp(-2\pi ft)\mathrm{d}t = \exp(2\pi \mathrm{j}f\tau)X(f)$$

The cross power spectrum of $x(t)$ and $y(t)$ is

$$Z(f) = X^*(f)Y(f) = |X(f)|^2 \exp(2\pi \mathrm{j}f\tau)$$

The information of time delay τ can be extracted from the cross power spectrum $Z(f)$. That is

$$2\pi f\tau = \arctan[\mathrm{Im}(Z(f)) / \mathrm{Re}(Z(f))] \tag{6.38}$$

If $x(t)$ is a stationary random signal with zero mean, then the cross power spectrum can be expressed as

$$K_{xy}(f) = \lim_{T\to\infty} \frac{1}{T} E[X_T^*(f)Y_T(f)] \tag{6.39}$$

where $X_T(f)$ and $Y_T(f)$ are the Fourier transform of the truncated signal $x(t)$ and $y(t)$

$$\begin{aligned} X_T(f) &= \int_0^T x(t)\exp(-2\pi \mathrm{j}ft)\mathrm{d}t \\ Y_T(f) &= \int_0^T y(t)\exp(-2\pi \mathrm{j}ft)\mathrm{d}t \end{aligned} \tag{6.40}$$

From Eqs. (6.39) and (6.40), we can see that the time delay value can be calculated by the signal in a limited duration.

In the digital form, suppose the data series are $x(0)$, ..., $x(kT_s)$, ...; $y(0)$, ..., $y(kT_s)$, ...

The DFT of these series, with the start point q and length N are

$$X_q(l) = \sum_{k=0}^{N-1} x((q+k)T_s)\exp(-2\pi \mathrm{j}kl/N)$$

$$Y_q(l) = \sum_{k=0}^{N-1} y((q+k)T_s)\exp(-2\pi \mathrm{j}kl/N)$$

Let $Z_q(l) = X_q^*(l)Y_q(l)$ $(l = 0,\ ...,\ N-1)$, we have

$$\varphi_l = 2\pi f_l \tau = \arctan[\mathrm{Im}(Z_q(l))/\mathrm{Re}(Z_q(l))] \tag{6.41}$$

In order to obtain a precise value of φ_l, it is necessary to have enough long data. A multiple times average is necessary. Suppose

$$Z(l) = \frac{1}{K}\sum_{m=1}^{K} Z_{mN}(l) \tag{6.42}$$

In any frequency bins, we can get the time delay estimation of τ. But it is worth showing that for any pairs of $(f_l,\ \varphi_l)$, the estimation value τ may be different. Data assimilation is necessary. The least mean square method can be used to obtain a new estimation value $\hat{\tau}_{\mathrm{ms}}$ of τ

$$\hat{\tau}_{ms} = \frac{1}{2\pi}\left[\left(\sum_{l=1}^{N-1} f_j\varphi_l\right)/\left(\sum_{l=1}^{N-1} f_l^2\right)\right] \tag{6.43}$$

The estimation error is given by [24]

$$\mathrm{Var}[\hat{\tau}_{\mathrm{ms}}] = \frac{1}{2\pi}\left(\sum_{l=1}^{N-1}\frac{f_l^2}{\mathrm{Var}(\varphi_l)}\right)^{-1/2} \tag{6.44}$$

If $\mathrm{Var}(\varphi_l)$ is independent of l, then

$$\mathrm{Var}[\hat{\tau}_{\mathrm{ms}}] = \frac{1}{2\pi}\left(\sum_{l=1}^{N-1} f_l^2\right)\sqrt{\mathrm{Var}(\varphi_l)} \approx \frac{T_s}{2\pi}\sqrt{3\mathrm{Var}(\hat{\varphi})/NK} \tag{6.45}$$

where $\hat{\varphi}$ is the estimation value of φ. From Eq. (6.45) we can see that the estimation error is proportional to T_s, but inversely proportional to $\sqrt{NK}$. This is easy to understand.

Finally, it is worth showing that the estimation value of φ is a periodic quantity. Generally speaking, the value of the arctangent function is in $[-\pi/2,\ \pi/2]$, therefore only for a small value of τ are there no ambiguity problems. This should be considered in sonar design.

6.2 Automatic Tracking and Multi-target Resolution

The performance of target tracking of a sonar system is strongly related to the ability of multiple target resolution since, if the resolution ability of the sonar is bad, when two targets are spatially closer to each other it is very hard to keep tracking these targets. In target tracking there are two different but dependent concepts: tracking distance and detecting distance.

The tracking distance indicates the furthest distance that a sonar system can steadily track a target until the target disappears. The detecting distance indicates the distance at which a target can be detected, but it is not guaranteed that the target can be steadily tracked thereafter.

6.2.1 Tracking Algorithm of Passive Sonar

For a modern passive sonar system, it is necessary to have the ability to track many targets in real time. Most of passive sonar can simultaneously track up to 8 targets or even more. There are several target tracking techniques that have been used in passive sonar, including fully automatic, semi-automatic, manual and preset tracking. In this section, the full and semi-automatic tracking techniques are described.

In order to explain the tracking problems in passive sonar, we will briefly discuss the tracking and detecting concept. This is quite important for a sonar operator.

A tracking target means the sonar system can continuously track the target and send the information of the bearing angle to the command / control system. In most cases, the tracking distance corresponds to the minimum detectable SNR. Usually, the detecting distance is larger than the tracking distance. The tracking distance and detecting distance are all solved from a sonar equation. It is usual to use the detection index as the threshold. We have

$$d = \frac{D|_{s+n} - D_n}{\sigma}$$

Based on the ROC (receive operation curve), we can calculate the relationship between the detection index and the detection probability (Table 6.2).

Table 6.2 Relationship of detection index and detection probability

D	1	2	3
$DT = 20\lg d$ (dB)	0	6	9
p_d	0.680	0.954	0.997

In the case of a tracking target, usually one takes DT = 6 dB and at this value the detection probability P_d = 0.954. However, in detecting a target mode, the

value of *DT* can be a little lower, therefore the detection probability will decrease. This is the reason why we might lose the target once a target is detected.

6.2.1.1 Fully Automatic Tracking

Fig. 6.12 illustrates the block diagram of fully automatic target tracking of passive sonar. The requirement of this tracking mode is that the sonar system working in a scenario where there are no operators on duty. Therefore, the sonar system must automatically detect the target and track it. This is actually a subsystem of automatic alarming of target intruding. The block diagram presented in Fig. 6.12 is a subroutine for one target detecting, locking and tracking. It is easily generated for multiple targets.

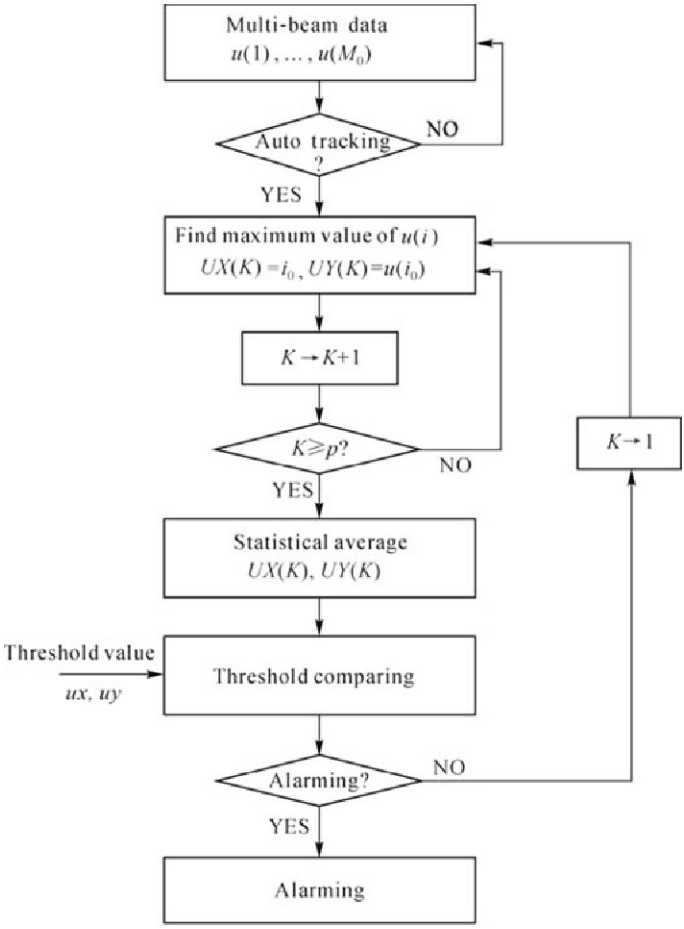


Fig. 6.12. Block diagram of automatic tracking of target

Suppose the beam data of the second accumulation are $u(1)$, …, $u(M_0)$. In the full automatic tracking mode, it is necessary to search the maximum value of $u(1)$, …, $u(M_0)$ and take away the wild value. In the p times continuous search, if the maximum value is on the same beam, for example, i_0, then one chooses this beam as the target DOA, and a subsequent tracking subroutine is started. The number p is determined by experience and the target movement parameters. The value of UX and UY in Fig. 6.12 should be averaged to compare with the threshold and then a final decision made, whether to send the intrude alarming signal.

Figs. 6.13 and 6.14 illustrate the flow chart of the algorithm for full automatic target tracking and the maximum value searching subroutine. In this algorithm, the beam data extension is necessary for maximum value searching. The parameter of p is selected as 3. In Fig. 6.13, the beam data is interpolated in order to have enough bearing accuracy.

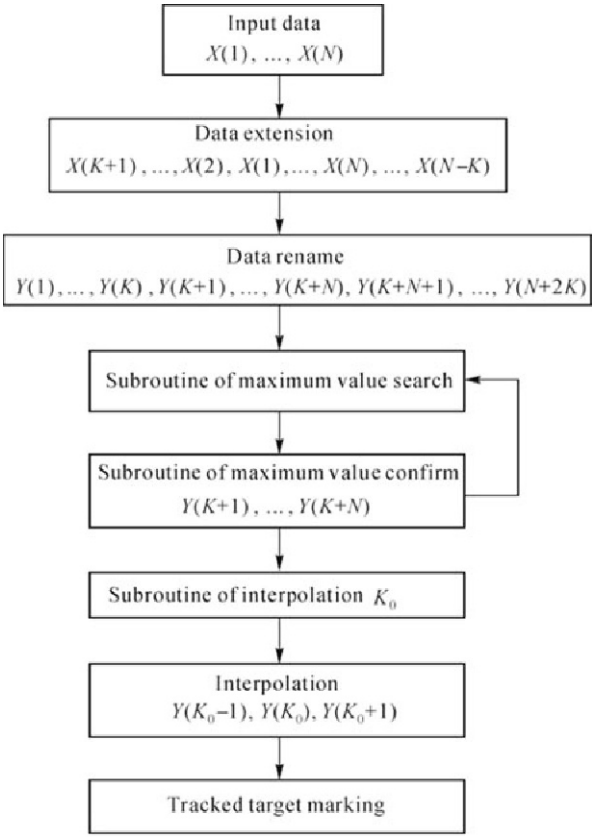


Fig. 6.13. Flow chart of automatic tracking for target

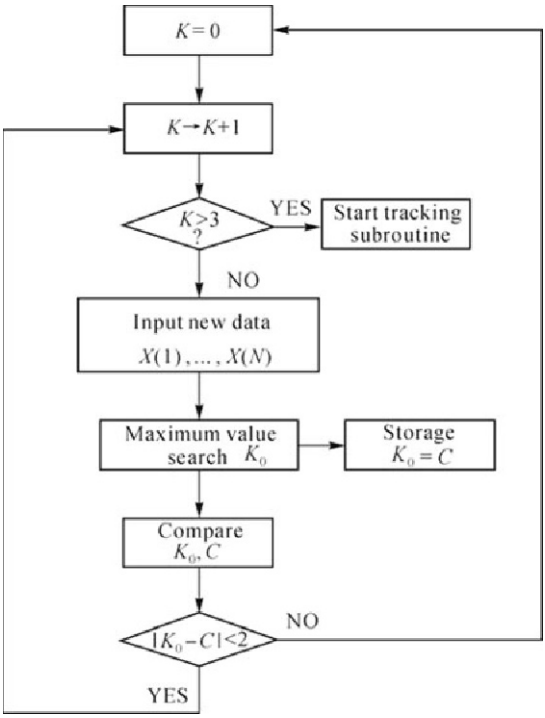


Fig. 6.14. Maximum value search subroutine in automatic target tracking

6.2.1.2 Semi-automatic Tracking

The semi-automatic tracking technique is the most popular technique used in passive sonar. Once the target is confirmed by the sonar operator, one should manually input the information to the sonar consol to mark the target and then switch the sonar system to automatic tracking mode. The marks indicated by the operator can have a specific color, symbols and digits. When the tracked target is lost, the operator can delete the specific marks and then begin new tracking. The principle of a semi-automatic tracking technique is illustrated in Fig. 6.15.

The operator looks at the time / bearing display or amplitude / bearing display. If a contact is to be confirmed, he / she should move the cursor to the right position, for example, beam K_0, and the sonar turns to the automatic tracking mode, i.e., chooses three beams around K_0: K_0-1, K_0, and K_0+1, or five beams: K_0-2, K_0-1, K_0, K_0+1 and K_0+2. The data from these beams are used to realize tracking. For example, in the case of three beams, by using $Y(K_0-1)$, $Y(K_0)$, and Y(K_0+1) , we get the precise bearing information

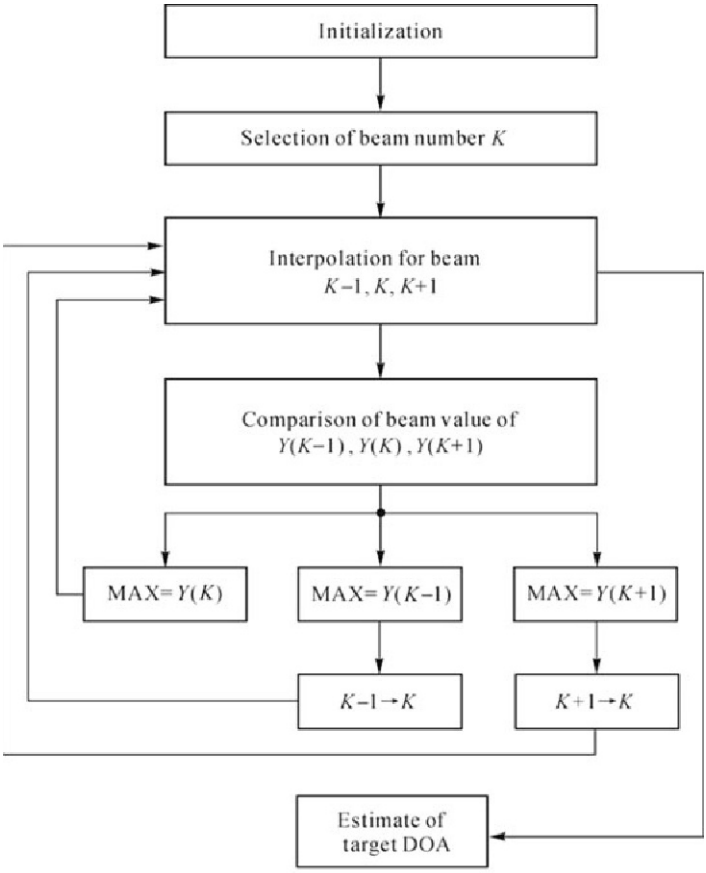


Fig. 6.15. Flow chart of manual tracking of the target

$$\hat{\theta} = K_0\theta_0 + \frac{Y(K_0+1) - Y(K_0-1)}{4Y(K_0) - 2Y(K_0-1) - 2Y(K_0+1)} \tag{6.46}$$

This information should be sent to the display system. At the same time it is necessary to search a new maximum value to compare with value K_0 and assign a new beam to replace $Y(K_0)$. The subroutine of the semi-automatic tracking technique is illustrated in Fig. 6.16.

If the input SNR is very small, or there is some random fluctuation in the comparison procedure of $Y(K_0-1)$, $Y(K_0)$ and $Y(K_0+1)$, then the tracking process may be interrupted and the target to be tracked would be lost. In this case, interposition by a sonar operator is necessary.

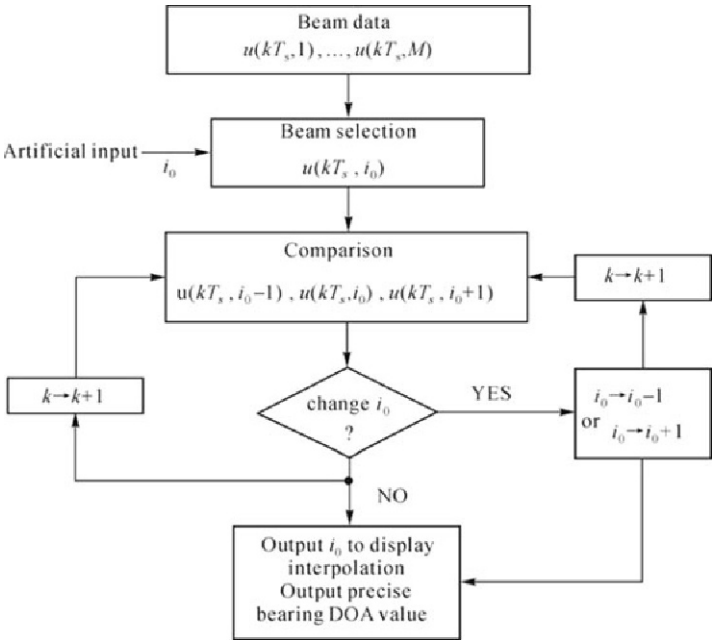


Fig. 6.16. Flow chart of passive tracking in the mode of artificial input beam number

When the target is locked and tracked, the sonar subsystem can begin to search a new target and try to track it if necessary. Usually there is a number p, which defines the maximum targets a sonar system can track. This procedure is illustrated in Fig. 6.17. For most of modern digital sonar, it can track up to 8 targets. The ability of sonar detection can further increase, which depends on the practical requirements.

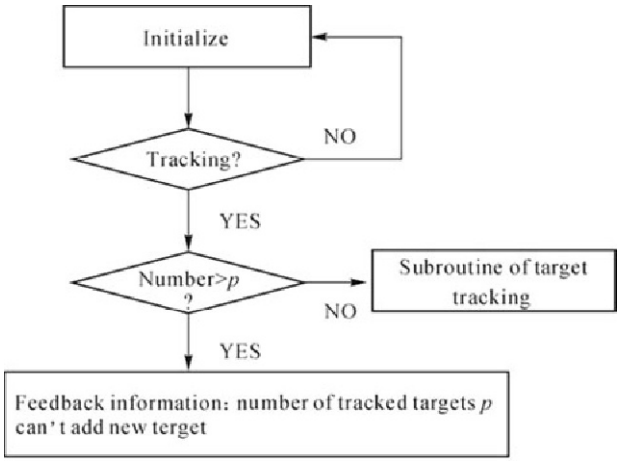


Fig. 6.17. Block diagram of manual driven target tracking

6.2.2 Multi-target Resolution

The resolution of multiple targets obviously depends on the source level of the targets. When we talk about resolution ability, we will limit our discussion to the case of two equal strength targets. That is, the resolution ability of multiple targets is defined as the minimum angle which the sonar system can distinguish for two equal strength targets, as we have shown in Fig. 6.1(b).

Suppose the directivity function of target A, in the neighborhood of the main lobe has the form of

$$D(\theta) = 1 - \alpha\theta^2 \tag{6.47}$$

where $\alpha > 0$ is the parameter to describe the width of the main lobe. As we know, the half beam width of 3 B is

$$\theta_0 = 0.5 \times \theta_{3\ \mathrm{dB}} = 0.54 / \sqrt{\alpha} \tag{6.48}$$

It means $D(\theta_0) = 0.707$.

If an equal strength target B, which is close to θ_0, appeared in θ' then the main lobe of the synthetic directivity function should be

$$G(\theta) = [D^2(\theta) + D^2(\theta - \theta')]^{1/2} = \{(1 - \alpha\theta^2)^2 + [1 - \alpha(\theta' - \theta)^2]^2\}^{1/2} \tag{6.49}$$

The straightforward calculation shows that the valley point of $G(\theta)$ is in $\theta'/2$ and

$$G(\theta'/2) = \sqrt{2}(1 - \alpha\theta'^2/4) \tag{6.50}$$

We will choose the value of $G(\theta'/2) = 0.707G(0)$ as the definition of the resolution ability. It is easy to prove that $G(1.3\theta_0) = 0.707G(0)$, and $\theta' = 2.6\theta_0$. It means that the resolution ability of multiple targets is roughly equal to the 1.3 times main lobe width of the directivity function. This conclusion looks simple but is very important in sonar design. Since, for the line array and circle array, the relationship between beam width and array aperture are already established, it is not difficult, therefore, to derive the relationship between resolution ability and array aperture.

Example

The resolution ability of line array with length L is

$$\Delta\theta \approx 1.3 \times 0.88\lambda / L = 1.14\lambda / L \tag{6.51}$$

The resolution ability of a circle array with radius r is

$$\Delta\theta \approx 1.3 \times 1.12\lambda / (\pi r) = 1.456\lambda / (\pi r) \tag{6.52}$$

6.2.3 Results of System Simulation

The theoretical results described above have been proved by system simulation. Some of the system simulation results are illustrated here. A circle array with radius 1.6 m is considered in this simulation. There are 32 elements that are uniformly distributed in the circle array. The directivity function of a single hydrophone is supposed to have exponential form

$$\xi(\theta) = \exp[-(\theta / 90)^2] \tag{6.53}$$

The unit of θ is degree. The working sector consists of $M = 16$. The results are illustrated in Figs.6.18 – 6.22.

Fig. 6.18 shows the directivity function of two equal strength targets, $\Delta\theta = 10°$. The power spectrum of white noise is flat over the frequency band. The value of the power spectrum of radiated noise of the submarine decreases with frequency 6 dB per octave. As we can see, the resolution ability of white noise is much better than that of the radiated noise of the submarine. The depth of the valley point is about 3 dB for white noise, but only 1 dB for radiated noise.

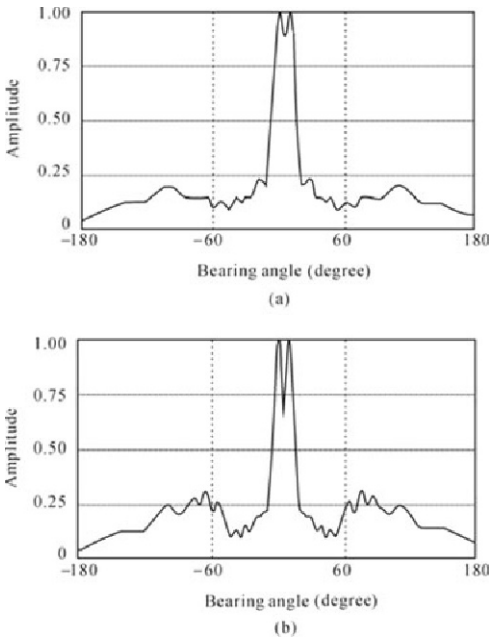


Fig. 6.18. Beam pattern of two equal strength targets ($N = 32$, $M = 16$, 4 – 10 kHz, circle array). (a) Radiated noise of surface ship $\Delta\theta = 10°$; (b) White noise$\Delta\theta = 10°$

Fig. 6.19 is the situation of the radiated noise of the submarine, in a frequency band 4 – 10 kHz, $\Delta\theta = 13°$ and 12°. We can see the depth of the valley point strongly depends on $\Delta\theta$. Fig. 6.20 shows another simulation result, the frequency band is

changed from 4 – 10 kHz to 2 – 8 kHz. Since the band has shifted to low frequency, the resolution ability decreases. This is understandable. Since the low frequency corresponds to a wider main lobe, therefore the resolution ability will decrease.

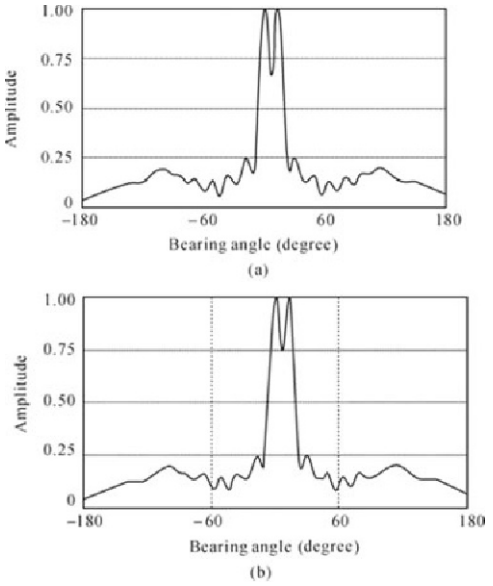


Fig. 6.19. Beam pattern of two equal strength targets (Radiated noise of surface ship, $N = 32$, $M = 16$, 4 – 10 kHz, circle array). (a) $\Delta\theta = 13°$; (b) $\Delta\theta = 12°$

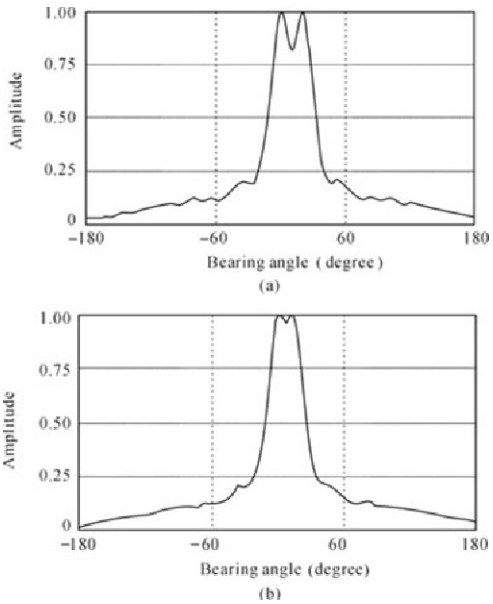


Fig. 6.20. Beam pattern of two equal strength targets (Radiated noise of surface ship, $N = 32$, $M = 16$, 2 – 8 kHz, circle array). (a) $\Delta\theta = 20°$; (b) $\Delta\theta = 15°$

It is worth showing that the resolution ability of a sonar system is also related to the strength ratio of two targets. Figs. 6.21 and 6.22 illustrate an example of two un-equal strength targets. Two targets have a strength ratio of 9.5 dB and the resolution ability depends on the frequency band. When the frequency band is 4 – 10 kHz, the two targets can be distinguished when $\Delta\theta = 15°$. But for a frequency band of 2 – 8 kHz or 2 – 10 kHz, two targets can be distinguished when $\Delta\theta = 25°$. We can think that one target is a weak signal to be extracted and the other target is strong interference. These simulation results show how a weak signal can be detected when there exists strong interference close to the signal.

Fig. 6.23 is a typical example of the display of a digital sonar system. This is a display scenario of a towed array sonar. At the top there is an amplitude bearing display which presents the response of a transient moment of the multi-beam output. And below the amplitude / bearing display is the time / bearing display. It is a 30-min bearing history of a towed array sonar. Each row of the time / bearing display is the intensity (brightness) conversion of the amplitude / bearing display. The newest row in the time / bearing display is just the amplitude / bearing display. We can see the correspondence relationship in Fig. 6.23.

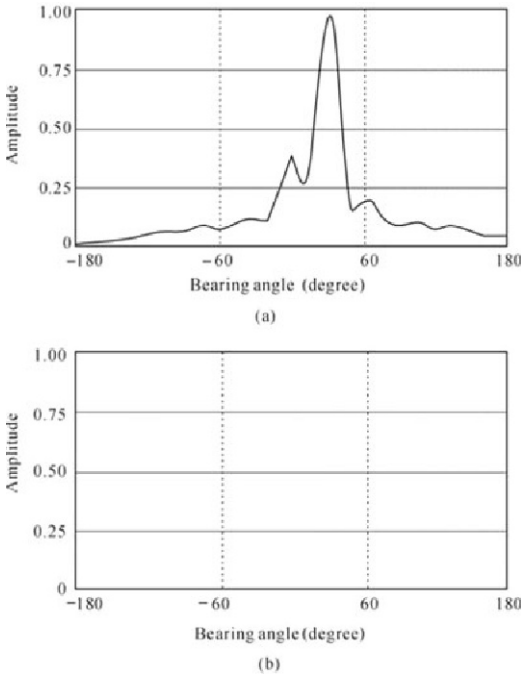


Fig. 6.21. Beam pattern of two unequal strength targets (Radiated noise of surface ship, $N = 32$, $M = 16$, ISR = 9.5 dB, circle array). (a) 2 – 10 kHz, $\Delta\theta = 30°$; (b) 2 – 8 kHz, $\Delta\theta = 25°$

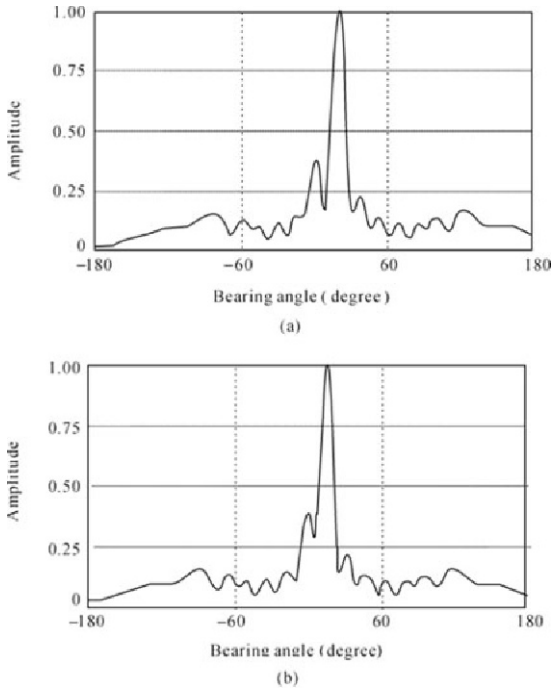


Fig. 6.22. Beam pattern of two un-equal strength targets (Radiated noise of surface ship, $N = 32$, $M = 16$, ISR = 9.5 dB, 4 – 10 kHz, circle array). (a) $\Delta\theta = 20$; (b) $\Delta\theta = 15°$

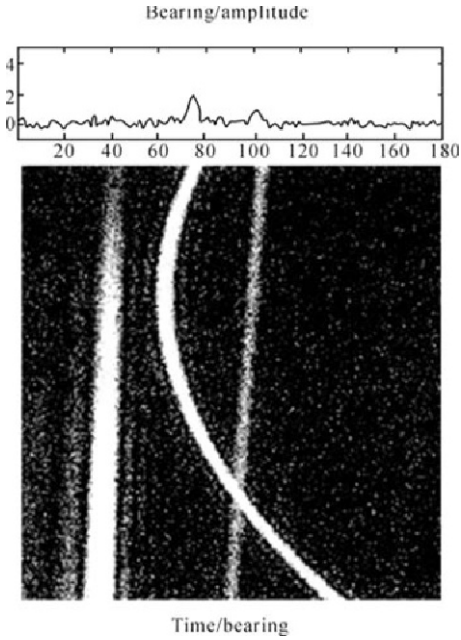


Fig. 6.23. The main screen picture of a digital sonar display

6.2.4 Method for Increasing the Resolution Ability of Multiple Targets

The theoretical analysis presented above shows that the resolution ability of multiple targets strongly depends on the directivity function of the sonar system. Therefore, if we want to increase the resolution ability in the frame of conventional beamforming, it is necessary to use a higher frequency. But the application of a higher frequency will result in a decrease in the ranging distance. This is a contradiction in sonar application. Many researchers have studied how to use other techniques to increase the resolution ability. Messer shows[119] that if two targets have different frequency components, then the frequency identification technique is a good choice for increasing resolution ability. But it is not always possible to use this technique since the target we are interested in usually has the same frequency band as the interference.

The technique introduced here is an application of adaptive beamforming theory. The key point is to form a directional beam, which has zero response to the direction of interference. Then the resolution ability is increased (see Fig. 6.24). We want to find a set of filters which steer in the interference direction. Therefore the interference beam is formed. The interference in each channel is subtracted by using the interference beam.

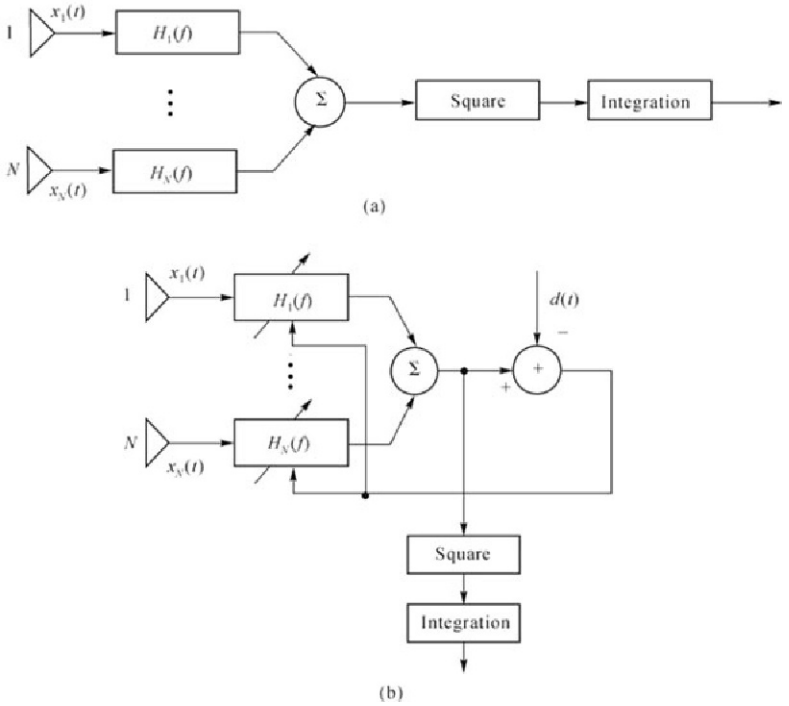


Fig. 6.24. Block diagram of CBF and ABF. (a) CBF (conventional beamforming); (b) ABF (adaptive beamforming)

In order to evaluate the performance of this system, we try to find the steady state resolution of the optimum filter and then calculate the directivity function of the adaptive beamforming system.

Suppose the input signals $x_1(t)$, ..., $x_N(t)$ are a stationary random process. In vector notation we have

$$\boldsymbol{x}(t) = (x_1(t),\ ...,\ x_N(t))^{\mathrm{T}} \tag{6.54}$$

Assume that the signal in the channel consists of signal M interference and isotropic noise. The incidental angle of the signal is θ_0; the incidental angle of interference is θ_1, ..., θ_M. The directional vector of signal and interference is expressed by the vector $\boldsymbol{A}(f,\ \theta_0)$ and $\boldsymbol{A}(f,\ \theta_m)$ $(m = 1,\ ...,\ M)$ respectively

$$\boldsymbol{A}(f, \theta_0) = [\exp(2\pi \mathrm{j} f \tau_1(\theta_0)),\ ...,\ \exp(2\pi \mathrm{j} f \tau_N(\theta_0))]^{\mathrm{T}}$$

$$\boldsymbol{A}(f,\ \theta_m) = [\exp(2\pi \mathrm{j} f \tau_1(\theta_m)),\ ...,\ \exp(2\pi \mathrm{j} f \tau_N(\theta_m))]^{\mathrm{T}}$$

The cross power spectrum of the input signal can be expressed by

$$\begin{aligned}\boldsymbol{K}_{xx}(f) = {} & K_n(f)\boldsymbol{I} + K_s(f)\boldsymbol{A}(f,\ \theta_0)\boldsymbol{A}^{*\mathrm{T}}(f,\ \theta_0) \\ & + \sum_{m=1}^{M} K_{rm}(f)\mathbf{A}(f,\ \theta_m)\boldsymbol{A}^{*\mathrm{T}}(f,\ \theta_m)\end{aligned} \tag{6.55}$$

where $K_n(f)$, $K_s(f)$ and $K_{rm}(f)$ represent the power spectrum background noise, the signal and m-th interference. $\boldsymbol{I}$ is the unit matrix. Suppose the transfer function of N filters is $H_1(f)$, ..., $H_N(f)$. In vector notation

$$\boldsymbol{H}(f) = [H_1(f),\ ...,\ H_N(f)]^{\mathrm{T}}$$

In the case of conventional beamforming

$$\boldsymbol{H}_{\mathrm{cv}}(f) = \boldsymbol{A}^{*}(f,\ \theta) \tag{6.56}$$

Therefore the directivity function for a single frequency signal is

$$D_{\mathrm{cv}}(\theta) = \boldsymbol{H}_{\mathrm{cv}}^{T}(f)\boldsymbol{K}_{\mathrm{ss}}(f)\boldsymbol{H}_{\mathrm{cv}}^{*}(f) \tag{6.57}$$

where $\boldsymbol{K}_{\mathrm{ss}}(f)$ is the signal power spectrum matrix and

$$\boldsymbol{K}_{\mathrm{ss}}(f)=K_{\mathrm{s}}(f)\boldsymbol{A}(f,\ \theta_0)\boldsymbol{A}^{*T}(f,\ \theta_0)$$

Substituting it into Eq. (6.57), we obtain

$$D_{\mathrm{cv}}(\theta)=K_{\mathrm{s}}(f)\,|\,\boldsymbol{A}^{*\mathrm{T}}(f,\ \theta_0)\boldsymbol{A}(f,\ \theta\ )\,|^2 \tag{6.58}$$

For adaptive beamforming, it is proved that we can get a solution of the optimum filter, similar to Eq. (3.11),

$$\boldsymbol{H}_{\mathrm{opt}}(f)=[\boldsymbol{K}^{*}_{\mathrm{xx}}(f)]^{-1}\boldsymbol{K}^{*}_{\mathrm{xd}}(f) \tag{6.59}$$

where $\boldsymbol{K}_{xd}(f)$ represents the cross power spectrum of the input signal vector $\boldsymbol{x}(t)$ with the desired signal (see Fig. 6.24(b)). Therefore we have the directivity function for adaptive beamforming

$$D_{\mathrm{ad}}(\theta)=K_{\mathrm{s}}(f)\,|\,\boldsymbol{A}^{*T}(f,\ \theta_0)\boldsymbol{K}^{-1}_{\mathrm{xx}}(f)\boldsymbol{A}(f,\ \theta\ )\,|^2 \tag{6.60}$$

By using the optimum transfer function, we should force the directivity function to have M zero points in the interference direction.

We can see that the key point in calculating $D_{\mathrm{ad}}(\theta)$ is to find $\boldsymbol{K}^{-1}_{xx}(f)$. For the environment field described in Eq. (6.55), it is proved that [120,121]

$$\boldsymbol{K}^{-1}_{xx}(f)=\frac{1}{K_n(f)}\times[\boldsymbol{I}-\sum_{i,j=0}^{M}B_{ij}\boldsymbol{A}_i\boldsymbol{A}_j^{*\mathrm{T}}] \tag{6.61}$$

where $\boldsymbol{A}_i$ denotes $\boldsymbol{A}(f,\ \theta_i)$ $(i=0,\ 1,\ ...,\ M)$, B_{ij} is the (i,j)-th element of matrix $\boldsymbol{B}$.

$$\boldsymbol{B}=(\boldsymbol{I}-\boldsymbol{DG})^{-1}\boldsymbol{D},\ \ (\boldsymbol{G})_{ij}=\boldsymbol{A}_i^{*\mathrm{T}}\boldsymbol{A}_j,\ \ (\boldsymbol{D})_{ij}=K_{ri}(f)\times\delta_{ij}/K_n(f) \tag{6.62}$$

δ_{ij} is Kronecker $i,\ j$.

Eq. (6.61) is convenient in theoretical analysis, but in practice we prefer to use the following recursive method, and the optimum transfer function will be calculated step by step. As an example, let $M=1$.

$$\boldsymbol{K}_{xx}(f)=K_n(f)\boldsymbol{I}+K_s(f)\boldsymbol{A}_0\boldsymbol{A}_0^{*\mathrm{T}}+K_r(f)\boldsymbol{A}_1\boldsymbol{A}_1^{*\mathrm{T}}$$

Suppose

$$\boldsymbol{Z} = \boldsymbol{I} + \frac{K_r(f)}{K_n(f)} \times \boldsymbol{A}_1 \boldsymbol{A}_1^{*\mathrm{T}}$$

It is easy to verify that

$$\boldsymbol{Z}^{-1} = \boldsymbol{I} - \frac{\boldsymbol{A}_1 \boldsymbol{A}_1^{*\mathrm{T}}}{N + K_n(f)/K_r(f)}$$

and

$$\boldsymbol{K}_{xx}^{-1}(f) = \frac{1}{K_n(f)} \times \left[\boldsymbol{Z}^{-1} - \frac{\boldsymbol{Z}^{-1} \boldsymbol{A}_0 \boldsymbol{A}_0^{*\mathrm{T}} \boldsymbol{Z}^{-1}}{\boldsymbol{A}_0^{*\mathrm{T}} \boldsymbol{Z}^{-1} \boldsymbol{A}_0 + K_n(f)/K_s(f)} \right] \tag{6.63}$$

Substituting Eq. (6.60) into Eq. (6.63), we obtain

$$D_{\mathrm{ad}}(\theta) = \alpha \times |\, q(\theta) \,|^2 \tag{6.64}$$

where α is a constant and

$$q(\theta) = \boldsymbol{A}_0^{*\mathrm{T}} \boldsymbol{Z}^{-1} \boldsymbol{A} \tag{6.65}$$

The behavior of the directivity function is determined by the function $q(\theta)$. It should have zero response in the interference direction and steer to the signal direction.

Suppose the maximum value of $q(\theta)$ is in $\hat{\theta}$. The calculation shows that when θ_0 and θ_1 are very close, then $\hat{\theta} \neq \theta_0$. That means we steer the beam in the interference direction θ_1 to force the response $q(\theta_1) = 0$, but the maximum value would not appear in the direction θ_0, the signal direction. The difference in $\theta_b = \hat{\theta} - \theta_0$ is called the bias angle. We conclude that the resolution ability of adaptive beamforming is much better than conventional beamforming. The cost is that the incidental angle of the signal will have a small bias to the true value. Fig. 6.25 shows the relationship of the bias angle to the interference direction. This is a line array with 15 elements and $d/\lambda = 0.2$. For this array, the resolution ability of CBF is about 25°, but for the ADF the resolution ability becomes 5°. The system gain of CBF G_{cv} and system gain of ABF G_{ad} are given in Fig. 6.26. the system gain of ABF is much better than CBF when the interference is very close to the signal, spatially. Therefore the adaptive beamforming technique is an effective method of improving the resolution ability, in the case where the angle between signal and interference is very small.

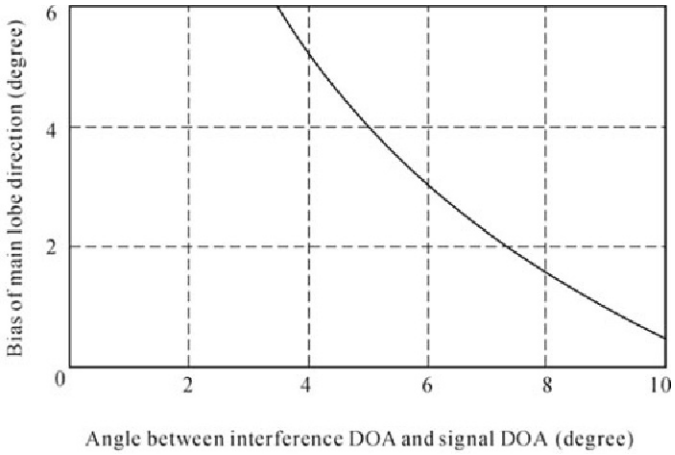


Fig. 6.25. Bias of main lobe direction due to strong interference suppression

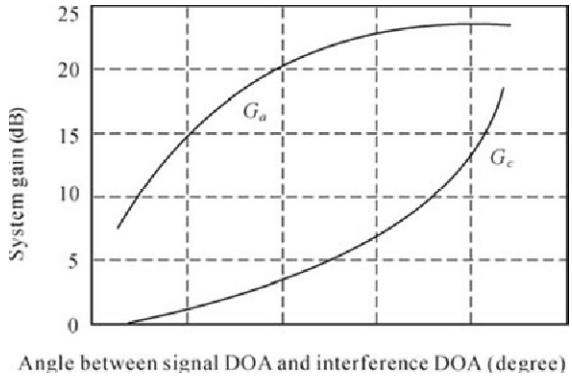


Fig. 6.26. Comparison of system gain of CBF and ABF in the case of strong interference

6.3 Moving Target Analysis: Pre-set Tracking

Target motion analysis (or target moving analysis) (TMA) is also called contact moving analysis (CMA). This is a hot topic in recent research in underwater acoustics. When a submarine cruises underwater, it seldom uses the active sonar. Therefore TMA is the main technique for tracking a submarine. Basically, the TMA problem is a three-dimensional problem. Some authors suggested using wave guide theory, singular value decomposition method, multiple signal classification (MUSIC), estimation of signal parameters via rotational invariance technique (ESPRIT) etc., for the TMA problem[2,35-38]. These algorithms are based on a special arithmetic algorithm, which was proposed by Volder in the 1950's[122].

It considerably reduces the coordinate rotation digital computer (CORDIC) matrix calculation quantity. However, until recently, there was no very effective method that could be used to steadily track an unknown object and extract its moving trace. In this section we will discuss some of the TMA techniques which are used in sonar engineering.

6.3.1 Concept of TMA

When the target is detected and tracked by the passive sonar system, it is hoped to extract more information of the tracking process. The information includes:
(1) The distance and movement information of target.
(2) The class of target, i.e., the identification information.
(3) The trace of moving target.

In this section we only discuss the problem of the target trace. Fig. 6.27 illustrates the scenario of target tracking. It is only limited to the case of a plane trace.

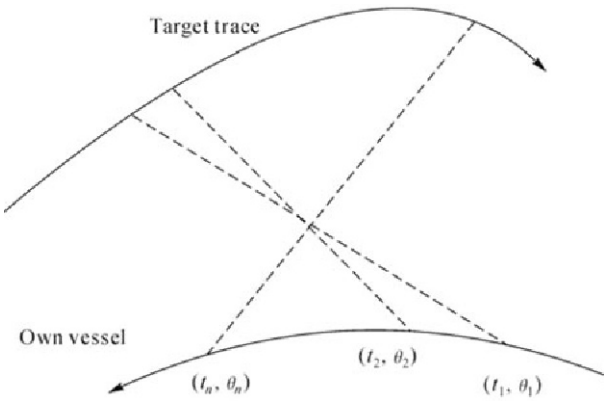


Fig. 6.27. Illustration of moving target tracking

The sonar platform and target to be tracked are moving along a certain trace. At any moment t_i, we can measure the incidental angle θ_i of the target, which is related to the platform axis, a straight line to bow and stern. Therefore the original data of TMA are actually a set of values (t_i, θ_i), since target tracking takes a long time in processing and therefore we must have enough data. Generally speaking, the observation time of TMA is from about 1 minute to several tens of minutes. The longer the observation time, the bigger the error in target tracking. Sometimes it is necessary to convert the relative incidental angle to the true value of the target bearing. Fig. 6.28 shows the relationship between the true value of the bearing

angle and the relative incidental angle. The angle of calibration is based on a north direction. Suppose the north direction is 0°, the platform true course is β, the target bearing angle, defined by the sonar system, is θ. Then the true angle value after calibration is $\alpha = \theta - \beta$.

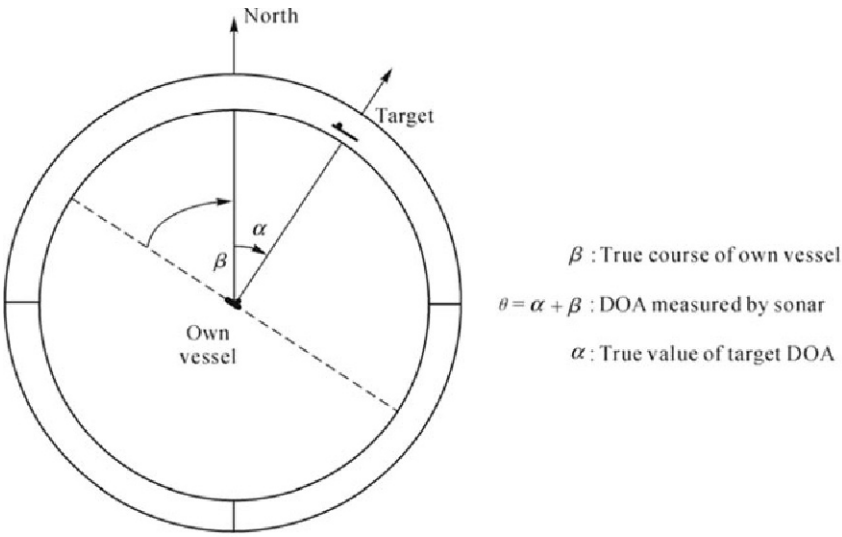


Fig. 6.28. True value of DOA in target tracking

6.3.2 *Pre-setting Tracking*

In the target tracking process, if there are multiple targets then the targets trace may be intersected spatially. Traditional tracking based on the maximum value algorithm may result in the loss of a weak signal. During this time, the data transferred to the command / consol will be interrupted. When the weak signal appears again, we have to renew the tracking procedure by manual interpose. It is necessary to avoid the interruption of the weak signal tracking procedure, otherwise the TMA will face serious problems.

There is a simple algorithm, called pre-setting tracking, which in most cases can guarantee the continuity of weak signal tracking. This topic copes with the target trace prediction. First of all, let us study the target trace in a sonar display (Fig. 6.29). We know that for the same target and the same form of movement, the display images of different sonar are totally different. Suppose a target is doing a circular movement around a sonar platform. For a cylindrical array, the trace on the display screen is a straight line (Fig. 6.29(a)). If the wet end of a sonar system is a linear array, then the trace displayed on screen is a broken line (Fig. 6.29(b)). Therefore, it is important to the sonar operator that the trace on the display screen

is not the actual target moving trace in the ocean. They have to remember the one-to-one corresponding relationship of the actual moving trace with the image displayed on screen. Fig. 6.30 is another typical example. The target is moving along a straight line, and the incidental angle determined by the sonar system is

$$\theta(t) = \arctan[(r \tan \theta_0 - vt) / r] \tag{6.66}$$

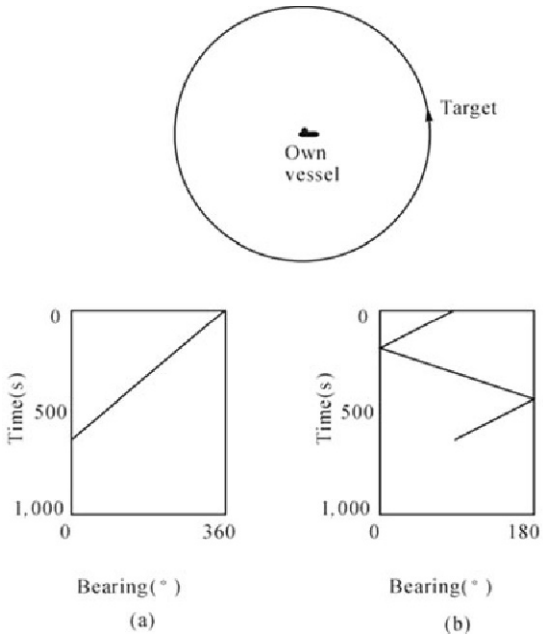


Fig. 6.29. Screen display for different sonar array. (a) Cylindrical array sonar; (b) Linear array sonar

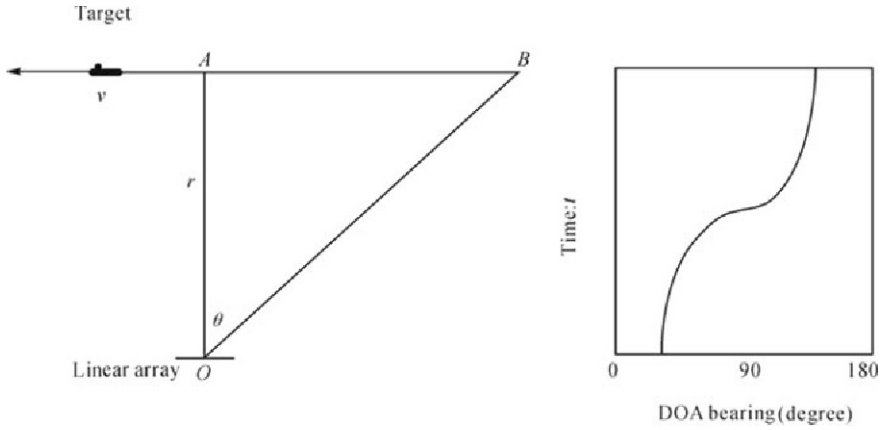


Fig. 6.30. Display example of time / bearing for linear array

It is obviously not the linear function of t. The movement trace on the display screen is a curve, but not a straight line.

The basic idea of pre-setting tracking is to predict target movement at moment t_{n+1} based on the observation data (t_i, θ_i) $(i = 1, ..., n)$. The algorithm assimilates the observation data (t_1, θ_1), ..., (t_n, θ_n) and gives the prediction value of $(t_{n+1}, \hat{\theta}_{n+1})$, $(t_{n+2}, \hat{\theta}_{n+2})$, ... The method to be used here is the least square algorithm. Suppose

$$\theta = At + B \tag{6.67}$$

to find A, B such that

$$J = \sum_{i=1}^{n} (At_i + B - \theta_i)^2 \tag{6.68}$$

is minimum.

Derivating J and letting the derivates be zero, so we can get

$$\begin{aligned} A &= \frac{\sum_{i=1}^{n} (t_i - \overline{t})(\theta_i - \overline{\theta})}{\sum_{i=1}^{n} (t_i - \overline{t})^2} \\ B &= \overline{\theta} - A\overline{t} \end{aligned} \tag{6.69}$$

where $\overline{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$, $\overline{\theta} = \frac{1}{n}\sum_{i=1}^{n} \theta_i$.

Substituting it into Eq. (6.67), we get

$$\hat{\theta}_{n+1} = At_{n+1} + B, \quad \hat{\theta}_{n+2} = At_{n+2} + B, \ .... \tag{6.70}$$

The flow chart of pre-setting tracking is given in Fig. 6.31. This is actually a part of the target track function. It is expected of the tracking procedure that, at any time, it is necessary to store L observation data (for example, $L = 30$). Once the pre-setting subroutine is started, we should stop the conventional tracking program for this target and then perform the pre-setting tracking subroutine until a new command is received. Fig. 6.32 is a concrete example of pre-setting tracking.

The assimilation method described above is a linear prediction; this does not mean the target moves along in a straight line, but it is hard to propose another kind of assimilation curve, since there is no reason to know what kind of moving trace is most likely to occur.

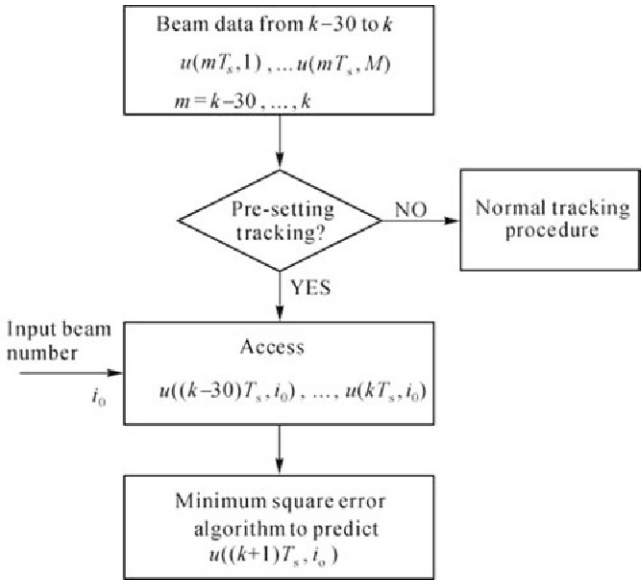


Fig. 6.31. Flow chart of pre-setting tracking subroutine

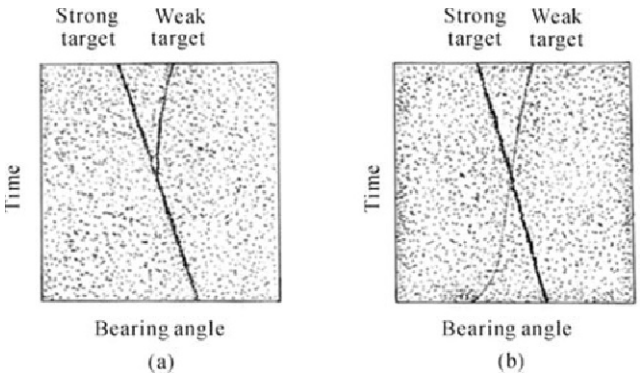


Fig. 6.32. Comparison of maximum value tracking and presetting tracking. (a) Maximum value tracking; (b) Pre-setting tracking

6.3.3 TMA Research

The main goal of TMA (target moving analysis) is to solve the target distance and moving parameters based on the observation data. Because we do not know the target moving trends, so it is difficult to make predictions in terms of a few data. For example, if the target is moving uniformly and along a straight line, then we can easily predict the target distance and moving speed. In Fig. 6.30, if the incidental angle of the target is θ_0 at position B, we obtain $\theta(t) =$

$\text{arc}[(r\tan\theta_0 - vt)/r]$. Based on a set of measurement data $(t_i,\ \theta_i)$, it is easy to solve the value of r and v. Of course, in the solving procedure we may have a multiple value problem and must take the necessary average. It is worth stating that when the target is far away, then integration necessarily takes a long time.

Generally, it is necessary to use the segment approximation method in TMA. The trace of the moving target is divided into many pieces. In each piece the target is supposed to move uniformly and along a straight line and also the sonar platform is fixed in this time duration. Then the prediction equation obtained in this segment can be used in the next segment. The prediction equation should be modified when we get new observation data. This algorithm looks like Kalman filtering, discussed in Chapter 2.

Now we are going to introduce the prediction equation of TMA (Fig. 6.33).

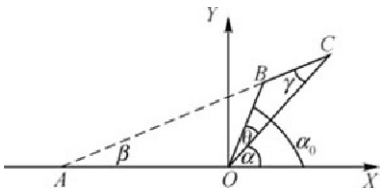


Fig. 6.33. Target moving analysis

Suppose BC is a piece of the target trace. At the moment $t = 0$, the target position is B, the sonar platform is in O. The direction OY steers north. AX is the axis of our own vessel. In a short time the target is uniformly moving along the straight line BC. The angle between BC and AX is β, as shown in Fig. 6.33. Suppose in time duration $[0, t]$ the sonar platform is fixed. When the target moves from B to C, the relative angle is changed from α_0 to α. The polar coordinate of B is (r_0, α_0). We want to find the relationship between θ and t. In $\triangle BCO$, based on the sinusoidal theorem

$$\frac{r_0}{\sin(\alpha-\beta)} = \frac{vt}{\sin(\alpha_0-\alpha)} \tag{6.71}$$

As Shown in Fig. 6.33, $\alpha_0 - \alpha = \theta$, $\alpha_0 - \beta = \gamma$. Substituting them into Eq. (6.71), we obtain

$$\frac{r_0}{vt} = \frac{\sin(\alpha-\beta)}{\sin(\alpha_0-\alpha)} = \frac{\sin(\alpha-\alpha_0+\alpha_0-\beta)}{\sin\theta}$$

Therefore

$$\cot\theta - \cot\gamma = \frac{r_0}{v\sin\gamma}\times\frac{1}{t} \tag{6.72}$$

If we set $y = \cot\theta$, $r_0/(v\sin\gamma) = a$, $\cot\gamma = b$, then Eq. (6.72) becomes

$$y = a / t + b \tag{6.73}$$

We conclude that y is the linear function of $1/t$. Eq. (6.72) is called a prediction equation. It is not difficult to solve the distance r_0 from Eqs. (6.71) and (6.72).

The general procedure for solving the prediction equation is as follows. Suppose we have a set of observation values (t_1, θ_1), ..., (t_n, θ_n). Denote $x_i = 1/t_i$, $y_i = \cot\theta_i$ $(i = 1, ..., n)$, calculate

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{6.74}$$

$$a = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \quad b = \overline{y} - a\,\overline{x} \tag{6.75}$$

In the next moment t_{n+1}, in terms of prediction equation $y = ax + b$, we have

$$\hat{y} = a \times \frac{1}{t_{n+1}} + b, \quad \hat{\theta}_{n+1} = \operatorname{arc\,cot}\hat{y}_{n+1} \tag{6.76}$$

It is worth showing that it is necessary to filter the wild value in processing the final data (θ, r) so that we can get basically smooth the prediction curve.

6.4 Fast Moving Target Tracking: Torpedo Alarming

The torpedo is one of the most threatening weapons for surface ships and submarines. Particularly in underwater battles, the survival of the submarine depends on its capability to detect a torpedo attack. Since the 1960's, the torpedo alarming problem has concerned the sonar designer. Earlier, the torpedo alarming topic belonged to the category of underwater acoustic countermeasures, i.e., the research focused on the decoy design. But the main problem, torpedo alarming, is a really challenging topic in anti-submarine warfare. The difficulties are: Firstly, the spectrum of torpedo noise has wider frequency band. If conventional sonar wants to detect a torpedo attack, big modifications are necessary; Secondly, the trace of the torpedo is three-dimensional, the speed of the torpedo is fast when close and therefore it is harder to detect and classify.

Nevertheless, there are several torpedo alarming techniques and equipment has been introduced on the market[40]. In this section we will introduce a

technique for torpedo alarming, which is based on the trace extraction technique of a fast-moving target[39].

6.4.1 Possibility of Torpedo Alarming

A torpedo is a fast-moving weapon underwater. The size is much smaller than that of a submarine. The normal speed is more than 40 kn (about 20 m / s). Because of the small size, there is more propeller blade and the vibration is relatively less than that of a submarine. Therefore the radiated noise of a torpedo is quite different from that of a submarine. From the passive detection point of view, some features can be useful in passive torpedo alarming:

(1) The importance of the impulse wave resulting from the torpedo launch or the torpedo entering the water.
(2) The importance of the fast change in the grazing angle when the torpedo is close to the target.
(3) The SNR increases for the receiving sonar when the torpedo is close to the sonar receiver.
(4) The frequency spectrum of the torpedo is much wider than that of a surface ship or submarine.
(5) The active target search signal of a torpedo is detectable by the sonar system.

The detection of the active target search signal belongs to the field of signal interception or signal capture. In this section, we will consider the passive detection problem in torpedo alarming.

In fact, the radiated noise of a torpedo can be used to detect a torpedo, since the source of the radiated noise of a torpedo has some different characteristics from those of a surface ship or submarine (Fig. 6.34).

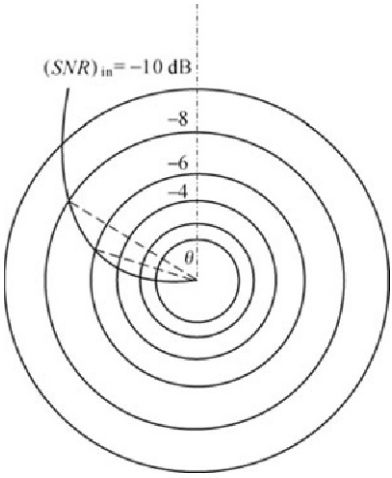


Fig. 6.34. The main parameters of a fast moving target

Suppose the distance between the torpedo and target is 10 km and the tangent speed of a torpedo is 40 kn. In 30 s, the torpedo has travelled about 0.6 km, and the grazing angle is 3.42°. For a sonar system, it is not difficult to detect this kind of change in the bearing angle. On the other hand, when the torpedo is approaching the target, the SNR will increase due to the increase in radiated noise. After time accumulation of a certain duration, the increase in the SNR can be extracted in real time. For example, if the torpedo is moving fast from a distance of 10 km to 9 km, this takes about 50 s and the SNR will increase by 1 – 2 dB. This change should be detected. As we have shown in Chapter 4, the change in the SNR will be appear as a jump in the direct current output (see Fig. 4.83). Therefore there is no doubt that the change in the SNR resulting from a fast approaching torpedo can be extracted.

It is necessary to show that there is a critical distance in torpedo alarming, since the submarine needs a safe time / distance to avoid the approaching torpedo.

Except for the change in grazing angle and the change in the SNR, there is another feature which can be used to detect a torpedo. This is the characteristic of torpedo noise. Since the rotation speed of the torpedo propeller is faster than that of a submarine, therefore the line spectrum components are usually much higher than those of a submarine. This is important information, which can be used to classify torpedo noise.

If the sonar system has a signal intercept or signal capture function in the high frequency band, for example more than 30 kHz, then it will be possible to detect the target search signal transmitted by the torpedo. In this situation, it is necessary to have a specific sonar wet end and dry end for detecting the high frequency impulse signal.

6.4.2 *Double Check Decision Strategy in Torpedo Alarming*

Fig. 6.35 shows the main idea of torpedo alarming. If there is specific sonar for intercepting or capturing the active signal of a torpedo, then it is necessary to have data fusion processing for conventional sonar and specific sonar. For example, if we want to intercept the active target search signal transmitted from a torpedo, the frequency must be higher than 30 kHz. Then the wavelength is 5 cm and the array aperture could be 20 cm. We need a small array and a much higher sampling frequency, for example at least 100 kHz. Nowadays most submarine sonar is not specific sonar for torpedo alarming. Therefore our discussion will be limited to the issue of the torpedo alarming function in a conventional sonar system.

The decision procedure is based on the three signal processing results: grazing angle analysis; target strength change analysis; spectrum analysis of torpedo noise.

For the torpedo alarming technique, the detection probability and false alarm probability are very important parameters. It is preferable to have a low false alarm probability. In order to decrease the false alarm probability and increase the detection probability, we can adopt a double check strategy. The block diagram of

this technique is shown in Fig. 6.36.

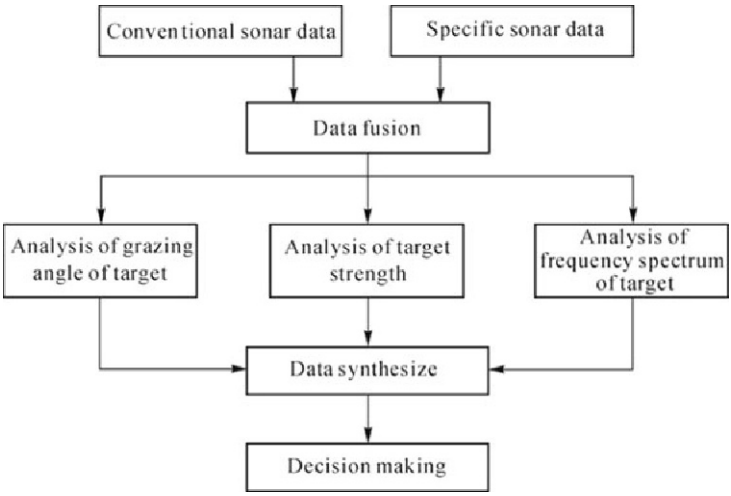


Fig. 6.35. Basic techniques of torpedo alarming

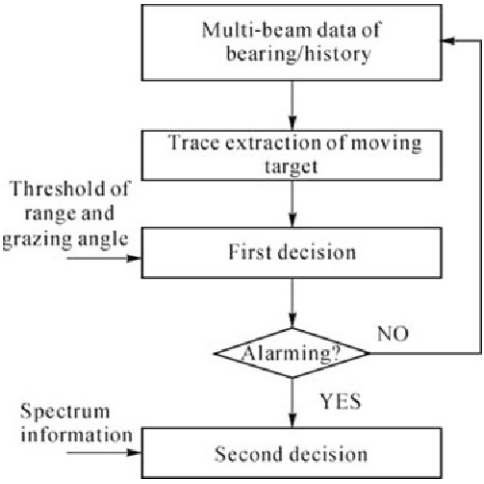


Fig. 6.36. Flow chart of torpedo alarming

From the time / bearing data we cut one frame image, denoted by $P[k, l]$, where k represents the order of row, and l represents the order of column. $k = 0, ..., K-1$; $l = 0, ..., L-1$. In fact, k is the time, the unit is the duration of post processing, for example 3 seconds; l is the bearing angle, the unit is the beam duration, for example 1°. So we have a frame image with $K \times L$ pixels. Perform two-dimensional image processing on this frame, in each 3 seconds complete the processing and take away the last row $P[K-1, l]$. Add a new row of data

$P[0, l]$ from the multi-beam output. Rename the frame $P[k, l]$ $(k < K-1)$ as $P[k+1, l]$, then we get a new frame. To repeat the processing we will extract the trace of a fast-moving target.

In processing the frame $P[k, l]$ $(k = 0, ..., K-1; l = 0, ..., L-1)$, it is necessary to calculate the change in grazing angle and the change in target strength. The pre-setting threshold values $\Delta\theta$ and ΔR are compared with the real time data. If one measured data is larger than the corresponding threshold value, then the first time decision is carried out that confirms the trace may be a torpedo approaching. The next step is to analyze the signal of this trace and the frequency spectrum is analyzed in detail. If the result further confirms it is a torpedo, then the second decision is carried out. Finally the torpedo alarming signal is sent to the display / consol.

The double check procedure will increase the confidence level of torpedo alarming and decrease the false alarm probability. The technique described here needs storage of a large amount of data, including the target trace history, candidate target trace and the spectrum of the candidate trace signal etc. This will be no problem for modern digital sonar, since the hardware is powerful enough to complete this task.

6.4.3 Target Traces Extraction Technique

Target trace extraction is a key point in torpedo alarming. Fig. 6.37 illustrates how to use data-sieves to extract the target trace, especially the fast moving target.



Fig. 6.37. Trace extraction technique for targets

For processing each frame, it may best to adopt a point-to-point process or block process. Therefore the data extension for each row is necessary. Suppose we have a frame of $P[k, l]$ $(k = 0, ..., K-1; l = 0, ..., L-1)$. $P[0, 0]$ is the top left pixel and $P[K-1, L-1]$ is the bottom right pixel. The first row of data is $P[0, 0], ..., P[0, L-1]$. Assume a data window with length $2M$ is selected. It is necessary to extend this row of data to $P[0, M], ..., P[0, 1], P[0, 0], ..., P[0, L-1], P[0, L-2], ..., P[0, L-1-M]$. The rule for extending a row data is: in the left end pixel, copy M data symmetry to $P[0, 0]$; in the right end pixel, copy M data symmetry to $P[0, L-1]$. The same procedure should be taken for each row and also performed on each column. Therefore the new frame actually is the frame with a side frame added of width M.

For the new frame, if we want to perform a point-to-point process, it won't result in any uncertainty about pixels, since the data is already extended. A technique called a data-sieve is adopted to extract the moving target trace. The main goal of data-sieving is to take away any random impulse-like flash and connect the broken point in trace-like pixels. There are three sieves adopted in this technique.

The first data-sieve is a global processing for the image. Random sparkle in the image will be removed by this sieve. This is the point-to-point processing. For each pixel, choose a block window centered at this point and then use the median value filtering to remove the impulse-like sparkle.

The second data-sieve begins in the first row. It is necessary to search the local maximum value and assign these values as the "trace seed", i.e., the starting point of the denominated trace. The method for searching the local maximum value can use the subroutine proposed in Chapter 5. Generally speaking, the identification rule for the local maximum value is two out of three or three subsequent occurrences. After confirming the start point of the denominated trace, the trace search will be operated in a column direction. It is necessary to pay attention to the break point, which may appear in the trace. Of course, the start point of the trace does not necessarily appear in the first row. It could appear in any row, depending on the real situation. Therefore the algorithm for searching the local maximum value should be performed in each row, such that we can renew the traces concerned.

The third data-sieve performs the trace identifying algorithm. It connects the trace we extracted in the previous two sieves and pads a suitable value to the break point, if necessary. And then we smooth and modify the trace such that the trace looks good.

The trace to be extracted may not be only one, and we should adopt the decision rule described above to confirm which one is a fast-moving target. Since the torpedo alarming function has to be done in one cycle of post accumulation, therefore the requirement of the operating time is serious. The pre-processing for torpedo alarming is about 30 – 60s; the moving distance of the torpedo is about 0.5–1 km.

6.4.4 Example of System Simulation

Fig. 6.38 illustrates the results of system simulation. The scenario of this simulation is:

(1) Equal-spaced towed array with 32 elements, $d/\lambda_0 = 0.4$, where λ_0 is the wave length of the center frequency.
(2) 256 beams, each frame has 256×128 pixels; beam duration 360° / 256≈0.703°.
(3) Time duration of adjacent row is 3.2 s, one frame history is about 128×3.2 s ≈409 s.
(4) Three targets:

- Target 1: a fast-moving target's relative strength is 0 dB, grazing angle changes from –15.6° to 44.1°.
- Target 2: a conventional target's relative strength is –7 dB, grazing angle changes from 14.0° to 0°.
- Target 3: A fast approaching target's grazing angle changes from –49.7° to –54.3°, the source level changes from –13 dB to 10 dB.

Since the beam pattern of a towed array is not uniform, so we can see from the figure that the broadside beam is narrow, but in the firing direction the main lobe of the beam pattern becomes wider.

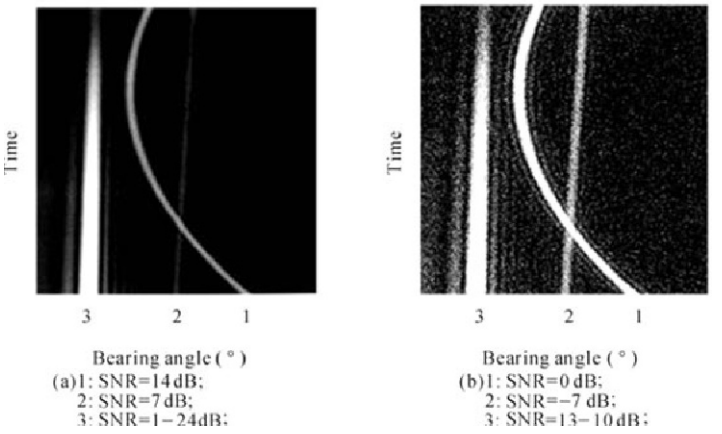


Fig. 6.38. Time / bearing display of moving targets of towed array sonar

The change rate of the grazing angle for target 1 is 59° / 410 = 0.14° / s. The change rate of signal to noise ratio from target 3 is 23 / 410≈0.05 dB / s.

Target 1 and target 3 are selected to be the denominated targets. It is worth showing that in the trace identifying process, the side lobe may be confirmed as traces and this should be avoided in the third sieve.

Finally, we have to mention the characteristic of the frequency spectrum of torpedo noise. Based on the resonant wave formula of a propeller, $f_m = m \times n \times r$, where m is the order number of the resonant wave, n is the blade number of the

propeller and r is the rotation time per second.

Since the value of $n \times r$ for a torpedo is much greater than that of a submarine, so it is easy to find the line spectrum of a torpedo whether in LOFAR or DEMON.

6.5 Passive Ranging of Target Based on the Time Delay Estimation

In order to locate the position of a target, the active sonar has to transmit a signal. This action can be dangerous because the signal transmitted may be detected by the hostile sonar. Therefore, this is a very bad thing for the attack submarine. Since the 1960's, the passive ranging technique has been studied by many researchers. From a theoretical point of view, three-point ranging is possible to locate the position of the target. But the real situation is much more complicated than what we expected. This is because the time delay estimation accuracy in passive ranging is of the order of micro seconds. And the fluctuation resulting from the underwater acoustic channel may be of the same order of magnitude. Therefore, it is necessary to have a technique which can modify the measured time delay value and increase the estimation accuracy.

The time-varying and spatial-varying phenomena has concerned researchers for a long time. Especially in shallow water, the acoustic wave is bounded by both the sea surface and sea bottom. Therefore the time delay estimation becomes more complex.

In this chapter we will introduce the principle of passive ranging and the implementation measures in a real situation.

6.5.1 Principle of Three-Point Ranging

The simplest passive ranging case is three-point locating (Fig. 6.39). There are three receivers; the distance between two receivers is d, where S represents the sound source, H_1, H_2 and H_3 are three receivers, located in a horizontal line.

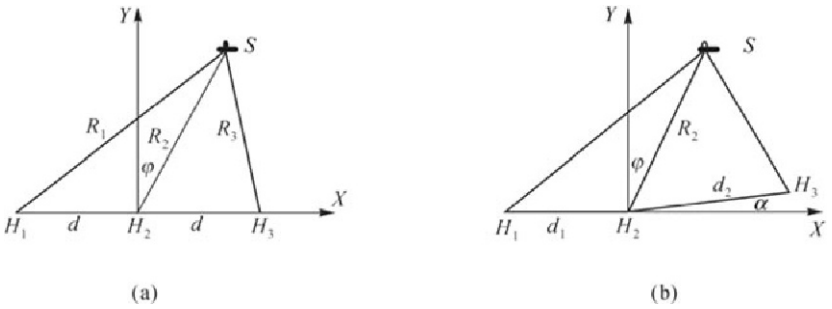


Fig. 6.39. Illustration of triangle ranging. (a) Simple model; (b) General model

Let $H_1H_2 = H_2H_3 = d$. We also denote $SH_1 = R_1$, $SH_2 = R_2 = R$, and $SH_3 = R_3$. We know that R is the distance we want to estimate. Generally speaking, $R >> d$, therefore, the differences among R, R_1 and R_3 are small. From the geometrical point of view, if we know the value of d, $R - R_1$ and $R_3 - R$, we can find the value of R. Therefore it is necessary to find the time delay τ_{12} between received signals at H_1 and H_2, and the time delay τ_{23} between received signals at H_2 and H_3.

For simplification, H_2 is selected as the origin of the coordinate (see Fig. 6.39(a)).

Assume $x_0 = R\sin\varphi$, $y_0 = R\cos\varphi$, and φ is the angle between SH_2 and y axis. It is obvious that $SH_1 - SH_2 = c\tau_{12}$, $SH_2 - SH_3 = c\tau_{23}$, where c represents the sound speed. We have

$$[(x_0 + d)^2 + y_0^2]^{1/2} - (x_0^2 + y_0^2)^{1/2} = c\tau_{12} \tag{6.77}$$

$$(x_0^2 + y_0^2)^{1/2} - [(x_0 - d)^2 + y_0^2]^{1/2} = c\tau_{23} \tag{6.78}$$

Since $x_0^2 + y_0^2 = R^2$, we obtain from Eqs. (6.77) and (6.78)

$$(x_0 + d)^2 + y_0^2 = (c\tau_{12})^2 + 2c\tau_{12}R + R^2 \tag{6.79}$$

$$(x_0 - d)^2 + y_0^2 = (c\tau_{23})^2 - 2c\tau_{23}R + R^2 \tag{6.80}$$

Summing up these two equations we have

$$R = \frac{2d^2 - c^2(\tau_{12}^2 + \tau_{23}^2)}{2c(\tau_{12} - \tau_{23})} \tag{6.81}$$

Eq. (6.81) shows how to find distance R from τ_{12} and τ_{23}. We also hope to estimate the incidental angle φ. It is necessary to make some simplified assumptions.

Based on our assumption, $R_1 = R + c\tau_{12}$, $R_3 = R - c\tau_{23}$, therefore

$$c\tau_{12} = R_1 - R = \left[d^2 + R^2 - 2Rd\cos\left(\frac{\pi}{2} + \varphi\right)\right]^{1/2} - R$$

$$= R[(1 + (d/R)^2 + 2(d/R)\sin\varphi)^{1/2} - 1]$$

Based on the Taylor expansion formula, we know that $(1 + x^2)^{1/2} \approx 1 + x/2$

$+x^2/8$, ($|x|<<1$). Assume $x = 2(d/R)\sin\varphi + (d/R)^2$, We have

$$c\tau_{12} \approx d\sin\varphi + d^2\cos^2\varphi/(2R) \tag{6.82}$$

The equation is valid in the assumption

$$d/R << 1 \tag{6.83}$$

Similar to Eq. (6.82), we have

$$c\tau_{23} \approx d\sin\varphi - d^2\cos^2\varphi/(2R) \tag{6.84}$$

Combining Eqs. (6.82) and (6.83), it follows

$$\varphi = \arcsin[c(\tau_{12} + \tau_{23})/2d] \tag{6.85}$$

$$R = d^2\cos^2\varphi/[c(\tau_{12} - \tau_{23})] \tag{6.86}$$

Eqs. (6.85) and (6.86) provide the basic formula in passive ranging sonar. The formula shows that if we can precisely measure the time delay τ_{12} and τ_{23}, then the target distance can be calculated. In the limit case, if the incidental signal is a plane wave, then $\tau_{12} = \tau_{23}$, and $R = \infty$. This is easy to understand. Let us estimate how precise a time delay is necessary in solving distance. See Eqs. (6.82) and (6.84), in the broadside direction, $\varphi = 0$. $c\tau_{12} \approx d^2/(2R)$, $c\tau_{23} \approx -d^2/(2R)$. If $d = 20$ m, $R = 10$ km, then $\tau_{12} \approx 13$ μs, $\tau_{23} \approx -13$ μs. That means the time delay accuracy should be in the order of μs.

It is worth showing that the greater the distance *R*, the smaller the time delay difference τ_{12} and τ_{23}. Also, when the incidental angle of the target is biased in a broadside direction, then τ_{12} and τ_{23} become smaller.

The situation described above is the simplest case; it is not difficult to derive the calculation formula in the general case. In Fig. 6.39(b), suppose $H_1H_2 = d_1$ and $H_2H_3 = d_2$. Three receivers are not located on the same horizontal line. H_2H_3 has an incline angle α to the horizontal axis. It is proved that[125]

$$R = \frac{d_1^2\cos^2\varphi + d_2^2\cos^2(\varphi+\alpha)}{2c(\tau_{12} - \tau_{23}) + 2l} \tag{6.87}$$

where

$$l = d_2\sin(\varphi+\alpha) - d_1\sin\varphi \tag{6.88}$$

These two equations are used to calculate the target distance. The incidental angle is solved from the following equations

$$c\tau_{12} = d_1 \sin\varphi \tag{6.89}$$

$$c\tau_{23} = d_2 \sin(\varphi + \alpha) \tag{6.90}$$

6.5.2 Four-Point Passive Ranging

The four-point passive ranging method is suitable for the large aperture array. In three-point ranging, the length of the base line of the time delay τ_{12} and τ_{23} is d. If d is very large, then τ_{12} and τ_{23} may have some kind of fluctuation, resulting from sound propagation. But if d is too small, then the time delay difference may be small too. This will decrease the ranging accuracy. There is a trade-off algorithm, that is the four-point passive ranging method.

In Fig. 6.40, four receivers H_i $(i = 1, 2, 3, 4)$ are used to measure the distance R between target S and receivers. Assume $H_1H_2 = H_3H_4 = l$, $H_1H_4 = L$, and H_0 is the center of H_1H_4.

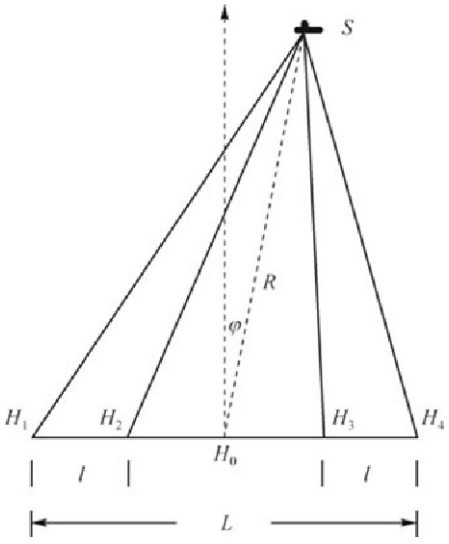


Fig. 6.40. Illustration of four-point array for passive ranging

According to Eq. (6.86), for the three-point array $H_1H_0H_4$, we have

$$R = \frac{L^2 \cos^2 \varphi}{4c(\tau_{10} - \tau_{04})} \tag{6.91}$$

Where $\tau_{10} = \tau_{12} + \tau_{20}$, $\tau_{04} = \tau_{03} + \tau_{34}$, therefore

$$R = \frac{L^2 \cos^2 \varphi}{4c(\tau_{12} - \tau_{34}) + 4c(\tau_{20} - \tau_{03})} \tag{6.92}$$

Similarly, for the three-point array $H_2 H_0 H_3$, we have

$$R = \frac{(L/2 - l)^2 \cos^2 \varphi}{c(\tau_{20} - \tau_{03})} = \frac{(L^2 - 4Ll + 4l^2)\cos^2 \varphi}{4c(\tau_{20} - \tau_{03})} \tag{6.93}$$

Substituting Eqs. (6.92) to (6.93) it follows

$$R = \frac{Ll \cos^2 \phi}{c(\tau_{12} - \tau_{34})} - \frac{l^2 \cos^2 \phi}{c(\tau_{12} - \tau_{34})} \tag{6.94}$$

In the case of $l << L$, Eq. (6.94) becomes

$$R \approx \frac{Ll \cos^2 \varphi}{c(\tau_{12} - \tau_{34})} \tag{6.95}$$

This is the formula of four-point passive ranging. Sometimes, four-point passive ranging is the sub-optimum method for measuring the target distance.

6.5.3 *Analysis of Passive Ranging Accuracy*

The accuracy of passive ranging is related to the target distance, array aperture and the time delay error. As an example, In Eq. (6.86), the estimate of target range is

$$\hat{R} = \frac{d^2 \cos^2 \varphi}{c(\tau_{12} - \tau_{23})}$$

It is proved that the variance of $\hat{R}$ is

$$\mathrm{Var}(\hat{R}) = c^2 \mathrm{Var}(\tau_{12} - \tau_{23})(R / L_e)^4 \tag{6.96}$$

where $L_e = d \cos \varphi$ is called the equivalent length of the array. Carter proved that Eq. (6.96) is close to the Cramer-Rao low bound[19].

From Eq. (6.96) we can derive a very important formula in passive ranging. It shows that the relative ranging error is equal to the relative time delay error. In fact, if we rewrite Eq. (6.96) as

$$\frac{[\mathrm{Var}(\hat{R})]^{1/2}}{R}=\frac{c[\mathrm{Var}(\tau_{12}-\tau_{23})]^{1/2}R}{L_e^2}=\frac{[\mathrm{Var}(\tau_{12}-\tau_{23})]^{1/2}}{d^2\cos^2\varphi/(cR)}=\frac{[\mathrm{Var}(\tau_{12}-\tau_{23})]^{1/2}}{\tau_{12}-\tau_{23}}$$

it is equivalent to

$$\Delta R/R=\Delta\tau/\tau \tag{6.97}$$

where $\Delta R=[\mathrm{Var}(\hat{R})]^{1/2}$ represents the average error of passive ranging, and $\Delta\tau=[\mathrm{Var}(\tau_{12}-\tau_{23})]^{1/2}$ represents the time delay error. We have shown before, when the target is far from the passive sonar, that the time delay value is small and therefore the relative error will increase.

Example

In section 6.1 we have shown that the accuracy of the time delay estimation depends on the integration time of the system and the SNR. Fig. 6.41 is an example. The frequency band of this sonar is 0.6 – 1.0 kHz. The optimum low bound of the time delay estimation is plotted in this figure. When the integration time is 10 s, the low bound of time delay estimation is $(\Delta\tau)_{\mathrm{opt}}\approx 2$ μs for $(SNR)_{\mathrm{in}}=0$. Based on this calculation, it is not difficult to derive the passive ranging accuracy.

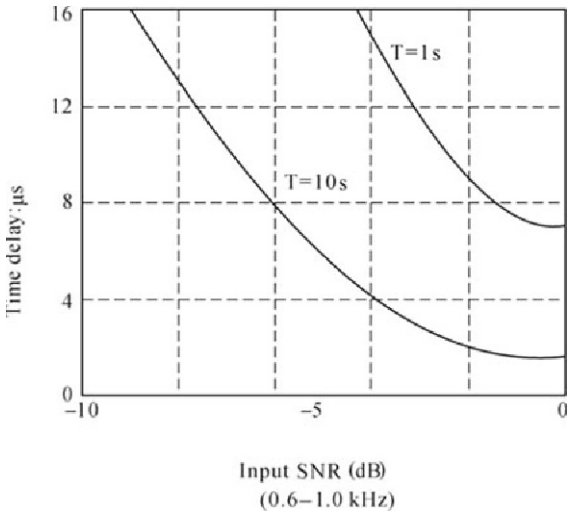


Fig. 6.41. The optimum low bound of error of time delay estimation

It is worth showing that the distance measurement error is also related to the incidental angle of the target. The relationship between the relative ranging error and incidental angle looks like the shape of a bathtub; sometimes it is called the bathtub curve of passive ranging sonar (Fig. 6.42). The theoretical explanation can be easily found in Eqs. (6.86) and (6.87).

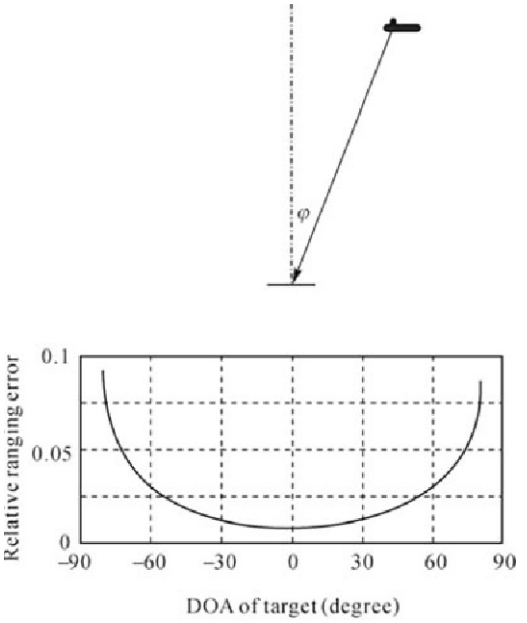


Fig. 6.42. Relationship between ranging error and DOA of target

Example

Suppose there is a three-point array, the distance of adjacent elements is $d = 30$ m. In the broadside direction ($\varphi = 0$), if $\Delta\tau = 10$ μs and $R = 5$ km, then $\Delta R \approx 600$ m and the relative ranging error is 12%. If the incidental angle changes to $\varphi = 30^\circ$, then the relative ranging error is about 20%.

Since the accuracy of passive ranging is very sensitive to the time delay estimation, the requirements of the array installation are very important. A 1-cm installation error may result in 6.7 μs time delay. So it is no surprise that there are some specific requirements in wet end installation for passive ranging sonar. Fig. 6.43 illustrates the basic plot in analyzing the effects of array installation.

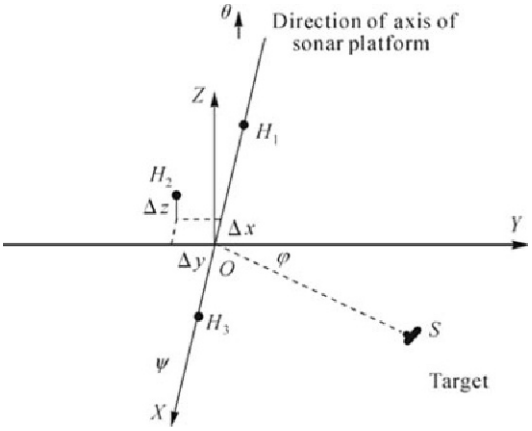


Fig. 6.43. Error in array installation

Suppose H_1, H_2, and H_3 are three receivers which are supposed to be installed in a straight line, but H_2 inclines the normal position due to installation problems. Assume the target is in the plane of $z = 0$ if the platform of the passive sonar has a rolling angle ψ and pitching angle θ. The installation errors are shown in Fig. 6.43, that are Δx, Δy and Δz. In this situation[124]

$$\phi = \arcsin[c(\tau_{12} + \tau_{23}) / (2d \cos\psi)] \tag{6.98}$$

$$R = \frac{d^2 \cos^2 \varphi \cos^2 \psi}{c(\tau_{12} - \tau_{23}) + 2l} \tag{6.99}$$

$$\begin{aligned} l = {} & \Delta x \sin\psi \sin\varphi + \Delta y(\sin\theta \sin\psi \sin\varphi + \cos\theta \cos\varphi) \\ & + \Delta z(\sin\theta \cos\varphi - \cos\theta \sin\psi \sin\varphi) \end{aligned} \tag{6.100}$$

If $\Delta x = \Delta y = \Delta z = 0$, then Eq. (6.99) becomes Eq. (6.86). In practice, the ideal installation doesnot exist; the calibration is very important. Generally speaking, the ranging error is most sensitive to the Δy, but not sensitive to the Δz. For Δx, the sensitivity depends on the incidental angle φ. With the decrease in φ, the sensitivity will decrease too.

Table 6.3 shows the typical calculation results[124].

Table 6.3 The effect of installation accuracy for passive ranging

d (m)	Δx (mm)	Δy (mm)	Δz (mm)	ψ (°)	θ (°)	φ (°)	R (km)	$\Delta R / R$ (%)
10	1			10	20	45	10	29
10		1		10	20	45	10	45
10			1	10	20	45	10	5

6.5.4 Realization of Passive Ranging

There are two kinds of passive ranging sonar in a submarine. One is conformal sonar where the six sub-arrays are installed on both sides of the submarine and each sub-array represents a receiver. This array is also called a Wide aperture array (WAA). Of course, the flank array can also be used for passive ranging. The other one is towed array sonar. Since the acoustic module of a towed array sonar is far away from propeller noise, therefore the interference is much smaller than WAA. And the aperture of towed array sonar has more flexibility, so it is possible to use a four-point passive ranging technique.

Fig. 6.44 shows the block diagram of three-point passive ranging sonar. The wet end of this sonar consists of three sub-arrays, and each sub-array has its own beamforming system. In the conduct of bearing sonar, the same steering beam of three sub-array systems is selected. The values of time delay differences τ_{12} and τ_{23} are calculated from the beam signal. Since the incidental direction of the target is unpredictable, therefore it is necessary to have a pre-time-delay algorithm which can manage 0° – 360° DOA. There are several methods that can be used to precisely measure the time delay value of τ_{12}, τ_{23}, for example, adaptive filtering, cross power spectrum method and correlation interpolation, etc. Usually it is necessary to divide the frequency band into many sub-bands and then take frequency weighting for the estimates. The flow chart of the ranging algorithm is illustrated in Fig. 6.45.

Example

Suppose the distance of the adjacent array is d, the incidental angle of the target is φ. Then the value of the pre-time-delay is $\Delta = INT[d \sin \varphi / T_s] \times T_s = n_0 \times T_s$, where T_S is the sampling time duration.

In order to calculate time delay value τ_{12} and τ_{23}, the signal from sub-array 2 should be delayed by Δ, and perform the correlation operation with the signal from sub-array 1. And the signal from sub-array 3 should be delayed by Δ, and perform the correlation operation with the signal from sub-array 2. The results of time delay differences τ_{12}, τ_{23} are used to solve target distance R.

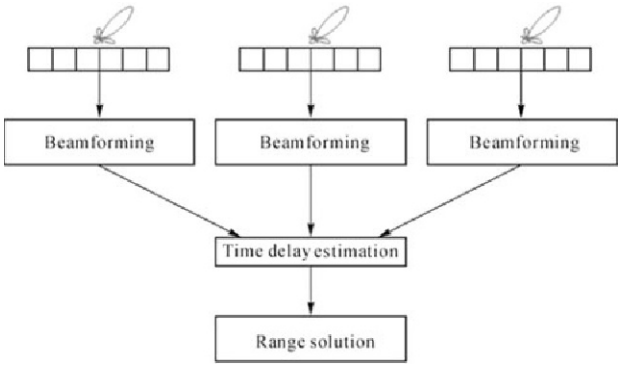


Fig. 6.44. Realization of passive target ranging

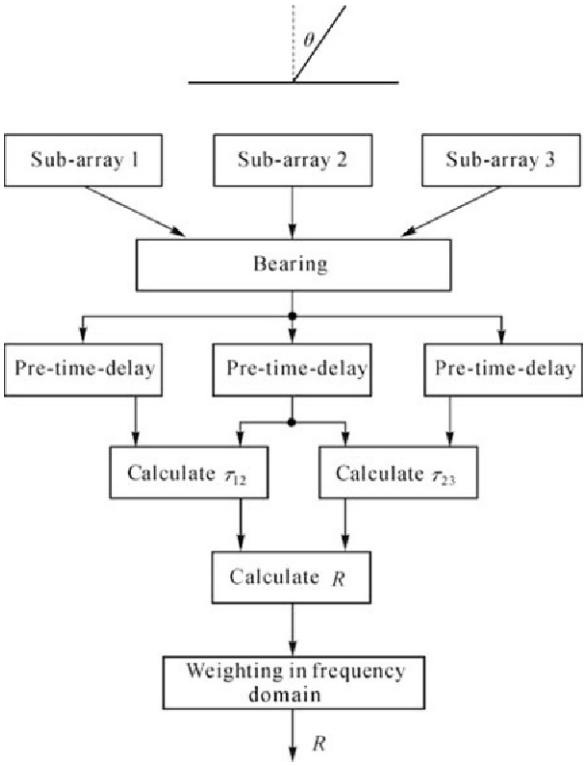


Fig. 6.45. Example of passive ranging algorithm

6.6 Target Classification and Recognition

Underwater target classification and recognition is a very attractive topic in sonar design. More precisely speaking, the topic we are concerned with is not only the target classification, but also the underwater acoustic signal classification. Therefore, the classification and recognition problems include the active sonar signal, echo signal, radiated signal and torpedo noise.

The theoretical results and technical achievements in speech recognition, fingerprint identification and human face recognition can be referenced in sonar signal processing[43, 47-49]. But the recognition of an underwater acoustic signal has specific characteristics; the noise of the propeller plays a very important role in radiated noise recognition. In a certain sense the recognition technique in underwater acoustics is more interested in the feature of the frequency domain.

The Naval Office of the USA began artificial neural network research earlier and some application software has already been introduced to the market[44-48].

Another effective technique for underwater acoustic signal recognition is to

use an expert system[51-58]. This is a branch of artificial intelligence. Among these algorithms the nearest neighborhood criterion, based on the cluster analysis concept developed in mathematical statistics, is used in target noise recognition[52].

6.6.1 Basic Requirement of Target Noise Classification

The basic requirement of target noise classification in digital sonar is to determine the signal category with a high confidence level. Therefore, it usually needs a higher SNR. And the procedure for target noise classification has to use the original signal rather than the processed signal. This is because the original signal contains necessary information about the target, and the processed signal may lose some information which is important in the target classification.

Fig. 6.46 shows that the position of the target noise classification problem is entirely sonar. The signal to be used in classification comes from the multi-beam system, of course the beam which steers to the target. The items we intend to classify are the active sonar signal, echo wave and radiated noise of the target, etc. The means of analysis can be various, such as time domain, frequency domain, or the combination of time / frequency domain.

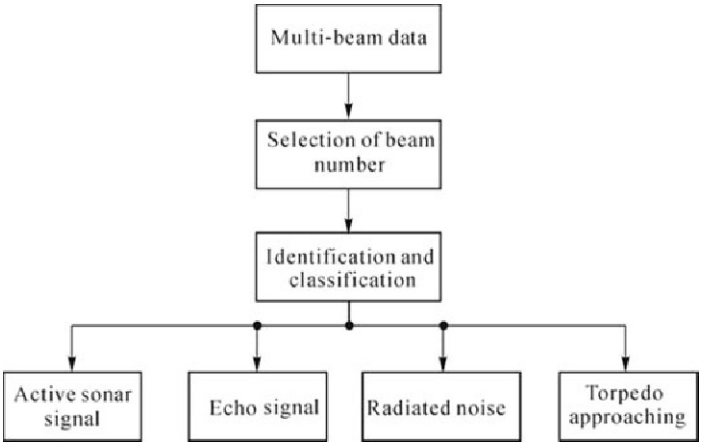


Fig. 6.46. Identification and classification in digital sonar

In studying the classification problem, it is natural to think of the signal processing mechanism in the human brain. It is obvious that the human brain has perfect audio, video and image processing ability. The capability of object classification and recognition is incredible. The human brain is a black box for us (Fig. 6.47).

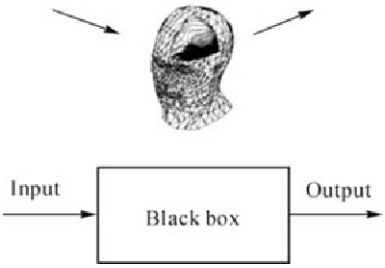


Fig. 6.47. Principle of black box

Based on the so-called "black box principle", we can try to understand some kinds of simple functions in an unknown system. For example, if the system input is 1, 2, 3, … and the output is 2, 4, 6, …, then we can guess that this system may be an amplifier, which doubles the input quantity. If the output of a system is 1, 3, 6, … for the input 1, 2, 3, …, we think it may be an accumulator. However, the mechanism of signal processing is impossible to explore in a simple test. Nevertheless, the research on artificial neural networks has already had some theoretical achievements and results. Some of it is helpful in developing applicable techniques in target noise classification.

6.6.2 Concept of Artificial Neural Network

The brain of a human being has the functions of association, memory, learning and searching. Artificial neural network studies try to simulate the function of our brain. It is estimated that there are about 10^{11} neutrons in our brain, and each neutron is connected to 10^4 other neutrons. This is a very complicated system, such that any models are too simple. Until now, about 50 models have been proposed by researchers in the artificial neural network field. Among them, about 15 models are regularly used in artificial intelligence research. Some of these can be used in target noise classification[45, 46].

(1) Hopfield / Kohonen model: This is a non-learning type model; the weight coefficients are set up beforehand. It is mainly used in image reconstruction.
(2) Back propagation model: This is a model that needs training and learning, used in speech synthesis.
(3) Boltzmann model: This model needs long time learning and training. The weight coefficients are adjusted to reach a so-called minimum energy state. It is mainly used in sonar and radar image recognition.
(4) Self mapping model: This model needs long time training and learning, mainly used in image recognition.
(5) Multi-layer sensing model: This is a self-learning model; the weight co-efficients in each layer can be adjusted. It is mainly used in noise recognition.
(6) Neo-cognition model: This is a model with massive conduct nodes. It is

mainly used in handwriting recognition.

we are going to introduce the multi-layer neutron models, the most popular application model.

Suppose the input data vector is $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$, the sensor network function of neutron is

$$u(\boldsymbol{x}) = w_0 + \sum_{i=1}^{n} w_i x_i \tag{6.101}$$

The output of this network is supposed to be threshold processing

$$y(\boldsymbol{x}) = \begin{cases} 1 & u(\boldsymbol{x}) \geq 0 \\ 0 & u(\boldsymbol{x}) < 0 \end{cases} \tag{6.102}$$

In a multi-layer model, there are several middle layers, which are hidden between the input and output layers. Their connection is also via the network function, but may not be linear. Fig. 6.48 illustrates a four-layer model. Except for the input and output layers, there are two middle layers. Assume they have M and H neutrons, respectively. This kind of model may contain a large amount of weight coefficients. The modeling goal of an artificial neutral network is to train and learn the samples, and finally find the weight coefficients. Of course, the training and learning procedure must guarantee that the input samples will result in known results. The consequent comparison will support the training and learning process.

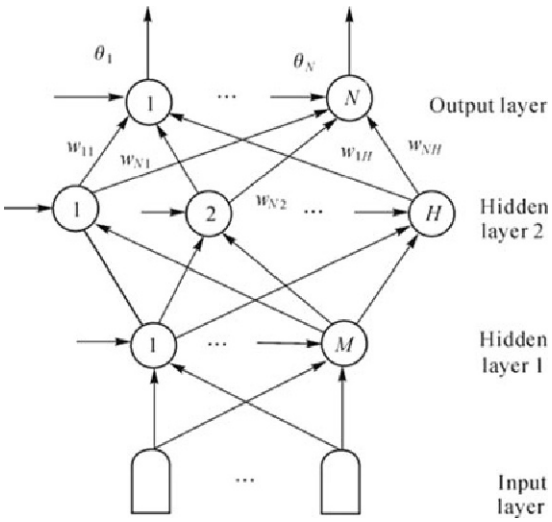


Fig. 6.48. Block diagram of multi-layer artificial neural network

For example, the training sample set is given by $\boldsymbol{x}(i)$, $d(i)$, $i \in I_r$. We also have a test sample set: $\boldsymbol{x}(i)$, $d(i)$, $i \in I_t$. The sets I_r and I_t do not intersect. The iteration formula for the weight coefficient is

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \eta(d(k) - y(k))\boldsymbol{x}(k) \tag{6.103}$$

After each iteration, the samples are renewed. $d(i)$ is the expectation output, which takes value 0 or 1. The value $y(k)$ and $d(k)$ is compared in the iteration process; the difference between them is used to adjust $\boldsymbol{w}(k)$. When the value of η is adopted well, Eq. (6.103) is convergent. It is proved that if

$$0 < \eta < \frac{1}{|\boldsymbol{x}(k)|_{\max}} \tag{6.104}$$

and $\boldsymbol{x}(k)$ has a separate property (something like independence of $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$), then Eq. (6.103) is convergent. The learning process will end in finite steps.

6.6.3 Simplified Expert System

Expert system theory is a branch of artificial intelligence. An expert system is a special system, which has an expert level of knowledge about some specific item. The knowledge database and deduction reasoning rule base can make a decision about the question to be solved.

An expert system is suitable for target classification and recognition. The self-learning ability is particularly useful in the database extension of underwater acoustics. Fig. 6.49 illustrates the basic structure of an expert system. There are two basic databases: signal features database and deductive reasoning rule database.

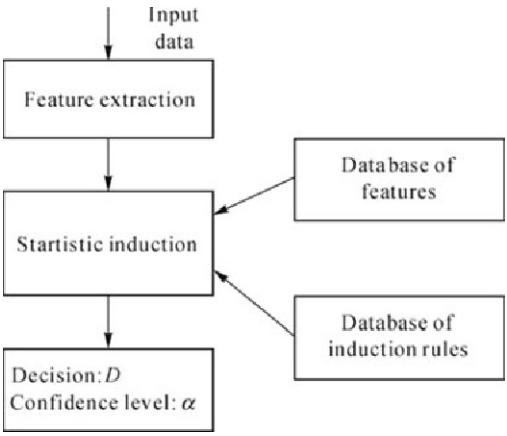


Fig. 6.49. Basic structure of expert system for target recognition

For the signal to be processed, the basic features of this signal are analyzed and extracted. The features are compared with the items stored in the database. The comparison procedure is based on the rule which stores in the rule database. Then, the system performs deduction reasoning and obtains the conclusion. And it also gives the value of the confidence level.

6.6.4 An Example of an Expert System Based on Nearest Neighborhood Criteria

The expert system based on the nearest neighborhood criteria, is supposed to be used in underwater target classification and recognition. The target types to be classified are: submarine, surface fighting ship, cargo and fishing boat. In the case where the SNR is too low to make a decision, it allows us to choose the conclusion: no targets (ambient noise only). Therefore the system describes five types of signal (submarine, surface fighting ship, cargo, fishing boat and ambient noise) (Fig. 6.50).

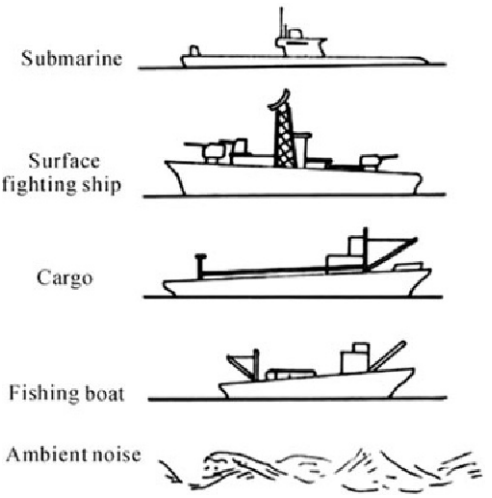


Fig. 6.50. Illustration of target classification and identification

The nearest neighborhood criterion is based on the theory of cluster analysis in mathematical statistics. The concept is easy to understand. Suppose the features of observation data can be extracted and expressed as a vector: $\boldsymbol{x} = (x_1, \ ..., \ x_N)^{\mathrm{T}}$. All vectors consist of a feature space X. For a different type of target, the feature class is different. For any given class $\boldsymbol{A}_i$, there is a representative vector, which can represent the features of this class. Assume this vector is

$$\boldsymbol{a}_i = (x_1^{(i)}, \ ..., \ x_N^{(i)})^{\mathrm{T}} \ (i = 1, \ ..., \ M) \tag{6.105}$$

The selection of $\boldsymbol{a}_i$ should, as far as possible, avoid an intersection. And also we have A_1+…+A_M = $\boldsymbol{X}$. That means there are no features which do not belong to any class.

For a given vector $\boldsymbol{x} \in \boldsymbol{X}$, how to determine if $\boldsymbol{x}$ belongs to $\boldsymbol{A}_i$? The most comprehensive method is to calculate the "distance" $\rho(\boldsymbol{x}, \ \boldsymbol{a}_i)$ between $\boldsymbol{x}$ and $\boldsymbol{a}_i$. If there is i_0, such that $\rho(\boldsymbol{x}, \ \boldsymbol{a}_i)$ minimum, then we can say that $\boldsymbol{x} \in \boldsymbol{A}_{i_0}$. That is to say, if $\rho(\boldsymbol{x}, \ \boldsymbol{a}_{i_0}) = \min\limits_i \{\rho(\boldsymbol{x}, \boldsymbol{a}_i)\}$, then $\boldsymbol{x} \in \boldsymbol{A}_{i_0}$. The confidence level α of decision $\{ D : \boldsymbol{x} \in \boldsymbol{A}_{i_0} \}$ is proportional to $1 / \rho(\boldsymbol{x}, \boldsymbol{a}_{i_0})\}$.

A very important question is how to define the distance between features. The feature vector can be abstract; the component of the vector may not be the real or complex number. But the distance between features must be a non-negative real value. It must satisfy the three axioms of Hilbert space that are:

(1) Non negative: $\rho(\boldsymbol{a}, \ \boldsymbol{b}) \geq 0$;

(2) Self consistency: $\rho(\boldsymbol{a}, \ \boldsymbol{a}) = 0$;

(3) Triangle inequality: $\rho(\boldsymbol{a}, \ \boldsymbol{c}) \leq \rho(\boldsymbol{a}, \ \boldsymbol{b}) + \rho(\boldsymbol{b}, \ \boldsymbol{c})$.

Fig. 6.51 is an explanation of cluster analysis.

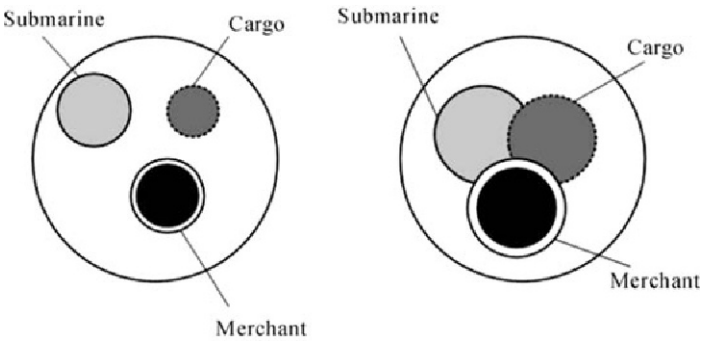


Fig. 6.51. Cluster analysis used in target recognition

Having the above preparative knowledge about cluster analysis theory, we are going to discuss the design details of an expert system.

As the representative vector of a class, it must satisfy the following conditions:

(1) The parameter of the feature described must be the result of a statistical average.

(2) Under the same conditions, the probability of feature occurrences is near to 1.

(3) The different classes in metric space must have the smallest intersection.

(4) The feature extraction algorithm must be easy to realize in hardware / software.

Based on the large amount of experimental and data analysis, we know the power spectrum of radiated noise of the target can be considered as the basis of feature extraction. Suppose there is a sinusoidal component with frequency f_0 in the signal $s(t)$. The N observation samples are

$$y_i(t) = s(t \mid f_0) + n_i \ (i = 1, \ ..., \ N) \tag{6.106}$$

where n_i represents the background noise. We have shown that the maximum likelihood estimate for f_0 is

$$\hat{f}_0 = s^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) \tag{6.107}$$

That means the average algorithm can result in the optimum estimation for frequency f_0. Because DFT is a kind of average operation, therefore the segment DFT for a signal can obtain the necessary estimates. Suppose there are J components which are related to the frequency features. Then all features consist of J dimension space, denoted by $\boldsymbol{R}_J$. Since the maximum likelihood estimation is an unbiased estimate, therefore the difference in the corresponding value of representative vector components can be considered as the definition of distance in space $\boldsymbol{R}_J$.

Assume $\boldsymbol{a} \in \boldsymbol{R}_J$, $\boldsymbol{b} \in \boldsymbol{R}_J$ and $\boldsymbol{a} = (a_1, \ ..., \ a_J)^T$, $\boldsymbol{b} = (b_1, \ ..., \ b_J)^T$, let

$$\rho(\boldsymbol{a}, \ \boldsymbol{b}) = \sum_{i=1}^{J} w_i \mid a_i - b_i \mid \tag{6.108}$$

where w_i are the set of weighting coefficients and $\sum_{i=1}^{J} w_i = 1, \ w_i \geq 0 \ (i = 1, \ ..., \ J)$. It is easy to prove that $\rho(\boldsymbol{a}, \ \boldsymbol{b})$ satisfy the three axioms of metric space. For any given signal, divide it into M segments. Each segment is of length N and is expressed as $x_m(0), \ ..., \ x_m(N-1) \ (m = 0, \ ..., \ M)$. DFT of m-th segment data is

$$X_m(l) = \sum_{k=0}^{N-1} x_m(k) \exp(-2\pi \mathrm{j} kl / N) \tag{6.109}$$

The average amplitude is

$$X(l) = \frac{1}{M}\sum_{m=1}^{M} \mid X_m(l) \mid \ (l = 0, \ ..., \ N-1) \tag{6.110}$$

Based on the dataset $\{ X(l) : \ l = 0, \ ..., \ N-1 \}$, we define six parameters and establish the feature vector:

(1) Strongest line spectrum position x_1;
(2) Average line spectrum distance x_2;
(3) Average line spectrum strength x_3;
(4) Number of line spectrum x_4;
(5) Maximum frequency band position x_5;
(6) Slope rate change of maximum frequency band x_6.

The exact quantitative definition of x_i, ($i = 1, \ldots, 6$) is as follows.

Suppose

$$A = \frac{2}{N} \sum_{l=0}^{N/2-1} X(l)$$

It is called the global average of the line spectrum. The line spectrum set is defined as $V = \{X(l) \mid X(l) \geq A + B_1\}$, where B_1 is a parameter to be determined.

The elements in set V are rearranged as $v_1, \ldots, v_p$ in the order of the frequency value from small to large. We have

$$x_1 = \max\{v_i \mid i = 1, \ldots, p\} \tag{6.111}$$

$$x_4 = p \tag{6.112}$$

Assume $X(l_i) = v_i \ (i = 1, \ldots, p)$, then

$$x_2 = \frac{l_p - l_1}{p - 1} \tag{6.113}$$

Select a positive integer B_2 and construct the local average

$$A_i = \frac{1}{2B_2} \sum_{k=1}^{B_2} [X(l_i - k) + X(l_i + k)]$$

The strength of i-th line spectrum is $Q_i = X(l_i) - A_i = v_i - A_i$, therefore

$$x_3 = \frac{1}{p} \sum_{i=1}^{p} Q_i \tag{6.114}$$

Select a positive integer B_3 and let

$$C_l = \frac{1}{B_3} \left[X(l) + \sum_{k=1}^{B_3} (X(l-k) + X(l+k)) \right]$$

Assume $C_{l_0} = \max\{C_l \mid l = 0, \ldots, N/2 - 1\}$, then

$$x_5 = l_0 \tag{6.115}$$

Select a positive integer B_4 such that

$$x_6 = \frac{1}{B_4}(C_{l_0} - C_{l_0+B_4}) \tag{6.116}$$

The expert system designed by the above algorithm is shown in Fig. 6.52.

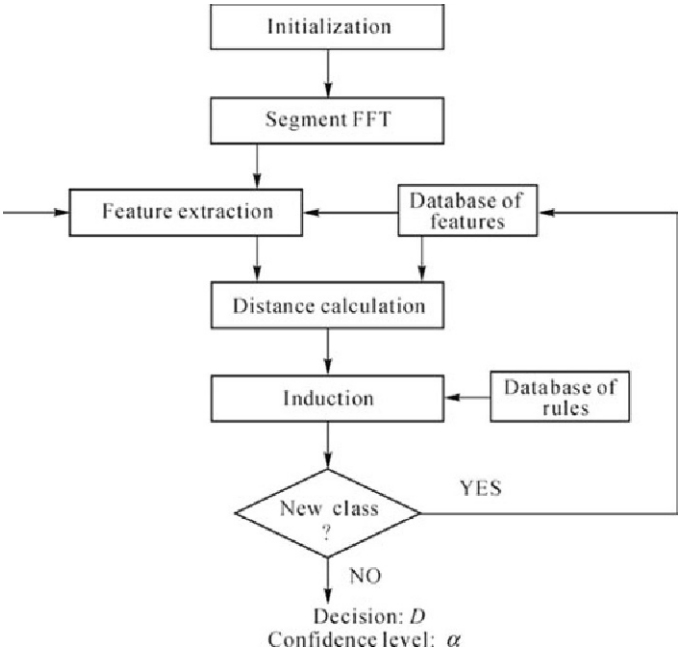


Fig. 6.52. Simplified expert system based on minimum neighborhood criteria

This system is a little different from the conventional expert systems; it allows a new type of class to be added into the database, i.e., if a new type of signal has a very strong line spectrum and we cannot assign it to any existing classes, then we can add this new type of signal to the database and extend the feature space to a new level. Therefore the expert system has a preliminary self-learning function.

Fig. 6.53 illustrates the operation procedure of an expert system. From this figure we can see that the feature extraction process is actually a dimension reduction process. In fact we reduce the 512 frequency bins to only 6 frequency bins. The 6 bins frequency feature is substantially representative of signal characteristics.

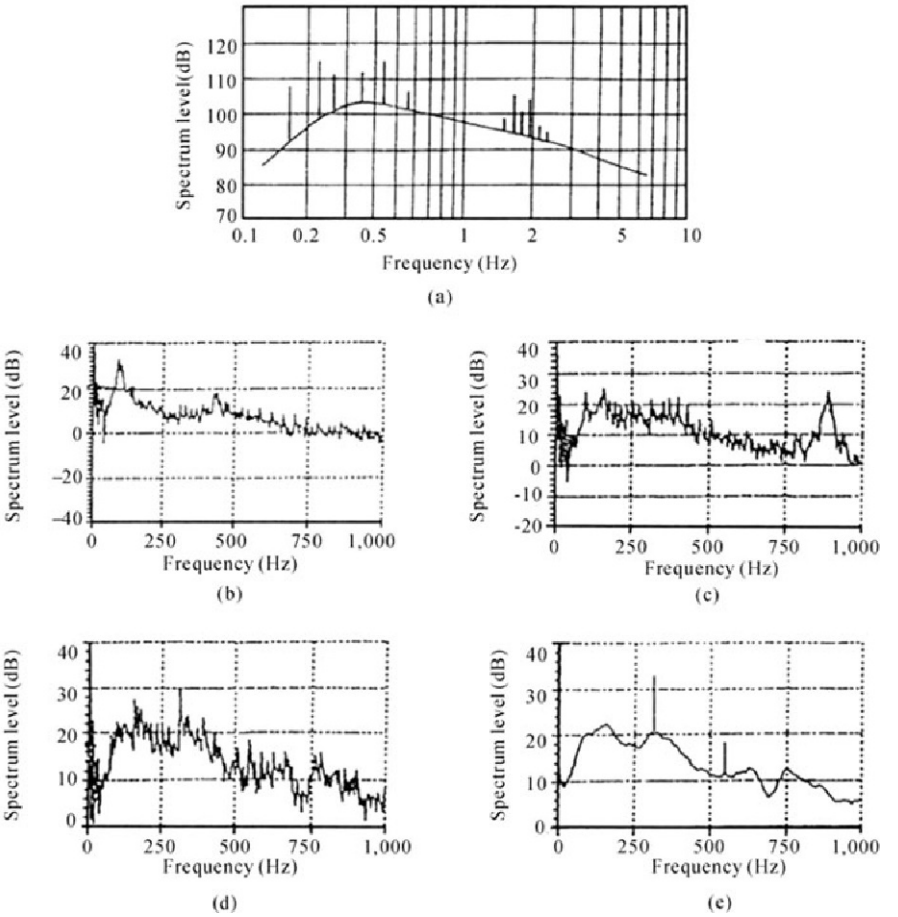


Fig. 6.53. Illustration example of feature extraction of radiated noise. (a) Spectrum of radiated noise; (b) Spectrum of radiated noise of cargo; (c) Spectrum of radiated noise of fishing boat; (d) Spectrum of radiated noise of submarine; (e) Feature extraction of radiated noise of submarine

6.7 Active Sonar Signal Capture and Analysis

The problem of signal capture or interception of active sonar is a relatively new topic in sonar design, although the signal capture problem belongs to the field of signal detection. But the method used in radiated noise detection may not be directly used in signal capture. This is because the active sonar signal has its own characteristics: the appearance time is random, the frequency is unknown, the pulse width is unknown, etc. Therefore, it is necessary to develop a new technique to capture an active sonar signal.

In this section, the time / frequency domain combined capture technique is introduced. This technique focuses on the long distance signal capture, therefore a long linear array is necessary. Some interception sonar in a submarine uses a single hydrophone to capture active sonar. It, of course, uses the technique described here, but the bearing function is limited due to the array aperture.

6.7.1 Theoretical Basis of Active Sonar Signal Capture

The frequency range of modern sonar is about 1 – 10 kHz. There exists the trend for the frequency of active sonar to move toward a lower frequency, although low frequency sonar needs a larger transducer. But the modern stealth submarine is forcing the utilization of low frequency.

The types of active sonar signals are various, but they have the same form: a narrow band signal. The signal to be transmitted may be of an amplitude modulated single frequency, linear frequency modulated, M-series, etc. The bandwidth is often of several hundred Hz. Therefore a narrowband receiver is necessary.

First of all, we are going to analyze how far an active sonar signal can be propagated. Suppose the signal to be transmitted is sinusoidal with frequency 4 kHz and the sound intensity is 1 kW (SL=200.5 dB). Assume the directivity index is 20 dB, then the source level is 200 dB. We can roughly estimate the propagation distance of this signal. Assume the spectrum level of background noise at 4 kHz is NL = 61 dB. If the bandwidth of the receiver is 100 Hz, then the background noise is 61+20 = 80 (dB). For the signal capture, we take DT = 10 dB, therefore FOM = 220−81−10 = 129 (dB). By using Eqs. (4.87) and (4.90) it is easy to find that the propagation distance is about 50 km.

From this analysis, we can know that signal capture for a long distance active sonar signal is possible. The single hydrophone can complete this mission.

If we also concern the incidental angle of the signal, then we should know what kind of array aperture is necessary. Table 6.4 shows the relationship between time delay differences and the distance. The results are calculated from Eq. (6.86) for $\varphi = 0°$.

Table 6.4 The relationship between distance and time delay

Source distance (km)	Array aperture d (m)	Time delay * (μs)
40	40	26
	100	166
10	40	106
	100	667

*: $\tau_{12} - \tau_{23}$

From this table we know that if the array aperture is d = 40 m, then the time accuracy in signal capture should be in the order of 1 μs. If the array aperture is

d = 100 m, then the time accuracy in signal capture should be in the order of 5 μs. Therefore the long line array has an obvious advantage in signal capture.

In frequency domain, we know the FFT analysis is an effective method for extracting sinusoidal and narrow band signals. From Fig. 5.26, we already know the long integration time will have enough gain to extract a sinusoidal signal. For a 300 Hz single frequency signal and sampling frequency 5 kHz, the 1,024-point DFT has a frequency resolution of 4.9 Hz. In this case, if the SNR is –20 dB, it is difficult to detect. But if we use a 16×1,024-point DFT, it will be easy to detect. Therefore the FFT algorithm is a good choice for capturing the active sonar signal in frequency domain.

6.7.2 A Combined Time/Frequency Domain Signal Capture Technique

Some examples of an active sonar signal are illustrated in Fig. 6.54. In Fig. 6.54(a), in the noise background, a sinusoidal signal of frequency f_0 = 1 kHz random appears, the impulse width is 10 ms and $(SNR)_{in}$ = 10 dB. In Fig. 6.54(b), f_0 = 2 kHz, the impulse width is 7.5 ms and $(SNR)_{in}$ = 6 dB. In both situations, the sampling frequency is f_s = 40 kHz.

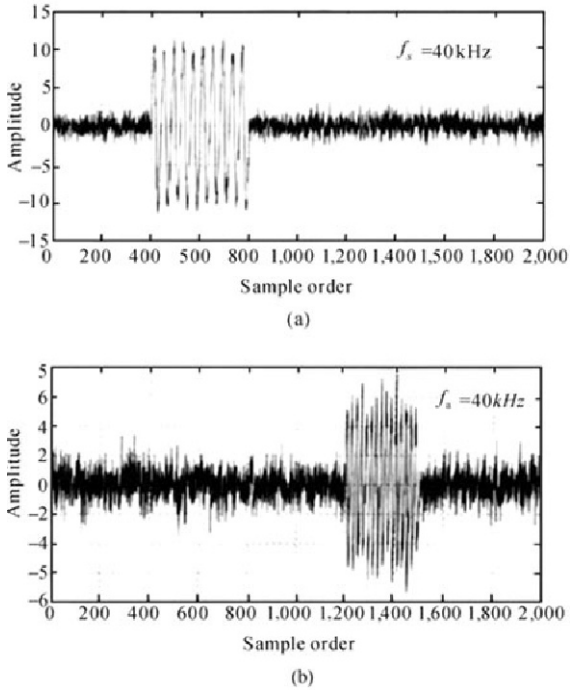


Fig. 6.54. Examples of active sonar signal. (a) Impulse width 10 ms, f_0 = 1 kHz, $(SNR)_{in}$ = 10 dB; (b) Impulse width 7.5 ms, f_0 = 2 kHz, $(SNR)_{in}$ = 6 dB

The signal capture is a quite difficult task. It is necessary to have enough SNR. And also the capture system should operate all the time, since the operator does not know where and when the active signal will appear. Therefore, the signal capture will operate completely automatically.

Fig. 6.55 shows the basic technique of signal capturing. Suppose there are three omni-directional hydrophones and each hydrophone has an independent data record and analysis system. This technique is based on the principle of two out of three, i.e., when at least two hydrophones capture the signal, the system will start to enter the time / frequency signal capture process, otherwise the system will be in stand-by status.

According to the capture requirements, it is necessary to detect the arriving signal in a wider frequency band. Therefore it is impossible to record an active sonar signal once it appears. The entire signal capture process divides into two stages: time domain capture and frequency domain capture.

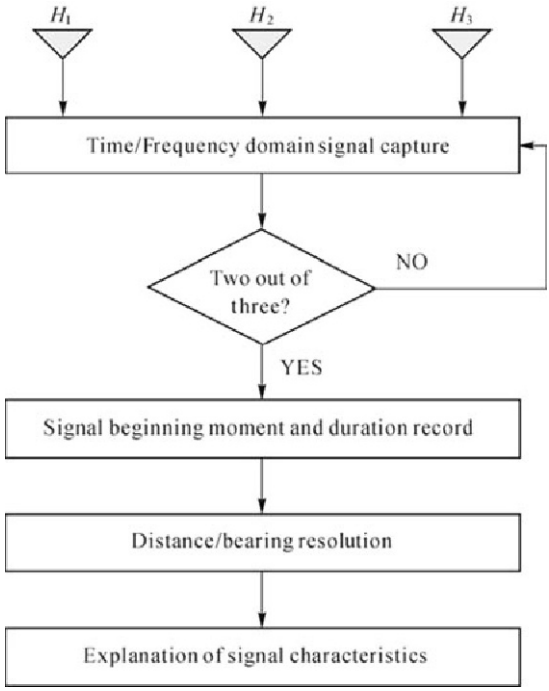


Fig. 6.55. Block diagram of active sonar signal capture

In order to have a high time resolution, it is necessary to adopt a relatively high sampling frequency. Suppose f_s = 125 kHz (T_s = 8 μs), take 1,024-point DFT, the frequency resolution is Δf = 125 kHz / 1,024 = 122 Hz. If in some frequency bins, for example, the 15th line spectrum, i.e., 14×122 Hz = 1,708 Hz. With a continuously appearing abnormally high amplitude value, we should store the received signal in a special memory, also remember the time of the starting point

and then turn to time domain capture.

The time domain capture is not for all frequency bins, but focuses on 1,708 Hz. An FIR filter will obtain a narrow band signal with a much higher SNR than the original input signal. The starting point and ending point of time domain capture determine the width of impulse. And then it is possible to take frequency domain analysis again for this segment signal captured in limited time duration. Therefore the signal feature is calculated by frequency analysis. Based on the time delay from three hydrophone signals, it is possible to estimate the incidental angle and source distance. A brief block diagram is shown in Fig. 6.56.

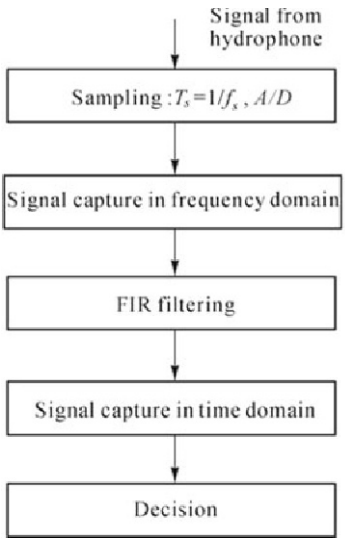


Fig. 6.56. Flow chart of signal capture in time / frequency domain

Now we are going to calculate the system gain of time / frequency domain capture.

For simplification, suppose the active sonar signal is a sinusoidal wave, therefore the received signal is

$$x(t) = A\cos(2\pi f_0 t + \varphi) + n(t) \tag{6.117}$$

where φ is a random phase with uniform distribution in $[0,\ 2\pi]$. The sample series of $x(t)$ is $x(0),\ ...,\ x(N-1)$, where N is the length of the sample series. The DFT of $x(k)$ $(k = 0,\ ...,\ N-1)$ is

$$X(l) = \sum_{k=0}^{N-1} x(k)\exp(-2\pi \mathrm{j}kl / N)\ (l = 0,\ ...,\ N-1) \tag{6.118}$$

The frequency resolution is

$$\Delta f = f_s / N \tag{6.119}$$

The input signal to noise ratio is

$$(SNR)_{\text{in}} = A^2 / (2\sigma_n^2) \tag{6.120}$$

where $\sigma_n^2 = E[n^2(t)]$.

In the neighborhood of frequency f_0, the width of the frequency bin is Δf, therefore the noise power in this frequency bin is

$$D_n = \frac{\sigma_n^2}{f_s / 2} \times \Delta f$$

Therefore the gain of frequency analysis is

$$G = \frac{\sigma_n^2}{D_n} = N / 2 \tag{6.121}$$

for $N = 1,024$, $G = 512$ (27 dB).

For large N, we can divide it into M segments, mutual non-intersection. Thus we increase the gain and simultaneously keep constant the width of the frequency bins. The input data is $x(0)$, $x(1)$, ..., $x(k)$, ... The m-th segment data is $x[(m-1)N + k]$ $(k = 0, \ ..., \ N-1)$. Take DFT for m-th segment data, we obtain

$$X_m(l) = \sum_{k=0}^{N-1} x[(m-1)N + k] \exp(-2\pi \mathrm{j} kl / N) \ (l = 0, \ ..., \ N-1)$$

The summation of $|X_m(l)|^2$ is

$$|X(l)|^2 = \frac{1}{M} \sum_{m=1}^{M} |X_m(l)|^2 \quad (l = 0, \ ..., \ N-1) \tag{6.122}$$

Since the noise and signal in $x(t)$ are independent, the noise in a different segment is also independent; therefore the gain can further increase. We already see this fact in Fig. 5.26. Fig. 6.57 is the flow chart of active sonar signal capture. It is necessary to have three memories for each channel signal. The data samples have some kind of overlap, the fraction of overlap can be 1/2 – 3/4. The reason for overlap sampling is that the signal to be captured may be a narrow impulse. If we do not use overlap sampling, we may lose too many useful samples. The three memories are indicated in Fig. 6.57 by RG_1, RG_2, and RG_3.

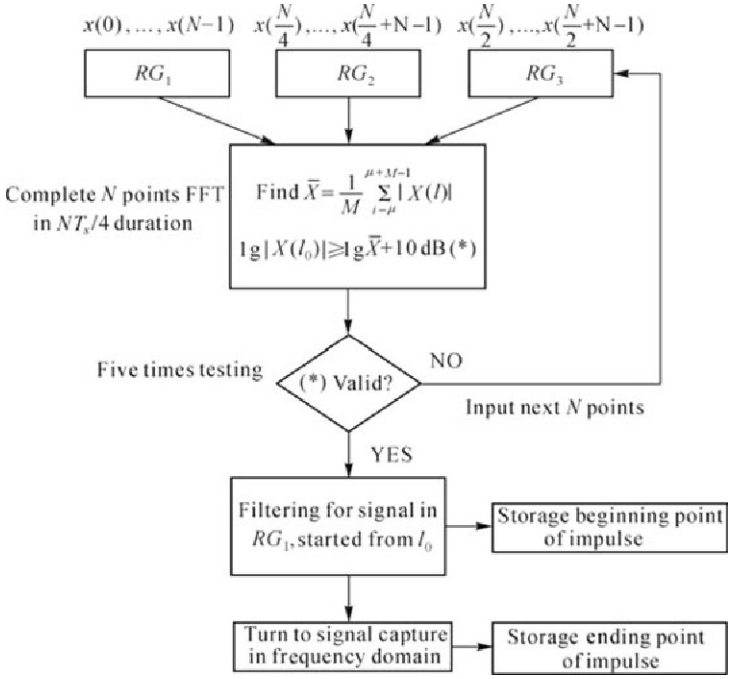


Fig. 6.57. Subroutine of active sonar signal capture in frequency domain

Example

Suppose the sampling duration of a signal capture system is $T_s = 8$ μs . Then the data length of 1,024 points will take 8.192 ms. If we use the non-overlap sampling technique in frequency analysis, then three segments will take 24.576 ms for a short impulse. This may result in signal loss in the capture process. But if we perform 1 / 2 overlap, three segments will take 12.286 ms and the situation will be much better than before.

In Fig. 6.57, the capture threshold is set to 10 dB. If some line spectrum, for example l_0, is continuously 5 times greater than 10 dB, then the value of l_0 is stored and also becomes the guide signal for the FIR filter. At the same time, the starting point of the impulse is recorded and we turn to time domain capture. Once we turn to time domain capture, because of the relatively high SNR, it is possible to record the width of the impulse. The flow chart of time domain capture is shown in Fig. 6.58.

A very important problem in time domain capture is the determination of the impulse rear margin. We can also use the 5 times threshold algorithm. In the case of failure to capture the signal in time domain, we should turn to the capture process in frequency domain again.

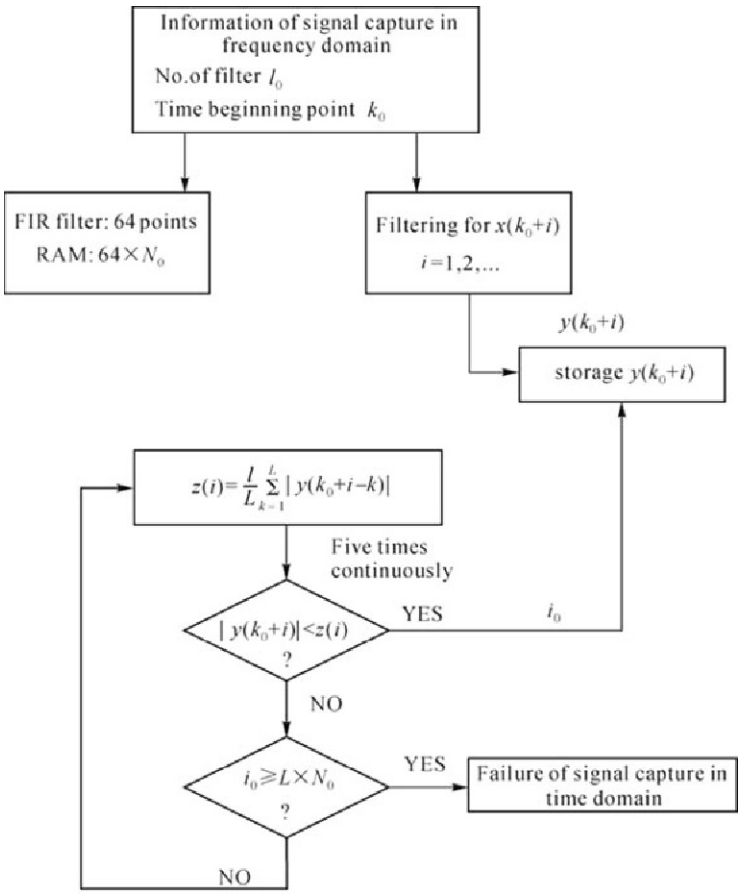


Fig. 6.58. Flow chart of subroutine of signal capture in time domain

6.7.3 Signal Feature Extraction of Active Sonar Signal

We are going to briefly discuss the signal feature in the frequency domain. When the active sonar signal is captured, generally speaking, there are at least several tens of milliseconds of signals that can be used. It is enough for frequency domain analysis.

(1) Single rectangle wave

$$x(t) = \text{rect}(\frac{t}{T}) = \begin{cases} 1 & 0 \leq |t| \leq T/2 \\ 0 & \text{otherwise} \end{cases} \tag{6.123}$$

The spectrum of $x(t)$ is

$$X(f) = T \times \frac{\sin(\pi fT)}{\pi fT} \tag{6.124}$$

The time wave and spectrum are shown in Fig. 6.59.

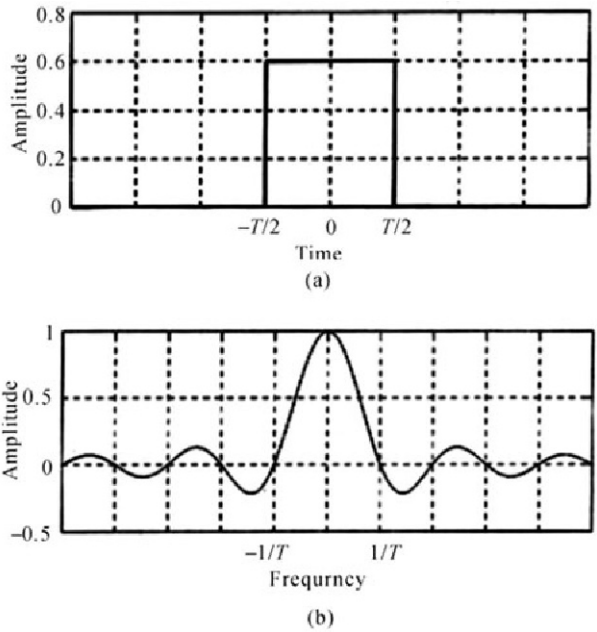


Fig. 6.59. Single rectangle impulse and its spectrum. (a) Waveform; (b) Spectrum

(2) Amplitude modulation signal

$$x(t) = \text{rect}(t / T)\cos(2\pi f_0 t) \tag{6.125}$$

$$X(f) = \frac{T}{2}\left[\frac{\sin(\pi(f - f_0)T)}{\pi(f - f_0)T} + \frac{\sin(\pi(f + f_0)T)}{\pi(f + f_0)T}\right] \tag{6.126}$$

The time wave and its spectrum are shown in Fig. 6.60. This is the most frequently used signal form in active sonar.

(3) Gaussian bell shape impulse

$$x(t) = A\exp(-(t / \alpha)^2 / 2) \tag{6.127}$$

$$X(f) = \sqrt{2\pi}\alpha A\exp(-2(\pi f)^2) \tag{6.128}$$

(4) Linear frequency modulation signal

$$x(t) = \text{rect}(t / T)\cos[2\pi(f_0 t + \mu t^2 / 2)] \tag{6.129}$$

where μ is the modulation coefficient. The spectrum of linear frequency modulation has no closed form expression; it can be expressed by Frenel integration.

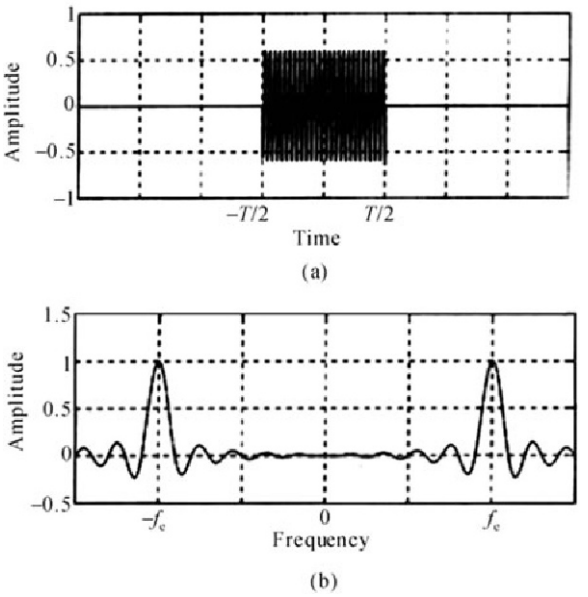


Fig. 6.60. Single rectangle impulse modulation and its spectrum. (a) Rectangle impulse modulation; (b) Spectrum

6.8 Underwater Acoustic Communication and Adaptive Equalization of Underwater Acoustic Channel

Acoustic communication sonar is one of extensive application of sonar. It has the functions of both a passive and active sonar system. But it is only used to communicate with a friendly user and not in attempting to intercept any hostile sonar operation. The utilization of acoustic communication sonar is also limited by the real environment, since any signal transmitted will result in the dangerous consequences of self-exposure.

Because the objective of acoustic communication is cooperation, therefore it can sufficiently use the acoustic channel characteristics; it can measure the performance of the acoustic channel before communication; it can use the favorite signal form in transmission.

Acoustic communication theory is a part of modern communication theory. Many techniques and algorithms in radio communication can be transplanted to

underwater acoustic communication, including coding techniques, spread frequency communication techniques and coherence detection. But it is worth showing that the media of underwater acoustic communication is completely different from radio waves. Some well-established techniques in wireless communication are not suitable for underwater acoustic communication.

This section will briefly introduce the basic theory and some new topics in underwater acoustic communication[74-76, 78-81].

6.8.1 Introduction of Underwater Acoustic Communication

Underwater acoustic sonar was developed with the birth of the submarine. Military requirements were the main motivation for acoustic communication sonar. In early times, in World War I, data transmission used the Morse code and the transmission data rate was very low. But the framework of underwater communication is the basis of modern communication sonar. With the development and exploration of ocean engineering, underwater acoustic communication interested many scientists and engineers.

Underwater communication sonar has been extensively used in ocean environment monitoring and prevention of disasters, deep sea oil / gas exploration, use of underwater robots. Some applications are not limited to point-to-point communication but are networks. The underwater GPS (global position system), studied in some countries, is a combination of wireless communication and underwater acoustics.

Fig. 6.61 shows the basic model of underwater acoustic communication. The information source contains the message to be sent to the receiver. It could be text, sound and image. The transmitter must convert the messages to an electrical signal and then transmit the electrical signal. Then the electrical signal has to be converted to an acoustic signal. An underwater channel is the media for transmitting acoustic signals. Usually, in the acoustic channel there exist various kinds of interference. The receiver, at the other end, should convert the acoustic signal to an electrical signal and extract the useful message, which is included in the interference in the received signal. The entire procedure is quite complicated. At all stages there are theoretical and practical examples.

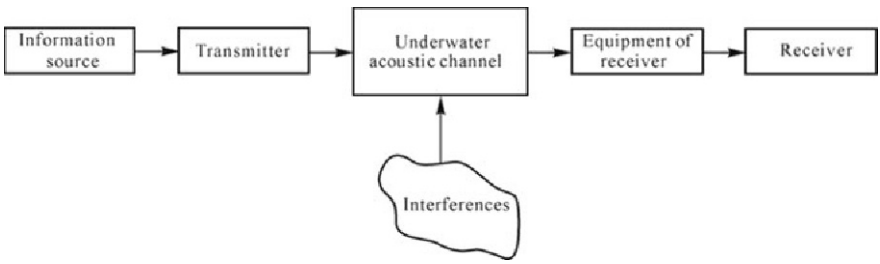


Fig. 6.61. Basic scheme of underwater acoustic communication

Fig. 6.62 shows the block diagram of the model of a digital underwater acoustic communication system. The main characteristic of this system is that it is digital. From coding in the transmitter to decoding in the receiver, we are using digital techniques.

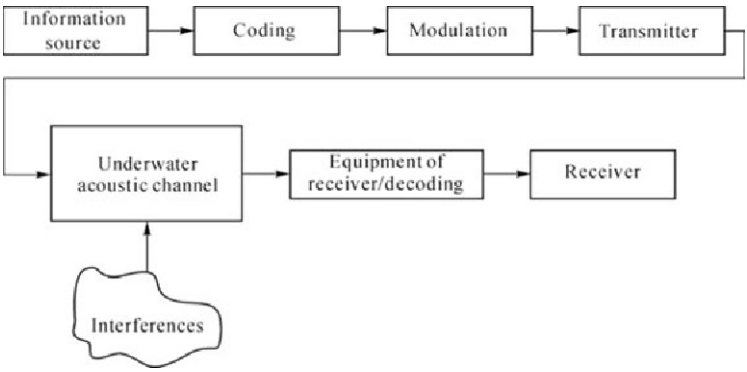


Fig. 6.62. Basic scheme of digital underwater acoustic communication

The transmission rate of digital communication is subject to the information source and communication distance. For the Morse code, the transmission rate is about 1 – 10 bps; for the human voice it is about 2,000 – 4,000 bps. At this transmission rate, a compressed voice signal can be transmitted underwater and can be recovered. As for the image signal, it is necessary to have more than a 1 Mbps transmission rate, even if the data is compressed.

Traditionally, the communication distance is categorized as near, middle and long distance (Table 6.5).

Table 6.5 Category of underwater acoustic communication

Item	Distance (km)	Frequency (kHz)
Very long distance	$\geqslant$1,000	$\leqslant$1
Long distance	10 – 100	Several
Middle distance	1 – 10	10
Short distance	0.01 – 1	100
Very short distance	$\leqslant$0.1	Several

Based on a large amount of experiments at sea, Baggeroer summarized the experimental results. An interesting result is shown in Fig. 6.63. It concludes that[80]

$$R(\text{Distance}) \times R(\text{Transmission rate}) \leq 40 \quad (\text{km·kbps}) \tag{6.129}$$

In this expression, the first R means the distance with unit km, and the second R represents the transmission rate with unit kbps. For example, in the 1990's, the Advanced Technology Demonstration project, sponsored by the US Office of Naval Research (ONR), used a transmission rate for FAX communication between surface ship and submarine of 0.4 kbps at a distance of 37 – 124 km.

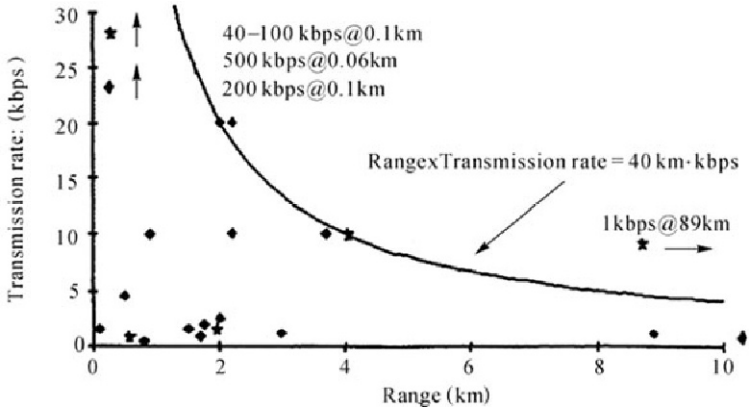


Fig. 6.63. Experimental results of underwater acoustic communication

There are some frequency division standards for underwater acoustic communication. Here is an example proposed by NATO (North Atlantic Treaty Organization):
(1) Beacon: 10 kHz
(2) Submerged bell: 37.5 kHz
(3) Commercial submerged vessel: 25 kHz
(4) Submarine: 42 kHz
(5) Diver: 40 kHz.

6.8.2 Coding and Decoding

The binary code cannot be directly transmitted underwater, since the front margin and rear margin of the binary code are both very sharp and will occupy a very wide frequency band. Therefore it is necessary to use a modulation technique. In an underwater acoustic channel, since there exist serious fluctuations and the effect of multiple path phenomena, the interference between codes may involve several or even several tens of symbols. Also, there may be a Doppler shift and therefore signal reception and recovery become more difficult.

In order to overcome the possible noise and interference in signal transmission, it is necessary to add some marker, or leave some blank segment, such that it is easy to identify the received message. The increase in redundancy will decrease the error code rate and increase the error correcting ability.

The coding technique is necessary not only for conventional underwater acoustic communication, but also for classified communication. Some code rearranging techniques or classified key code techniques are extensively used in underwater acoustic communication. Decoding is the inverse procedure to coding. In a general case, when we receive the transmitted signal, we have to first find the marker code or information code and then decode the receiving signal.

Now we are going to introduce some examples of modulation code. Fig. 6.64

is the ASK (amplitude shift keying) code. This code is also called the on / off keying code. It controls the amplitude of the carrier wave by the single polar non zero code. When the code unit equals one, the carrier wave is transmitted. When the code unit is equal to zero, the carrier wave is not transmitted.

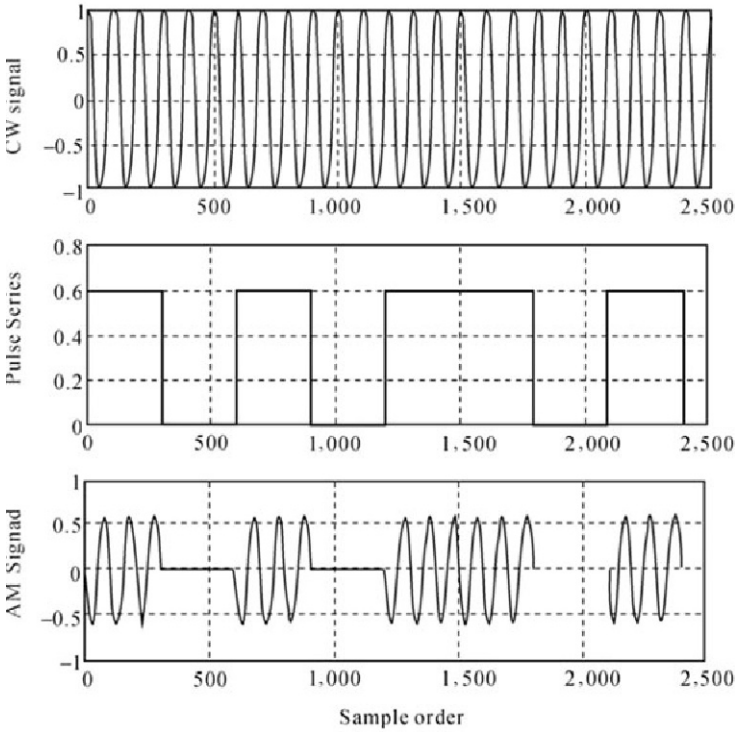


Fig. 6.64. Modulation of ASK. From top: carrier, envelope, modulation signal

If we use the rectangle function to express a single impulse

$$g(t) = \text{rect}[(t + T/2)/T] = \begin{cases} 1 & 0 \le t \le T \\ 0 & \text{otherwise} \end{cases} \tag{6.130}$$

Then the mathematical expression of the ASK code is

$$x(t) = \sum_{k=-\infty}^{\infty} a_k g(t - kT)\cos(2\pi f_0 t + \varphi) \tag{6.131}$$

The value of a_k is 1 or 0 and φ is the random phase, uniformly distributed in $[0,\ 2\pi]$. The anti-interference ability of this kind of modulation is relatively low and the utilization of effective power is limited. Therefore it is only applied in the case of a low transmission rate.

Fig. 6.65 illustrates the FSK (frequency shift keying) modulation code. When the code value is 1 then the carrier frequency is f_0 and when the code value is 0 the carrier frequency is f_1. Compared with the ASK code, this kind has a higher power utilization ratio.

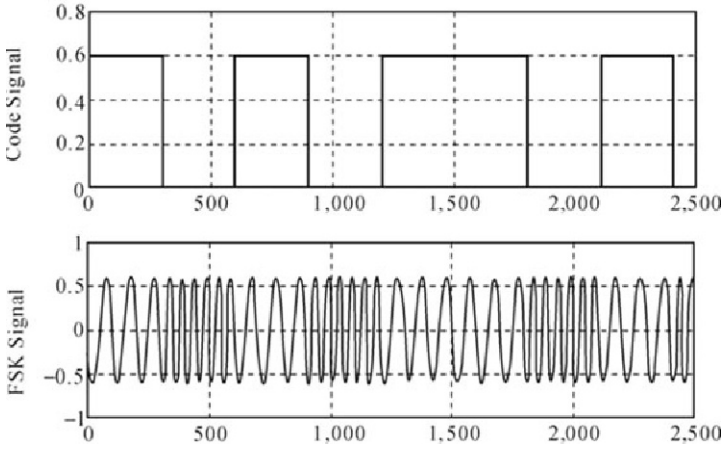


Fig. 6.65. Modulation of FSK coding. From top: envelope, modulation signal

Fig. 6.66 shows an example of PSK (phase shift keying) modulation code. Suppose the code value changes from 1 to 0, the phase will change 180°.

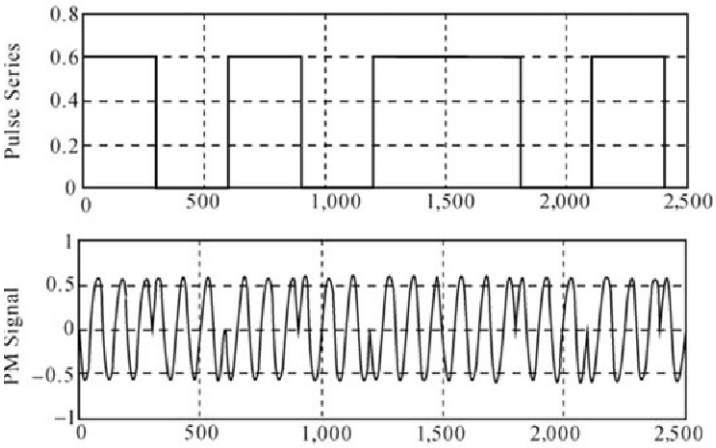


Fig. 6.66. Modulation of PSK coding. From top : envelope, modulation signal

In the ideal underwater acoustic channel (i.e., the channel that only has amplitude attenuation), PSK has a higher transmission rate, and therefore it is

usually used in a vertically direction transmission in deep water, where the vertical acoustic channel can be approximately considered as an ideal channel.

In the decoding process, it is necessary to utilize the coding information. The very important thing is to find the time delay value of the coding signal and then to find the frame synchronize code. The frame correlation technique is useful in finding the peak value of the frame signal and reference signal.

The conventional time delay estimation algorithm can be used to find the time delay of the coding signal. For the ideal channel, the matched filter of frequency modulation is recommended.

Assume the input signal of the receiver is

$$x(t)=\begin{cases} A\sin[2\pi f_0(t-\tau_0)+\mu(t-\tau_0)^2/2] & \tau_0 \le t \le \tau_0+T \\ 0 & \text{otherwise} \end{cases}$$

The reference signal is

$$r(t)=\begin{cases} A_0\sin[2\pi f_0 t+\mu t^2/2] & 0 \le t \le T \\ 0 & \text{otherwise} \end{cases}$$

It is straightforward to calculate that the output of the matched filter is

$$R(\tau)=\begin{cases} \dfrac{2AA_0T}{\mu(\tau-\tau_0)}\times\sin\left[\dfrac{1}{2}\mu(\tau-\tau_0)\left(1-\dfrac{|\tau-\tau_0|}{T}\right)\right]\cos(2\pi f_0(\tau-\tau_0)) & |\tau-\tau_0|\le T \\ 0 & \text{otherwise} \end{cases}$$

for $\tau=\tau_0$, $R(\tau)$ reaches the maximum value. Subtracting the propagation time we obtain the code time delay value. Based on the peak value of the frame synchronizer, the decoding algorithm is performed in each frame by using a fast correlation technique. The reference code corresponding to the peak value of the correlation function is the signal to be decoded. Since all the code to be processed is the symbol 0 – 1, and the reference matrix actually is a matrix consisting of the symbol 0 – 1, we can adopt the Adama matrix transform to simplify the calculation.

6.8.3 Frequency Spread Technique

In Chapter 3, when we discussed the coding theorem and channel capacity in information theory, the Shannon basic theorem about channel capacity was expressed as (Eq. (3.64)): $C=W\log_2(1+S/N)$, where C represents the channel capacity or transmission ability and the unit is bps; W is the frequency band width with unit Hz and S/N represents the SNR. It is also denoted by σ_s^2/σ_n^2. From this equation, we can see a noise-free channel has infinite channel capacity. For a

channel with interference and noise, if we keep C as a constant, then a larger bandwidth will correspond to a lower signal-to-noise ratio. That means the communication system will benefit from frequency spread. That is the main reason for using the frequency spread technique.

In the communication channel, if there is white noise interference, then the gain from the frequency spread technique is limited. Suppose n_0 is the noise density, then the total power in the frequency band is $N = Wn_0$. It is proved that[107] $\lim_{W\to\infty} C = 1.44S / n_0$. This means that the channel capacity has a limit in this kind of communication channel. Therefore the effect of the frequency spread technique for the white noise interference channel is limited. But for the channel with narrow band noise, the frequency spread technique will be useful, as is shown in the Shannon theorem.

In underwater acoustic communication, the frequency spread technique has been successfully used[111-113]. The time width per code is compressed and repeatedly transmitted. The most popular frequency spread code is the M series. This is a quasi-random code driven by time-based impulse using a shift register. We can define the M series as the longest linear feedback series of a shift register. As we know the longest length of 0 – 1 series generated by an n stages shift register is $2^n - 1$, therefore an n stages m series with length $2^n - 1$. Fig. 6.67 shows the block diagram of a 4-stage M series generator. The linear feedback format is

$$a_4 = a_0 \oplus a_1 \tag{6.133}$$

where the symbol $\oplus$ represents the summation of module 2.

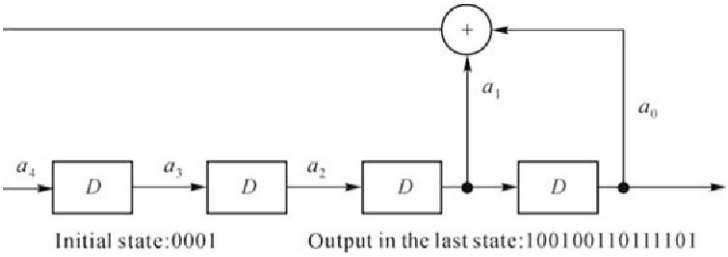


Fig. 6.67. Fourth order M series generator

Since the M series has the following excellent performance, it is extensively used in many application areas (Fig. 6.68).

(1) For any $n \geq 1$, there exists an n stages shift register, which can generate the M series.

(2) In any period of the M series there are 2^{n-1} of 1 and 2^{n-1} of 0.

(3) If we define the correlation function of the M series as $R(i) = A - B$, where

A = the number in $x(k)$ and $x(k+i)$ with the same symbol; B = the number in $x(k)$ and $x(k+i)$ with a different symbol; then

$$R(i) = \begin{cases} 2^{n-1} & i = 0 \\ -1 & \text{otherwise} \end{cases} \tag{6.134}$$

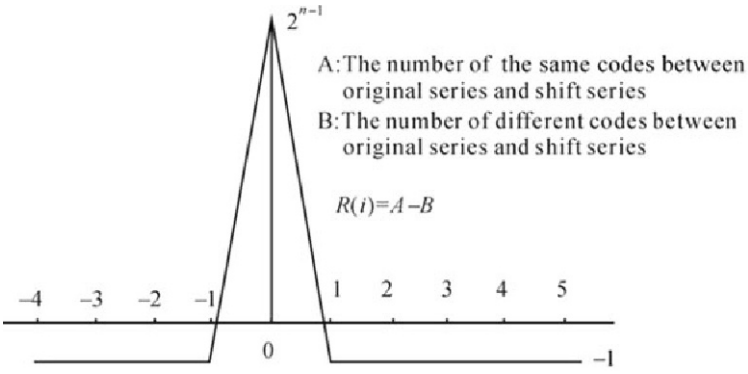


Fig. 6.68. Auto-correlation function of M series

Beside the M series, there are other kinds of code used in underwater acoustic communication, for example the Gold code and the Huffman code, etc.

6.8.4 *Equalization of Underwater Acoustic Channel*

The propagation performance of an underwater soundwave strongly depends on the characteristics of the underwater acoustic channel. If we can measure the performance of the underwater acoustic channel in real time, this will provide a very important basis for sonar design. In this field, underwater communication provides the most favorable conditions since both the transmitter and receiver are collaborative users of the underwater acoustic channel. Before establishing communication between the transmitter and the receiver, one can test the specification and performance of the underwater acoustic channel to be used. The most suitable test signal for a real ocean environment can be selected. And the channel match technique can be used to increase communication reliability and efficiency.

Because of the time / special varying characteristics of the underwater acoustic channel, research work on channel matching and equalization is at an early stage. Fig. 6.69 is the principal explanation of channel equalization.

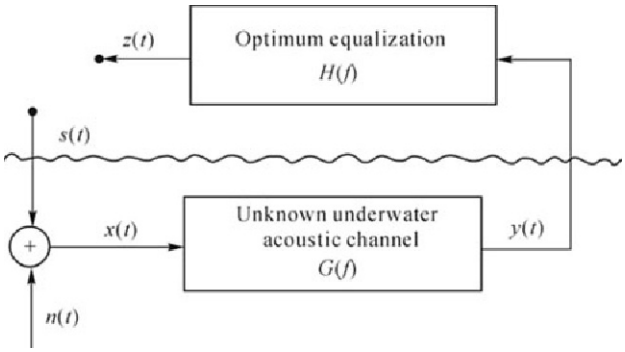


Fig. 6.69. Block diagram of equalization of underwater acoustic channel

Suppose the original transmitted signal is $s(t)$, the ambient noise $n(t)$ is added to this signal at the transmitted end. The signal is $x(t) = s(t) + n(t)$. The signal $x(t)$ propagated via the underwater acoustic channel $G(f)$ becomes $y(t)$ at the receiving end. This is a very ideal description of the underwater acoustic channel. The channel is considered as a linear system and $G(f)$ is the transfer function of this channel. In the real ocean environment, the channel is time varying and spatially varying and also has some kind of frequency spread and multi-path effect. The characteristics of the underwater acoustic channel may be quite complicated and cannot be described by simple $G(f)$. But studies of this simplified model will be useful in understanding a more complicated situation.

The goal of channel equalization is to find a filter $H(f)$, such that the output of this filter is $z(t) = s(t)$, when the input is $y(t)$. The role of equalization looks likely to inverse the effect of $G(f)$, which performs on $x(t)$. It is not an easy way to find the optimum equalization system. If the underwater acoustic channel is stationary and reliable, then it is possible to find the optimum transfer function and realize it.

Fig. 6.70 shows the block diagram of the adaptive channel equalization principle. The aim of this adaptive system is to measure the channel transfer function $G(f)$ by using a known signal $s(t)$. Once we find $G(f)$, the inverse of it can be used as the equalization transfer function. In Fig. 6.70, the error function $\varepsilon(t)$ is used to adjust the optimum filter. We have $\varepsilon(t) = z(t) - y(t)$. It is obvious that $H_{\text{opt}}(f) = G(f)$. In this case, $E[\varepsilon^2(t)]$ reach a minimum. When we obtain $H_{\text{opt}}(f)$, it is easy to calculate $1/G(f)$. Of course we should take care of the pole points in $G(f)$.

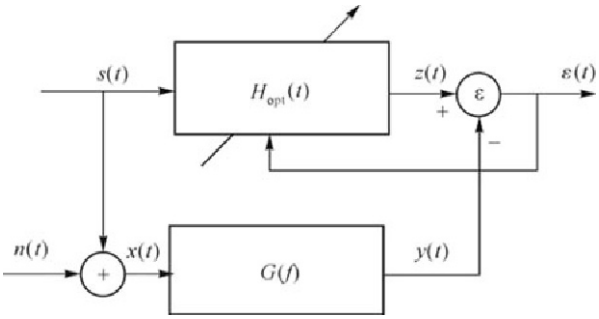


Fig. 6.70. Block diagram of adaptive equalization of underwater acoustic channel

In order to implement the optimum transfer function $H_{opt}(f)$, the tapped delay line can be a good choice. This is because it is easy to realize in hardware and software (see Fig. 6.71). The time delay duration can be chosen to equal the sampling duration, and the delay line should be long enough, so that the tapped delay line can approximate the optimum filter.

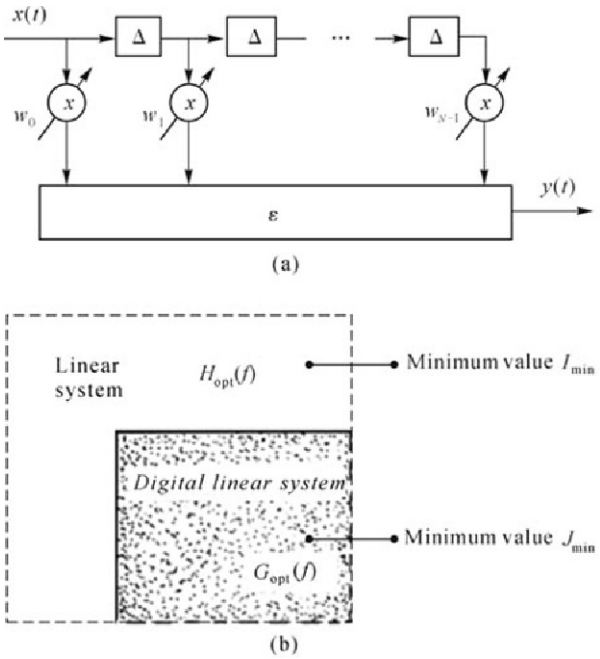


Fig. 6.71. Realization of adaptive filtering: tapped delay line

But it is necessary to know that the tapped delay line is only one kind of special linear system, and $H_{opt}(f)$ is the optimum transfer function for all linear

filters. If the minimum mean square value of $E[\varepsilon^2(t)]$ is $I_{\min}$ for all linear filters and the minimum mean square value of $E[\varepsilon^2(t)]$ in the condition of all digital tapped delay lines is $J_{\min}$, then $J_{\min} \geq I_{\min}$. We should select the right parameter such that the difference between $I_{\min}$ and $J_{\min}$ is as small as possible. It is proved that[123]

$$J_{\min} = I_{\min} + \int |H_{\text{opt}}(f) - G_{\text{opt}}(f)|^2 \, \mathrm{d}f \tag{6.135}$$

Fig. 6.72 illustrates the relationship between $I_{\min}$ and $J_{\min}$.

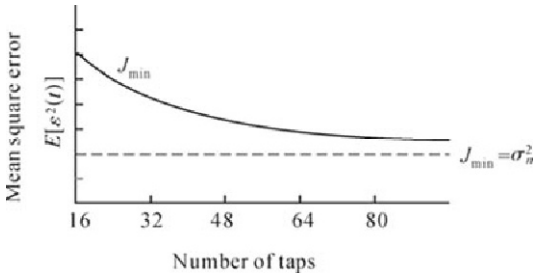


Fig. 6.72. Illustration of $I_{\min}$ and $J_{\min}$

6.9 Introduction of Multistatic Sonar

The traditional sonar system is based on monostatic design philosophy. The operation mode of this kind of sonar is based on the assumption that the transducer and the receiver of the sonar system are placed in the same position, i.e., the transmitting and receiving signals are in the same location. Therefore, for active sonar, the reverberation caused by the transmitting signal may become the main factor to affect the target detection performance.

Multistatic sonar has become a hot topic for the application of the sonobuoy in anti-submarine warfare[127-129]. It also appeared in the formation operation of surface ships. The sonar system on each surface ship should coordinate operations and share information received in a collaborative operation.

Generally speaking, one can think that only active sonar has the problems of a multistatic scenario, since purely passive multistatic problems can be categorized in the data fusion field and theoretical study results in multisensor or distributed network sensor areas can be applied in purely passive multistatic detection [130-135].

6.9.1 General Description of Multistatic Sonar Problem

Fig. 6.73 illustrates the detection scenario of monostatic sonar. This is a typical operation mode of active sonar. The source and the receiver are placed in the same position. Once the sonar transmits an active signal, the receiver is waiting to receive an echo signal, reflected from the target. The propagation path of the transmitting signal and echo signal are basically the same, except for some reverberation or scattering from the sea boundary or sea bottom.

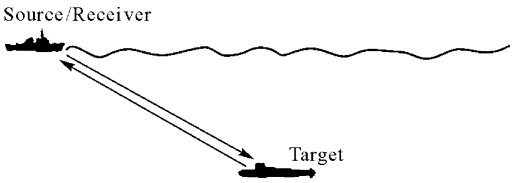


Fig. 6.73. Illustration of monostatic sonar

Figs. 6.74 and 6.75 is the detection scenario of bistatic sonar and multistatic sonar. In Fig. 6.74, for bistatic sonar, except for active sonar, there is also a passive sonar which is separately placed in another position. It receives the echo signal from the target and of course the propagation path is different from the original echo signal to the place of the active sonar. In Fig. 6.75, there are five sonobuoys, only one sonobuoy transmits a signal and the other four sonobuoys are waiting to receive an echo signal from the target.

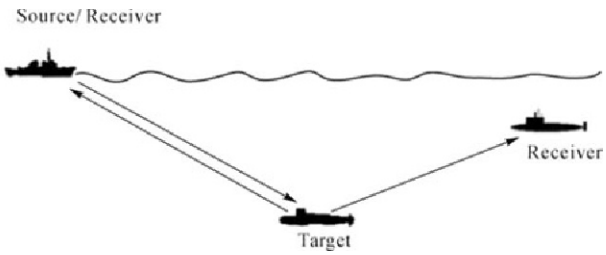


Fig. 6.74. Illustration of bistatic sonar

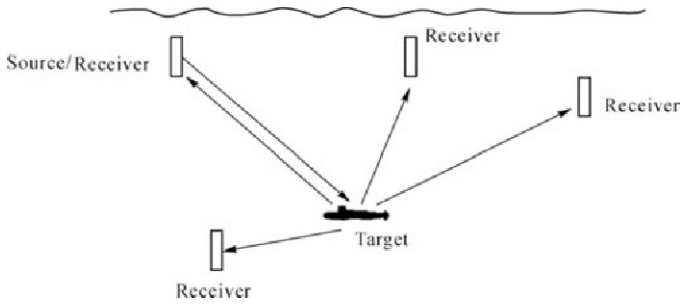


Fig. 6.75. Illustration of multistatic sonar

There are obvious advantages in using multistatic sonar in anti-submarine warfare.

(1) Wider range of sonar detection;
(2) An increase in the probability of target detection;
(3) An increase in the bearing accuracy in the case of multiple detections;
(4) More efficient utilization of detection resources of sonar systems.

The utilization of multistatic sonar requires more complicated methodologies to analyze the data coming from various receivers.

Fig. 6.76 shows how bistatic sonar can increase the detection range of a sonar system.In the case of monostatic sonar (Fig. 6.76 (a)), the detection range of active sonar is determined by the reverberation level and ambient noise level, where r_1 represents the detection range of active sonar, which works under reverberation limitation conditions, and r_2 represents the detection range of active sonar, which works under ambient noise limitations, usually $r_1 \leq r_2$.

In Fig. 6.76(b) there are two separate receivers. One is active / passive dual function sonar and the other is pure passive sonar. For the latter, reverberation is not a limitation factor for this sonar system. At most, the directly transmitted signal from active sonar can be considered as another target, which is obviously different from echo signals. Therefore, the detection range for this sonar should be further than that of the original active sonar, which is limited by the reverberation level.

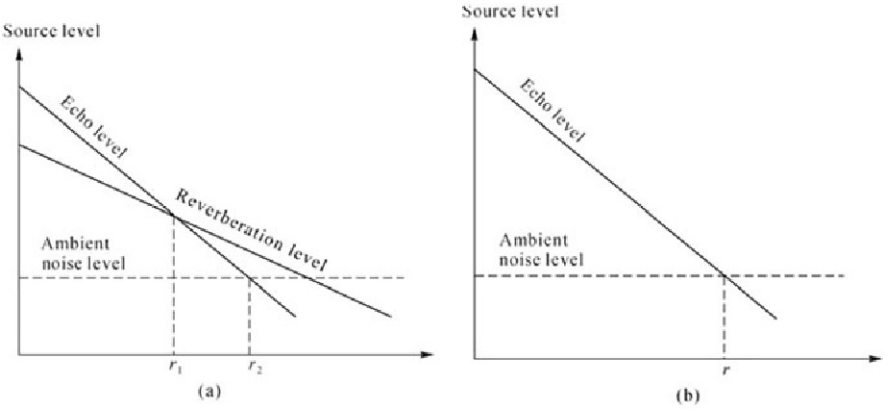


Fig. 6.76. (a) Ranging distance of monostatic sonar; (b) Ranging distance of multistatic sonar

6.9.2 *Basic Formula for Calculating Detection Range in Multistatic Sonar*

Fig. 6.77 illustrates the main scenario for calculating the ranging distance of multistatic sonar. Suppose S is the source location of active sonar, R is a separate receiver of passive sonar. An unknown target is located in position T. The distance

r_0 between S and R is known. For simplification, a straight line RK is chosen as calculating reference. For example, RK could be the axis line from aft to bow of a surface ship R. In this case the bearing angle of passive sonar in R can be easily used to calculate the target distance.

Define the transmitting moment as $t = 0$. The signal travel time from S to R is supposed to be $t_0 = SR / c = r_0 / c$, where c represents the sound velocity. The travel time from S to T and from T to R is respectively defined as $t_1 = ST / c = r_1 / c$, $t_2 = RT / c = r_2 / c$. It is necessary to show that we can only measure the value of $t_1 + t_2$, but not t_1 and t_2 separately.

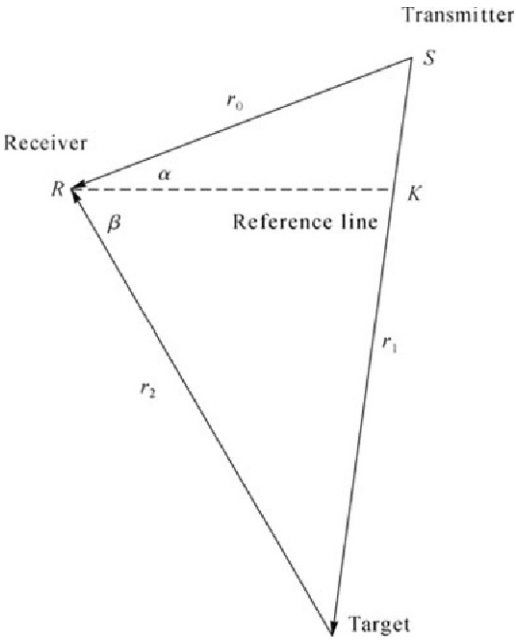


Fig. 6.77. Range calculation in multistatic sonar

The bearing angle for S and T is α and β respectively. According to the assumption that t_0 and $t_1 + t_2$ is known, we should find the algorithm which can express r_2 by α, β, t_0 and $t_1 + t_2$. It is easy to show that

$$r_1^2 = r_2^2 + r_0^2 - 2r_0 r_2 \cos(\alpha + \beta) \tag{6.136}$$

In order to express r_2 by α, β, t_0 and r_1+r_2, it is necessary to find the relationship between r_2 with α, β, r_0 and r_1+r_2. Since $r_1^2 = (r_1 + r_2)^2 - 2r_1 r_2 - r_2^2$,

we have

$$(r_1 + r_2)^2 - 2r_1r_2 - r_2^2 = r_2^2 + r_0^2 + 2r_0r_2\cos(\alpha + \beta) \tag{6.137}$$

Based on Eq. (6.137), it is not difficult to show that

$$\begin{aligned}(r_1 + r_2)^2 - r_0^2 &= 2r_1r_2 + 2r_2^2 - 2r_0r_2\cos(\alpha + \beta) \\ &= 2r_2[(r_1 + r_2) - 2r_0\cos(\alpha + \beta)]\end{aligned} \tag{6.138}$$

From Eq. (6.138), we obtain

$$r_2 = \frac{(r_1 + r_2)^2 - r_0^2}{2[(r_1 + r_2) - r_0\cos(\alpha + \beta)]} \tag{6.139}$$

It can be rewritten as

$$r_2 = \frac{(t_1 + t_2)^2 - t_0^2}{2[(t_1 + t_2) - t_0\cos(\alpha + \beta)]} \tag{6.140}$$

This is the formula we expected.

Suppose $t_0 = 3.5$ s, $t_1 + t_2 = 17.6$ s, $\alpha = 25^\circ$, $\beta = 30^\circ$ and c = 1,500 m / s. Substituting them into Eq. (6.140), then $r_2 = 14.33$ km.

6.9.3 Realization of Bistatic Sonar Function by Using Monostatic Sonar

Althongh the performance of bistatic sonar and multistatic sonar is sometimes much better than monostatic sonar, it is possible to take advantage of bistatic sonar by using monostatic sonar. That is to say, we can adopt some kind of special treatment such that the function of bistatic sonar can be realized by using monostatic sonar, and of course we have to pay some cost for this procedure.

Fig. 6.78 is an example of bistatic sonar funcition using monostatic sonar. Suppose R is a surface ship with passive sonar, and it can also launch distance. The shell explosion wave under the sea can propagate to R. The echo from target T is also received by R. If one can measure the incidental angle α of the explosive signal from S and the incidental angle β of the echo signal from T, then the target distance r_2 is solved by Eq. (6.139). But we have to find the relationship between r_0, r_1+r_2 with the explosive moment and travel time of the explosive wave and

echo signal. It is not difficult, when we choose a suitable reference point, to calculate the start moment.

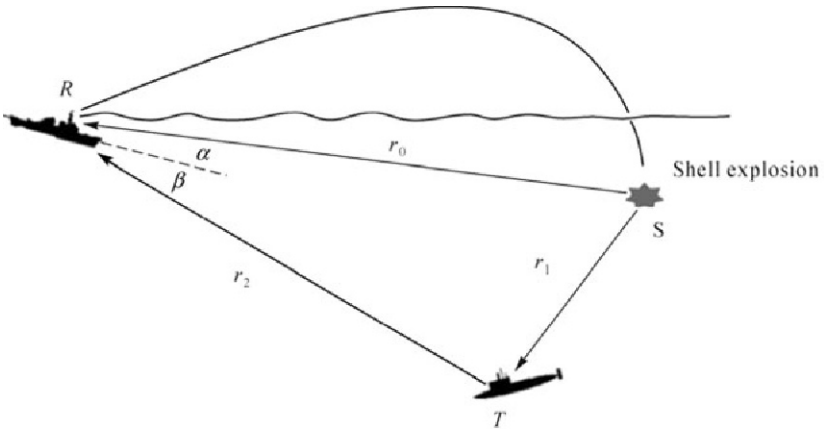


Fig. 6.78. Realization example of bistatic sonar function using monostatic sonar

References

[1] Anderson, V. C.: The first twenty years of acoustical signal processing. J. Acoust. Soc. Amer., 51, pp. 1062-1080 (1972)
[2] Stergiopoulos, S. (Eds.): Advanced Digital Signal Processing Handbook. CRC Press, Boca Raton (2001)
[3] Chen, C. T. (Eds.): The past, present and future of underwater acoustic signal processing. IEEE Signal Processing, 15-4, pp. 21-53 (1998)
[4] Buck, B., Macaulay, V. A.: Maximum Entropy in Action. Oxford Science Pub. (1991)
[5] Green, T. J. Jr.: Robust passive sonar. DARPA Tech 2000, Dallas (2000)
[6] Tolstoy, A.: Matched Field Processing For Underwater Acoustics. World Scientific Pub. New Jersey (1993)
[7] Dillard, R. A., Dillard, G. M.: Detectability of Spread Spectrum Signals. Artech House, Norwood (1989)
[8] Davanport, R. L.: The influence of the ocean environment on sonar system design. EASCON Record 1975, pp. 200-207 (1975)
[9] Apel, J. R., Badiey, M., Chiu, C. S., et al.: An overview of the 1995 SWARM shallow water internal wave acoustic scattering experiment. IEEE Oceanic Eng., 22-3, pp. 465-500 (1997)
[10] Proc. of IEEE Special issues on Real Time systems, Vol.82 (1994) No.1
[11] Brillinger, D. R.: Some basic aspects and uses of high order spectra. IEEE Trans., SP-36-3, pp. 239-250 (1994)
[12] Swingler, D., Walker, R.: Line array beamforming using linear prediction for

aperture interpolation and extrapolation. IEEE Trans., ASSP-37-1, pp. 16-30 (1989)
[13] Li, Q. H., Zhang, C. H.: A new architecture of spatial filter for suppressing strong directional interference. In: Proc. Of IMDEX'99, Singapore, pp. 87-96 (1999)
[14] Trider, R. C.: A fast Fourier transform (FFT) based on sonar signal processor. IEEE Trans., ASSP-26, pp. 15-20 (1978)
[15] Jarvis, S., Morrissey, R.: High rate underwater acoustic telemetry with naval applications. Sea Technology, 40-5, pp. 67-70 (1999)
[16] Li, Q. H., Yin, L.: A new algorithm of background equalization in digital sonar. J. of Chinese Acoustics, 18-3, pp. 211-216 (1999)
[17] Rickard, J. T., Zeidler, J. R.: Second order output statistics of adaptive line enhancer. IEEE Trans., ASSP-27, pp. 31-39 (1979)
[18] Hahn, W. R.: Optimum signal processing for passive sonar range and bearing estimation. J.Acoust.Soc. Amer., 58, pp. 201-207 (1975)
[19] Carter, G. C.: Time delay estimation for passive sonar signal processing. IEEE Trans., ASSP-29, pp. 463-470 (1981)
[20] Piersol, A. G..: Time delay estimation using phase data. IEEE Trans., ASSP-29, pp. 471-477 (1981)
[21] Robertson, G. H.: A fast amplitude approximation for quadrature pairs. Bell Syst. Tech. J., 50, pp. 2849-2852 (1971)
[22] Wax, M., Kailath, T.: Decentralized processing in sensor arrays. IEEE Trans., 73, pp. 1123-1129 (1985)
[23] Barton, P.: Digital beam forming for radar. IEEE Proc., 127-F-4, pp. 266-277 (1980)
[24] Li, Q. H.: A precise time delay estimation for split beam array system. In: Proc. ICASSP'86, Tokyo, pp. 1816-1819 (1986)
[25] Ianniello, J. P.: Large and small error performance limits for multi path time delay estimation. IEEE Trans., ASSP-34, pp. 245-251 (1986)
[26] Quaiz, A. H.: An overview of the time delay estimation in active and passive system for target localization. IEEE Trans., ASSP-29, pp. 527-533 (1981)
[27] Johnson, D. H.: Improving the resolution of bearing in passive sonar arrays by eigenvalue analysis. IEEE Trans., ASSP-30-4, pp. 638-647 (1982)
[28] Cadzow, J. A.: Multiple source location – the signal subspace approach. IEEE Trans., ASSP-38, pp. 1110-1125 (1990)
[29] Satish, A., Kaskyap, R. L.: Multiple target tracking using maximum likelihood principle. IEEE Trans., SP-42, pp. 1577-1695 (1995)
[30] Cox, H.: Sensitivity consideration in adaptive beamforming. In: Signal processing, Griffith, J. W. R. (Eds.), Academic Press, Boston (1973)
[31] Li, Q. H.: An application of DSP technique in underwater target automatic tracking. In: Proc. of ICSPAT'93, pp. 1574-1577 (1993)
[32] Jabrikian, J., Messer, H.: Three dimensional source localization in a waveguide. IEEE Trans., SP-44-1, pp. 1-13 (1996)
[33] Janffret, G. G., Musso, C. J.: New results in target motion analysis. In: Proc. UDT'91, Paris, pp. 239-244 (1991)

[34] Li, Q. H.: Signal separation theory using adaptive array. J. of Intl. Adaptive Control and Signal Processing, 2, pp. 121-130 (1988)
[35] Doisy, Y.: General motion estimation from correlation sonar. IEEE J. of Oceanic Engr., 23-5, pp. 127-140 (1998)
[36] Smidt, R. O.: Multiple emitter location and signal parameter estimation. In: Proc. of RSDC, Spectral Estimation Workshop, pp. 243-256 (1979)
[37] Roy, R., Kailath, T.: ESPRIT – estimation of signal parameters via rotational invariance techniques. IEEE Trans., ASSP-37, pp. 984-995 (1989)
[38] Viberg, M., Ottersten, B., Kailath, T.: Performance analysis of the total least square ESPRIT algorithm. IEEE Trans., SP-39, pp. 1122-1135 (1991)
[39] Li, Q. H.: A trace extract technique for fast moving target in underwater acoustics. In: Proc. of UDT'99, pp. 177-180 (1999)
[40] Diracca, C., Crandruilemin, B.: Torpedo warning receivers. In: Proc. of UDT'94, London, pp. 44-46 (1994)
[41] Li, Q. H., Li, S. Q.: Range directivity – a new concept in passive source bearing. In: Proc. of UDT'92, pp. 7-12 (1992)
[42] Kummert, A.: Fuzzy technology implemented in sonar systems. IEEE J. of Oceanic Engr., 18-4, pp. 439-446 (1993)
[43] Campbell, J. P. Jr.: Speaker recognition: a tutorial. In: Proc. of IEEE, 85-9, pp. 1437-1462 (1997)
[44] Trearen, P., Pacheco, M., Vellasco, M.: VLSI architectures for neural networks. In: IEEE Micro., pp. 8-27 (1989)
[45] Widrow, B., Lehr, M. A.: 30 years of adaptive neural networks: perception, adaline and backpropagation. In: Proc. of IEEE, 78, pp.1415-1441 (1990)
[46] Miller, W. et al.: Office of naval research contribution to neural networks and signal processing in ocean engineering. IEEE J. of Oceanic Engr., 17-4, pp. 299-307 (1992)
[47] Yu, P. T., Coyle, E. J.: The classification and associative memory of stack filters. IEEE Trans., SP-40, pp.2483-2497 (1992)
[48] Grenie, M.: Acoustic detection of propeller cavitation. In: Proc. ICASSP'90, pp. 2907-2910 (1990)
[49] Deller, J. R.: Set membership identification in digital signal processing. IEEE ASSP Magazine, 6-1, pp. 4-20 (1990)
[50] Fox, M. S.: AI and expert system myths, legends and facts. IEEE Expert, 5-1, pp. 8-22 (1990)
[51] Parsay, K., Chignell, M.: Expert systems for experts. John Wiley, New York (1988)
[52] Li, Q. H.: The nearest neighborhood criteria in underwater target recognition. In: Proc. of UDT'96 (1996)
[53] Tarlton, M.: The use of an expert system toolkit of sonar system design. Naval Forces, X-XII, pp. 62-66 (1994)
[54] Lourens, J. G.: Passive sonar ML estimator for ship propeller speed. IEEE J. of Oceanic Engr., 23-4, pp. 448-453 (1998)
[55] Maksym, J. N.: Expert systems for ship noise interpretation. In: Adaptive methods in underwater acoustics, Urban, H. G. (Eds.), pp. 653-670 (1985)

[56] Lucas, P., Van der Gaag, L.: Principles of expert systems. Addison-Wesley, New Jersey (1991)
[57] Boashash, B., O'Shea, P.: A methodology for detection and classification of some underwater acoustic signals using time-frequency analysis techniques. IEEE Trans. on Acoustics, Speech, and Signal Processing, 38-11, pp. 1829-1841 (1990)
[58] Bregman, A. S.: Auditory scene analysis: the perceptual organization of sound. The MIT Press, London (1990)
[59] Brown, G. J., Wrigley, S. N.: Feasibility study into the application of computational auditory scene analysis techniques to sonar signals, Technical report, University of Sheffield, Department of Computer Science (2000)
[60] Song, H. C., Kuperman, W. A., Hodgkiss, W. S.: Recent results from ocean acoustic time reversal experiments. In: Proc. of ECUA'2002, Gdansk, pp. 279-284 (2002)
[61] Kuperman, W. A., Hodgkiss, W. S., Song, H. C. et al.: Phase conjugation in the ocean: experimental demonstration of a time reversal mirror. JASA, 105, pp. 1597-1606(1999)
[62] Seong, W., Byun, S. H.: Robust matched field processing algorithm based on feature extraction. IEEE J. of Oceanic Engr., 27-3, pp. 642-652 (2002)
[63] Gershman, A. B., Sidiropoulos, N. D.(Eds.): Space-time processing for MIMO communications. Wiley, New York (2005)
[64] Chandran, S. (Eds.): Advances in Direction of Arrival Estimation. Artech House, Boston (2005)
[65] Kuperman, W. A., D'Spain, G. L. (Eds.): Ocean Acoustic Interference Phenomena and Signal Processing. AIP Press, New York (2001)
[66] Jackson, D. R., Dowling, D. R.: Phase conjugation in underwater acoustics. JASA, 91, pp. 3257-3277 (1991)
[67] Stojanovic, M.: Recent advances in high-speed underwater acoustic communications. IEEE J. Oceanic Eng., 21, pp. 125-136 (1996)
[68] Yang, T. C., Siderius, M.: Temporal coherence and fluctuation of acoustic signals in shallow water. In: Proc. of the Fifth European Conference on Underwater Acoustics, Lyon, pp. 63-68 (2000)
[69] Kim, S., Edelmann, G. F., Kuperman, W. A. et al.: Spatial resolution of time reversal array in shallow water. J. Acoust. Soc. Amer., 110, pp. 820-829 (2001)
[70] Hawkes, M., Nehorai, A.: Acoustic vector sensor correlations in ambient noise. IEEE J. of Oceanic Engr., 26-3, pp. 337-347 (2001)
[71] Nehorai, A., Paldi, E.: Acoustic vector-sensor array processing. IEEE Trans. Signal Processing, 42, pp. 2481-2491 (1994)
[72] Tolstoy, A.: Matched Field Processing for Underwater Acoustics. World Scientific Pub. New Jersey (1993)
[73] Lee, N., Tracey, B.: Matched field processing – How much does it really benefit? . In: UASP'2001, West Greenwich (2001)
[74] Kilfoyle, D., Baggeroer, A. B.: The state of the art in underwater acoustic telemetry. IEEE J. Oceanic Eng., 25-4, pp. 4-27 (2000)

[75] Proakis, J. G., Sozer, E. M., Rice, J. A., Stojanovic, M.: Shallow water acoustic networks. IEEE Communications Magazine, 39-11, pp. 114-119 (2001)
[76] Freitag, L., Stojanovic, M., Grund, M., Singh, S.: Analysis of channel effects on direct-sequence and frequency-hopped spread-spectrum acoustic communication. IEEE J. of Oceanic Engr., 26-4, pp. 586-593 (2001)
[77] Robinson, A., Lermusiaux, P.: Prediction systems with data association for coupled ocean science and ocean acoustics. In: Proc. of ICTCA, Honolulu (2003)
[78] Quazi, A., Konrad, W.: Underwater acoustic communications. IEEE Com. Magazine, pp. 24-29 (1982)
[79] Jobst, W.: Measurement of the temporal, spatial and frequency stability of an underwater acoustic channel. J. Acoust. Soc. Amer., 63, pp. 62-69 (1979)
[80] Baggeroer, A. B.: Acoustic telemetry – an overview. IEEE J. Oceanic Engr., 9, pp. 229-235 (1984)
[81] Stojanovic, M.: Recent advances in high rate underwater acoustic communications. IEEE J. Oceanic Engr., 21, pp. 125-136 (1996)
[82] Whitman, E. C.: Submarine in Network Centric Warfare. In: Sea Power Archives, www.navyleague.org (2004)
[83] Pruitt, T.: Maritime homeland security for ports and commercial operation. Sea Technology, 45-11, pp. 20-26 (2004)
[84] Akyildiz, I. F., Dario, I. A., Pompili, D. Melodia, T.: Underwater acoustic sensor networks: research challenges. Ad Hoc networks, 3, pp. 257-279, http://www.seciencedirect.com (2005)
[85] Fink, M., Kuperman, W. A., Montagner, J. P., Tourin, A.: (Eds): Imaging of Complex Media with Acoustic and Seismic Waves. Springer, New York (2002)
[86] Stojanovic, M. et al.: Adaptive multi-channel combining and equalization for underwater acoustic communications. J. Acoust. Soc. Amer., 94, pp. 1621-1631 (1993)
[87] Stojanovic, M., Catipovic, J. A., Proakis, J. G.: Phase-coherent digital communications for underwater acoustic channels. IEEE J. Oceanic Eng., 19-1, pp. 100-111 (1994)
[88] Westervelt, P. J.: Parametric acoustic array. JASA, 35, pp. 535-537 (1963)
[89] Rasmussen, R. A.: Studies related to the design and use of time / bearing sonar display. AD 690540 (1968)
[90] Neri, D. F., Kobus, D. A., Luria, S. M. et al.: Effect of background and foreground color coding on detection: acoustic data analysis. AD-A 154105 (1985)
[91] Butler, W. B., McKemie,W. M.: Engineering guidelines for the use of color on the sonar display. AD 782597 (1974)
[92] Deuser, M., Middleton, D.: On the classification of underwater acoustic signal: I and II. J. Acoust. Soc. Amer., 58 (1975)
[93] Roddy, A. R., Stosz, J. D.: Fingerprint features – statistical analysis and system performance estimates. In: Proc. of IEEE, 85-9, pp. 1437-1462 (1997)

[94] Hartgan, J. A.: Clustering Algorithms. Wiley, New York (1975)
[95] Anderberg, R. A.: Cluster Analysis for Applications. Academic Press, Boston (1973)
[96] Turban, E., Watkins, P. R. (Eds.): Applied Expert Systems. North Holland, Amsterdam (1988)
[97] Sari, H., Woodward, B.: Digital acoustic voice communications. Sea Technology, 39-5, pp. 31-36 (1998)
[98] Curtin, T. R.: ONR program in underwater acoustic communications. Sea Technology, 40-5, pp. 17-27 (1999)
[99] Forbes, A.: Acoustic monitoring of global ocean climate. Sea Technology, 35-9, pp. 65-69 (1994)`
[100] Bucaro, J. A.: Optical hydrophones for sonar. EASCON'78 Record, pp. 298-302 (1978)
[101] Olmo, G., Magli, E., Presti, L. L.: Joint statistical signal detection and estimation Part I : theoretical aspects of the problem. Signal Processing, 80, pp. 57-73 (2000)
[102] Song, H. C., Kuperman, W. A., Hodgkiss, W. S.: A time reversal mirror with variable range focusing. J. Acoust. Soc. Amer., 103, pp. 3234-3240 (1998)
[103] Kim, S., Edelmann, G. F., Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T.: Spatial resolution of time reversal array in shallow water. J. Acoust. Soc. Amer., 110, pp. 820-829 (2001)
[104] Clark, A., Sekino, H.: A multidisplinary deep sea long term observatory in Japan. In: Proc. Oceans 2001 Conference, Honolulu, pp. 1290-1295 (2001)
[105] Detrick, R., Frye, D. Collins, J., Gobat, J.: DEOS Moored Buoy Observatory Design Study. WHOI, Technical Report (2000)
[106] Butler, R., Chave, A. D., Duennebier, F. K., et al.: Hawaii-2 observatory pioneers opportunities for remote instrumentation in ocean studies, EOS. Trans. American Geophysical Union, 81-15, pp. 162-163 (2000)
[107] Ballard, R. D., Yoerger, D. R., Stewart, W. K., et al.: ARGO / JASON: a remotely operated survey and sampling system for full-ocean depth. In: Proc. Oceans 91, 1, pp. 71-75 (1991)
[108] Proakis, J. G.: Digital Communications, 3rd Edn. McGraw-Hill, New York (1997)
[109] Freitag, L., Stojanovic, M., Grund, M., et al.: Analysis of channel effects on direct-sequence and frequency-hopped spread-spectrum acoustic communication. IEEE J. of Oceanic Eng., 26-4, pp. 586-593 (2001)
[110] Robinson, A., Lermusiaux, P.: Prediction systems with data association for coupled ocean science and ocean acoustics. In: Proc. Of ICTCA, Honolulu, www.deas.harard.edu (2003)
[111] Wei, R. C., Chen, C. F., Newhall, A. E., et al: Lin, Po-Chang A preliminary examination of the low-frequency ambient noise field in the south china sea during the 2001 ASIAEX experiment . IEEE Journal of Oceanic Eng., pp. 1308-1315 (2001)
[112] Quazi, A., Konrad, W.: Underwater acoustic communications. IEEE Com.

Magazine, pp. 24-29 (1982)
[113] Fischer, J. H., Bennet, K. H.: High data rate, underwater acoustic data communication. Sea Technology, 34-5, pp. 10-13 (1993)
[114] Gershman, A. B., Sidiropoulos, N. D.(Eds.): Space-Time Processing for MIMO Communications. Wiley, New York (2005)
[115] Rion, O., Vetterli, M.: Wavelets and signal processing. IEEE SP Magazine, pp. 14-38 (1991)
[116] Shensa, M. J.: The discrete wavelet transform: Wedding the Trous and Mallat algorithm. IEEE Trans., SP-40, pp. 2462-2482 (1992)
[117] Fanily, F.: Fractals. DSP Applications, 2-12, pp. 50-58 (1993)
[118] Ramnsley, M. F.: Fractals Everywhere. Wiley, New York (1990)
[119] Knapp, C. H., Carter, G. C.: The generalized correlation method for estimation of time delay. IEEE Trans., ASSP-24, pp. 320-327 (1976)
[120] Messer, H.: The use of spectral information in optimal detection of a source in the presence of a directional interference. IEEE J. of Oceanic Eng., 19-3, pp. 416-421 (1994)
[121] Li, Q. H., Schwartz, S. C.: A Kind of Pre-processing of Robust Wiener filter. Department of EE, Princeton University (1985)
[122] Li, Q. H., Schwartz, S. C.: A Combined Architecture of Adaptive Noise Canceller and Line Enhancer. Department of EE, Princeton University (1985)
[123] Volder, T. E.: The CORDIC Trigonometric Computing Technique. IRE Trans., EC-8 (1959)
[124] Li, Q. H.: An Introduction to Sonar Signal Processing. Ocean Pub., Beijing (1985)
[125] Heimdal, P., Bryn, F.: Passive ranging techniques. In: Proc. of NATO ASI on Signal Processing, New York (1972)
[126] Sutton, R.: Optimization of human / machine control systems. In: Proc. of UDT'1990, London, pp. 243-248 (1990)
[127] Comon, P. et al.: Sonar interception from a single beam or sensor. In: Proc. of UDT'1996, London, pp. 258-261(1996)
[128] Edwards, J. R., Schmidt, H., LePage, K. D.: Bistatic synthetic aperture target detection and imaging with an AUV. IEEE J. of Oceanic Eng., 26-4, pp. 690-699 (2001)
[129] Coraluppi, S.: Multistaic sonar localization. IEEE J. of Oceanic Eng., 31-4, pp. 964-974 (2006)
[130] De Pasquale, S., Albano, F.: Passive and active multistatic / multisonar data processing with / without predefined coherent and / or impulsive single / multi source. In: Proc. of UDT'2009, Cannes (2009)
[131] Varshney, P. K.: Distributed Detection and Data Fusion. Springer, New York (1996)
[132] Hall, D. L., Llinas, J.: An introduction to multisensor data fusion. In: Proc. of IEEE, 85-1, pp. 6-23 (1997)
[133] Viswannathan, R., Varshney, P. K.: Distributed detection with multiple sensors: Part I – fundamentals. In: Proc. of IEEE, 85-1, pp. 54-63 (1997)

[134] Tenny, P. R., Sandel, N. R.: Detection with distributed sensors. IEEE Trans., AES-17-4, pp. 501-510 (1981)
[135] Li, Q. H.: The optimum linear data fusion of independent observation data. Chin. J. Acoust, 20-1, pp. 18-24 (2001)
[136] Li, Q. H.: The optimum linear data fusion of dependent observation data. Chinese J. of Acoustics, 20-2, pp. 97-102 (2001)

7

System Simulation Techniques in Digital Sonar Design

When we design modern digital sonar according to tactical and technical requirements, we naturally face a very complicated situation: many dependent parameters must be selected, and many decisions concerning acceptance / rejection techniques must be determined. We have to evaluate whether to increase the complexity of the equipment, if necessary, in order to enhance performance. And also we have to make a trade-off between various benefits or costs. But we cannot manufacture the hardware at the design phase, since this will result in a waste of labor and materials, and postpone the period of development.

Fortunately, with the development of computer techniques, we are able to simulate our design on a computer. System simulation is a digital technique, which carries out experiments on the computer. It includes some kinds of mathematical and logical models which describe a real time system.

There is no doubt that system simulation will considerably save the cost of sonar design and reduce the period of manufacture.

Of course, simulation is not the same as reality. In particular, the ocean environment of underwater acoustics is very complicated. Exact and accurate simulation of the ocean environment is difficult. The motivation for us is to continuously make efforts to improve system simulation techniques.

In this chapter we will introduce the basic techniques of system simulation for sonar design[1-12] and also briefly discuss the design principles of a sonar simulator.

7.1 General Concept of System Simulation

In the computer science field, system simulation represents the realization of some kind of dynamic system or phenomenon by using a computer. The goal of system simulation is to implement laboratory testing or performance prediction. System simulation has already become the most powerful and most prospective area in

computer applications. System simulations are extensively used in industry, science, administration management and the economy, etc., including space flight, nuclear explosions, weather forecasting, human population prediction, birth control, micro management / control of the economy, and so on.

7.1.1 Brief Introduction of System Simulation Technique

System simulation in the sonar design field is defined as realizing certain physical or logical systems using a computerized model.

First, we have to precisely describe the system to be simulated. It is necessary to decompose and categorize the specifications or functions with which we are concerned. Second, it is necessary to establish a mathematical model which can be used on the computer. And the operational environment should be logically correlated. Third, we compile the computer program. Any programs, whatever the computer language used, must exactly meet the requirements of the modeling. We must be able to adjust the parameters of the program and it must be easy to change the flow chart direction so that it is easy to test. Fig. 7.1 illustrates the block diagram of a system simulation procedure. At last, we run the compiled program, the necessary output data and figures should be used to evaluate the simulation results. And the final analysis and conclusions are derived after system simulation.

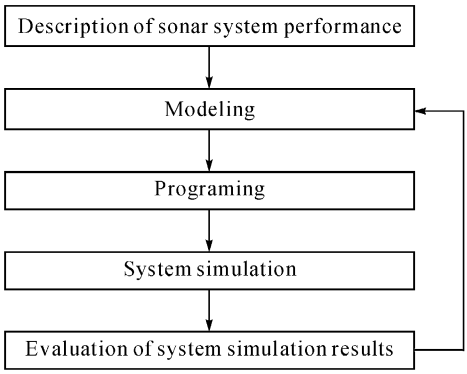


Fig. 7.1. General concept of system simulation

A modern digital simulation technique was proposed by von Neumann. He predicted the important role of digital data in system modeling. In 1944, the method of a random number experiment in solving theoretical problems was named the Monte Carlo method. In 1952, Dr. McLeod organized the Society for Computer Simulation. This was the first professional association in the computer simulation area[1,9].

In 1777, the French scientist Buffon suggested using a random number experiment to calculate the value of π. This may be the earliest application of the

Monte Carlo method. The experiment suggested by Buffon is shown in Fig. 7.2. There are a series of parallel straight lines with unit length apart. And the length of a needle is $l \le 1$. Suppose we drop the needles on the plane and account for the number of needles which cross one of the lines. If the total number of dropped needles is N, and the number of needles which cross one line is M, then it is proved that $\pi = 2lN / M$. It is easy to show that

$$\frac{M}{N} = \frac{2}{2\pi}\int_0^{\pi} l \sin\alpha \mathrm{d}\alpha = \frac{2}{\pi} l$$

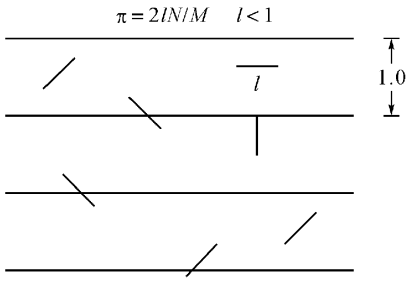


Fig. 7.2. Example of Monte Carlo method

Some scientists have carried out this interesting experiment and the results are shown in Table 7.1.

Table 7.1 Buffon's needle experiment

Tester	Year	l	N	M	π
Wolf	1850	0.80	5000	2532	3.15926
Smith	1855	0.60	3204	1218	3.15665
Fox	1884	0.75	1030	489	3.15951
Lazzarini	1925	0.83	3408	1808	3.1415992

7.1.2 Development of Modern System Simulation Technique

In a general sense, system simulation can be understood as a study process for an existing system or a developing system. In order to evaluate the inner performance of this system and the interactive effect with the outer environment, it is necessary to carry out the experiment on a computer. The modern system simulation technique is based on information, the computer and other related professional knowledge. It is a synthesis of theory and technique.

The system simulation technique can be categorized in different classes. Based on the simulation objects, it can be divided into continuous, discrete and hybrid system simulation. Based on the structure of the simulation system and means of

implementation, it can be divided into digital simulation, hardware simulation and software simulation. Based on the application area of the simulation, it can be divided into training simulation and engineering simulation.

A large system is usually very complicated. The parameters to be adjusted may be in the order of hundreds. Therefore the system simulation is also very complicated. Sometimes we have to break down the entire system into several subsystems and then carry out the system simulation. Large scale system simulation can be carried out in several stages. Therefore system modeling should be considered very carefully in order to meet the requirements of various simulation stages. System simulation can be carried out not only on a computer but also on a cluster of computers or distributed computer system, even on a network of computers.

7.2 Theoretical Basis of System Simulation: Monte Carlo Method

The goal of sonar system simulation on a computer is to verify the feasibility of design schema. The preliminary task is to simulate the ocean environment field. That is to say we have to generate various signals, background noise, ambient noise and interference. All these signals have to be completely similar to the real life situation. The signals and noise simulated on a computer are input to the simulated sonar system. Therefore, it is necessary to find an effective method to generate random signals which have an expected performance. The most powerful means of doing that is by the Monte Carlo method.

The ability to generate random signals and noise that are perfectly similar to the real situation is the key point in sonar system simulation, since, if the simulated signal and noise are far from the real situation, then the verification of the sonar system performance is unachievable.

Of course, this is not to say that all the physical parameters of the environment can be simulated. Some physical parameters of the ocean environment cannot be simulated. For example, we cannot provide system simulation of a sound pressure test for a sensor. This kind of test can be only carried out in a real laboratory.

7.2.1 Method for Generating a Random Number

Fig. 7.3 illustrates the main contents of sonar system simulation. The simulation objectives are divided into three classes: environment simulation, signal processing simulation and post processing simulation.

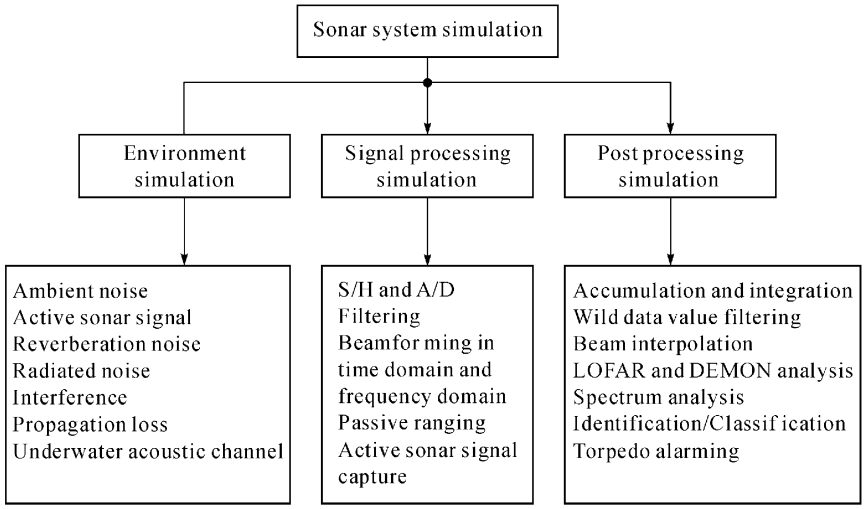


Fig. 7.3. Classification of sonar system simulation

Environment simulation is to simulate signals, noise and interference, which is the sonar utilization environment. The later two classes simulate the sonar itself. We want to construct a virtual sonar scenario on a digital computer and verify the performance of this sonar. First, we have to simulate the ocean environment since the signal, noise and interference we are going to simulate are either deterministic or stochastic. Therefore, the preliminary work in sonar simulation is to study how to generate deterministic and stochastic signals. Before discussing the method of generating various signals, we will explain how to use the Monte Carlo method to implement a specific task on a digital computer. See Fig. 7.4 and suppose there is an inscribed circle in a square and the side of this square is 1. It is obvious that the area of the square is 1 and the area of the circle is $\pi / 4$. On a digital computer we can generate two independent random numbers *RND*_1 and *RND*_2 which are uniformly distributed in (0.5, 0.5). Denote

$$Z = (RND_1)^2 + (RND_2)^2 \tag{7.1}$$

It is obvious that if $\sqrt{Z} \le 1$, then point Z is in the inside of the circle. If N is the total number of independent pairs (*RND*_1, *RND*_2) and M is the number of $\sqrt{Z} \le 1$, then we have $M / N \approx \pi / 4$ and $\pi \approx 4M/N$. When N is large enough then we can obtain a very accurate approximate value of π.

In this example, the uniform distribution random variable plays an especially important role. It is the basis for generating many different kinds of random numbers. The most popular algorithm for generating an independent random number $RN(k)$, which is uniformly distributed in segment (0, 1), is the so-called "hybrid co-residual method". It is easy to find the subroutine based on this method

in a signal processing software package. The basic principle of this method is described here.

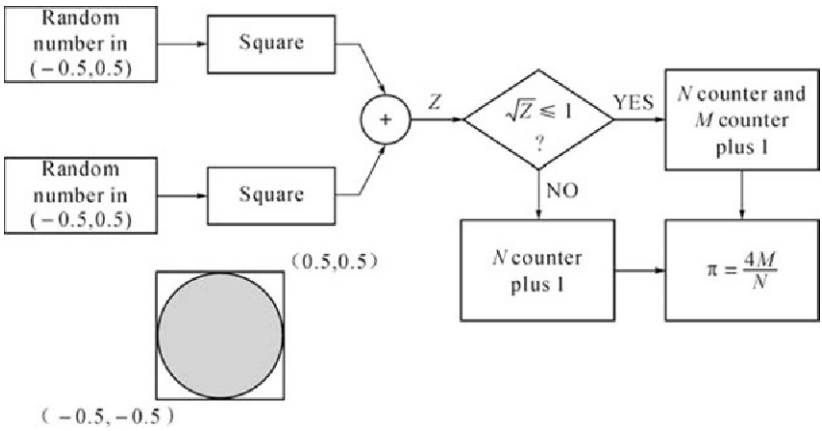


Fig. 7.4. A Monte Carlo method for calculating π

Suppose x_k is the k-th random number, then the $(k+1)$ -th random number is given by

$$x_{k+1} = ax_k + c(\text{mod } M) \tag{7.2}$$

where a, c, and M represent non-negative numbers. And mod M represents the number that is co-residual with M. Usually the value M takes the form of $M = 2^m$. The selection of values a, c, and m should satisfy the requirements of independence and the long period of $RN(k)$.

We obtain

$$RN(k) = x_k / M \tag{7.3}$$

$RN(k)$ is the random number which is uniformly distributed in (0, 1). There are many well verified values for a, c, and m. A typical example is a = 257; c = 1,2,3,4,5,6,7,8,9,11; m = 1,2,1,3,7,5,16,14,15.

Based on the random number of $RN(k)$, we can generate a random number with any given probability density function. The following theorem is the basis of this conclusion.

Theorem 7.1 *Suppose r is the random number with uniform distribution in* (*0*, *1*). *For a given probability distribution density function* $p(x)$, *let*

$$F(x) = \int_{-\infty}^{x} p(t)\mathrm{d}t \tag{7.4}$$

Then the probability density function of random variable $u = F^{-1}(r)$ *is* $p(x)$.

Proof

In terms of the definition of a probability distribution function, we have

$$F_u(x) = P\{u \le x\} = P\{F^{-1}(r) \le x\} = P\{r \le F(x)\} = \int_0^{F(x)} \mathrm{d}t = F(x) \quad (7.5)$$

This proves our theorem.

Example:

Suppose $RN(k)$ is the random number with uniform distribution in (0, 1) and $p(x) = q\exp(-qx)$ $(q > 0)$. This is a negative exponential probability distribution. We have

$$F(x) = \int_{-\infty}^{x} p(t)\mathrm{d}t = 1 - \exp(-qx)$$

The inverse function is $x = -\ln[1 - F(x)]/q$. Based on the above theorem, we know that $u(k) = -\ln[1 - RN(k)]/q$ obeys the exponential probability distribution. Because $RN(k)$ and $1 - RN(k)$ are both uniformly distributed in (0, 1), therefore

$$z(k) = -\ln[RN(k)]/q \quad (7.6)$$

is the random variable which has the probability density function $p(x) = q\exp(-qx)$.

7.2.2 Basic Time Series in Digital System Simulation

For sonar system simulation, the discrete time series appears in various data flows of the program. These kinds of time series consist of several basic time series. The uniform distributed random series is one of the basic time series. The simulation of a deterministic signal is in the form of

$$x(k) = A\cos(2\pi k f_0 / f_s + \varphi) \quad (7.7)$$

where f_0 represents the signal frequency, f_s represents sampling frequency. A is the amplitude of the signal and φ is the random phase, uniformly distributed in $(0, 2\pi)$. For the ordinal number $k = 0, 1, 2, \ldots$ it obtains the deterministic

signal samples. It is proved that $x(k)$ in Eq. (7.7) is stationary and an ergodic random process. From pure random process theory, only in the case where φ is an irrational number and $x(k)$ is stationary and ergodic, but since an irrational set almost all exists in $(0, 2\pi)$, therefore, for simplification, we acknowledge that $x(k)$ is stationary and ergodic for all φ.

The strength of signal $x(k)$ is $A^2/2$. In the simulation process, if we want to change the SNR, one method is to change the value of A.

The discrete sampling of a continuous signal is shown in Fig. 7.5. As we can see, the sampling duration must be short enough to guarantee small distortion. The discrete time samples are quantized in amplitude and then become a complete digital signal. The system simulation technique in this stage must choose the right sampling frequency and bit number of the A/D converter, which are exactly the same as what we designed in the digital sonar system.

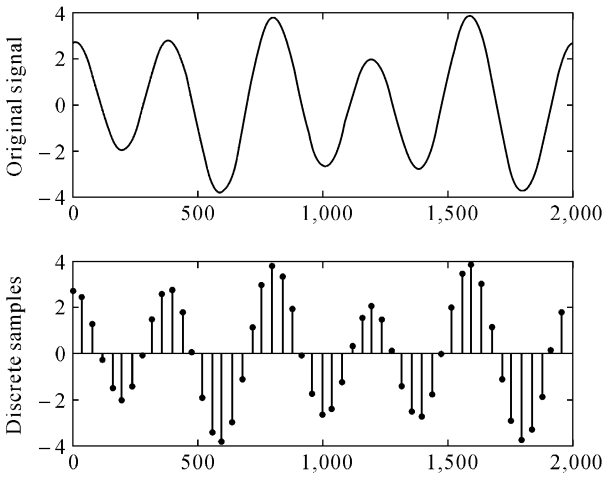


Fig. 7.5. The discrete sampling of a continuous signal

Now we turn to discuss the FIR filtering simulation. First, we should determine the length of the FIR filter. Suppose the frequency response of an analogue filter is $H(f)$, the DFT of this frequency response is

$$w(0),\ ...,\ w(N-1) \tag{7.8}$$

If the input signal of this FIR filter is $x(k)$, then the output is

$$y(k) = \sum_{i=0}^{N-1} w(i)x(k-i) \tag{7.9}$$

Fig. 7.6 illustrates the coefficients of a DFT filter for an ideal rectangle frequency

response. There is a basic principle in choosing FIR coefficients: it is preferable to take a longer discrete FIR rather than a shorter one, since the longer FIR is closer to corresponding analogue filtering.

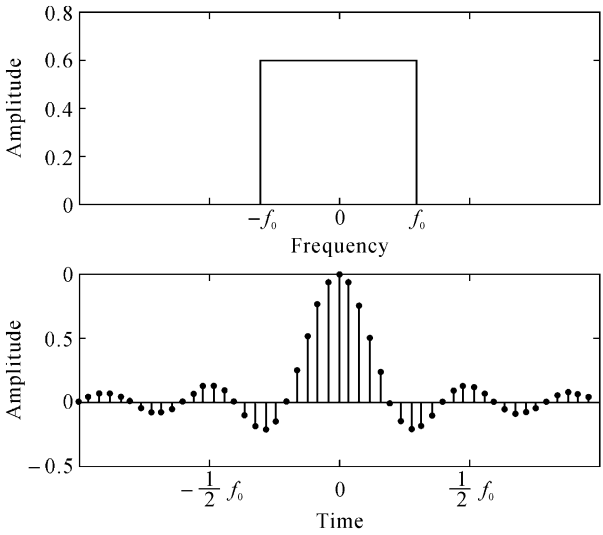


Fig. 7.6. FIR filtering coefficients

In system simulation, the most popular signal form is a Gaussian or Gaussian generated signal, because the central limit theorem in probability theory shows that the summation of a large amount of independent random effects will result in a Gaussian random process. The mechanism of ambient noise and underwater target radiated noise, in a certain sense, can be considered to be the result of a large amount of independent random disturbs. Therefore, in most cases, it is supposed that the ambient noise of the ocean environment is a Gaussian random process.

We can use theorem 7.1 to generate a Gaussian random number, but there is no closed form of expression for the inverse of a Gaussian probability distribution function. I is not possible to directly use an inverse function.

There are two frequently used methods for generating a Gaussian random number. One is a moving average algorithm. The theoretical background to this algorithm is the central limit theorem. Fig. 7.7 shows the block diagram of this algorithm.

Based on the central limit theorem described in Chapter 2, we know that if R_1, ..., R_N are N independent random variables with the same distribution, then for a big enough N

$$G = \frac{(R_1 + ... + R_N)/N - \mu}{\sqrt{\sigma^2 / N}} \tag{7.10}$$

tends to have Gaussian or normal distribution, where $\mu = E(R_k)$, $\sigma^2 = \mathrm{Var}(R_k)$, $(k = 1, ..., N)$. Then, it is easy to obtain a Gaussian random series by using a random number, which is uniformly distributed in (0, 1).

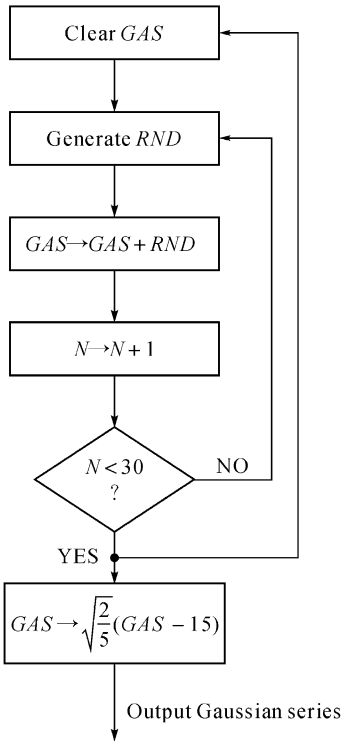


Fig. 7.7. A method to generate a Gaussian number

Suppose $RN(k)$ are independent random variables which are uniformly distributed in (0,1), then $\mu = E[RN(k)] = 0.5$, $\sigma^2 = \mathrm{Var}[RN(k)] = 1/12$. In terms of Eq. (7.10) we know

$$GAS(k) = \left[\sum_{i=1}^{N} RN(kN+i) - N/2\right]\sqrt{12/N} \tag{7.11}$$

is the independent Gaussian signal series or white noise. Here the number N plays an important role. It is proved that for $N \geq 30$, $GAS(k)$ is close to having a Gaussian probability distribution function.

Fig. 7.8 illustrates an example of a Gaussian number. It is very close to the ambient noise record in outside view.

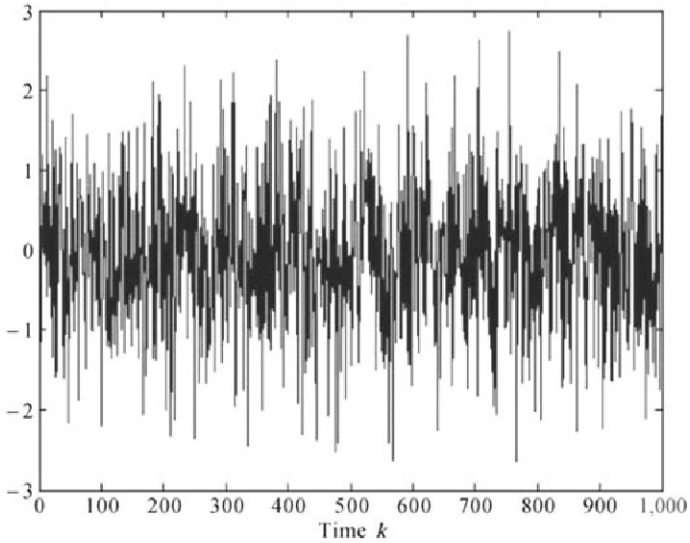


Fig. 7.8. Normal random number series (zero mean and variance 1)

The other algorithm for generating a random Gaussian independent series is to use two independent random numbers *RND*_1 and *RND*_2

$$\begin{aligned} X_1 &= (-2\ln RND_1)^{1/2}\cos(2\pi RND_1) \\ X_2 &= (-2\ln RND_2)^{1/2}\cos(2\pi RND_2) \end{aligned} \tag{7.12}$$

It is proved that X_1 and X_2 are independent Gaussian variables with mean value 0 and variance 1.

There are two important digital features of the random series, one is the auto-correlation function and one is the cross-correlation function. It is necessary to verify these two correlation functions when we generate a random series.

Suppose $x(k)$ ($k = 0, 1, 2, \ldots$) is a random series. The auto-correlation function of $x(k)$ is defined as

$$R_x(k) = \frac{1}{N}\sum_{i=0}^{N-1} x(i)x(k+i)\ (k = 0,\ 1,\ 2,\ ....) \tag{7.13}$$

If $x(k)$ is white noise, then for $k \neq 0$, $R_x(k) = 0$. That means the auto-correlation function is δ function.

If $y(k)$ ($k = 0, 1, 2, \ldots$) is another random series, then the cross-correlation function of $x(k)$ and $y(k)$ is defined as

$$R_{xy}(k) = \frac{1}{N}\sum_{i=0}^{N-1} x(i)y(k+i)\ (k = 0,\ 1,\ 2,\ ...) \tag{7.14}$$

Fig. 7.9 illustrates an example of the auto-correlation function of a white noise random series. Fig. 7.9(a) is the situation when N = 1,024 and Fig. 7.9(b) is the situation when N = 10,240. It is obvious that the fluctuation in the autocorrelation decreased with an increase of N.

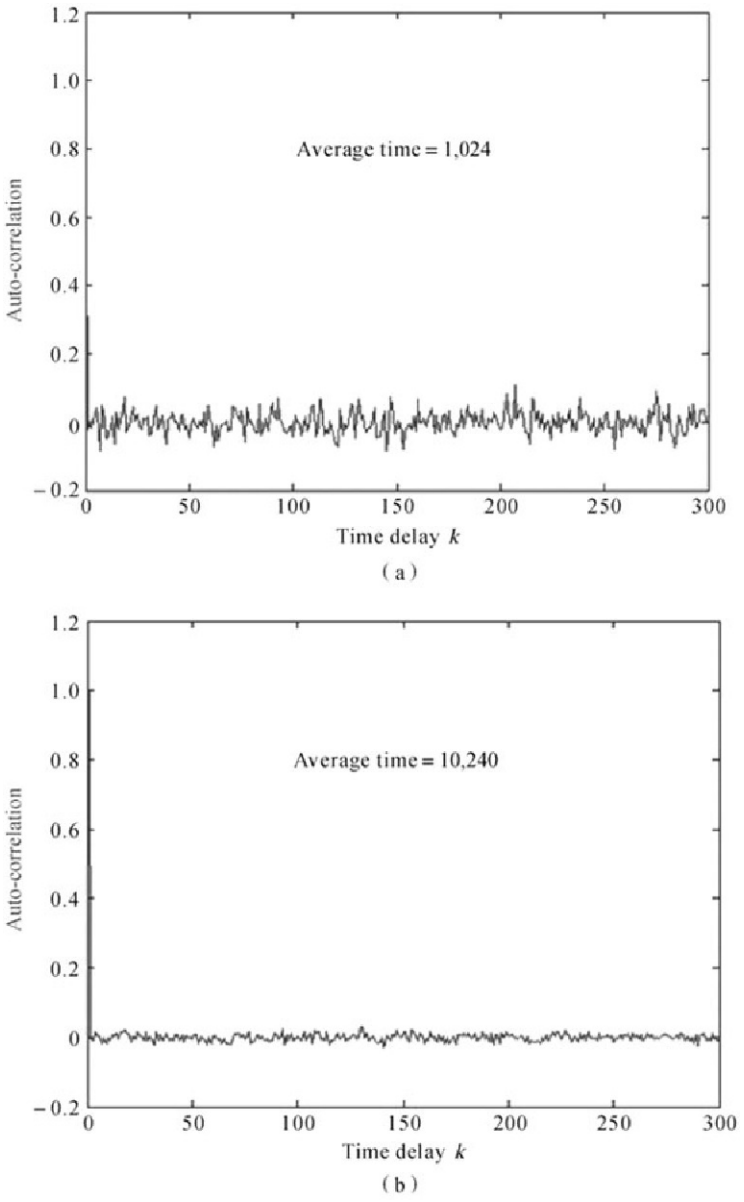


Fig. 7.9. Auto-correlation function of a Gaussian random signal. (a) Average time = 1,024; (b) Average time = 10,240

Fig. 7.10 illustrates the cross-correlation function of two independent Gaussian random series.

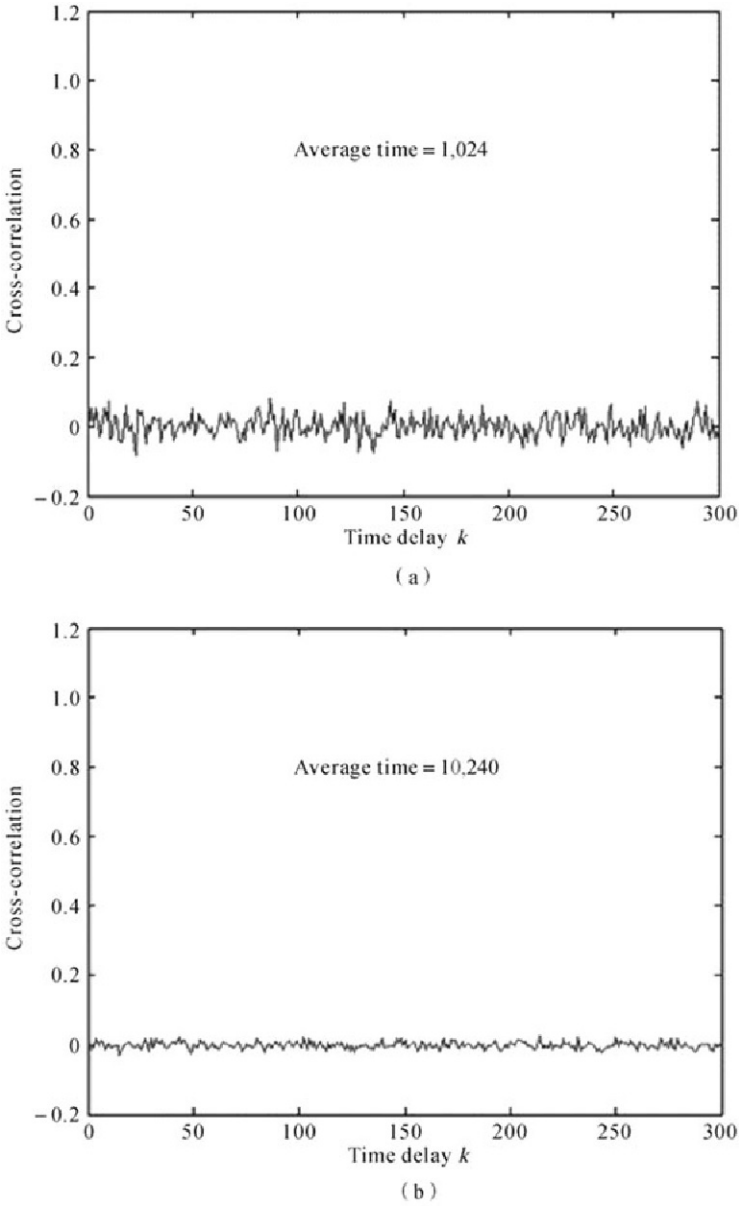


Fig. 7.10. Cross correlation function of a Gaussian random signal. (a) Average time = 1,024; (b) Average time = 10,240

The calculation of the power spectrum of a random series is a little more

complicated than for the correlation function. There are two methods that can be used to calculate the power spectrum. One is to calculate the DFT for a random series. The other is to calculate the correlation function and then take DFT for the correlation function. Both of them are effective. It is necessary to show that the data is long enough so that we can obtain a precise spectrum value. We can also divide a long data series into several non-intersection segments, for each segment perform DFT and then sum up the DFT value in the same bins to obtain average results. This is the true power spectrum in the sense of a statistical average.

Suppose the random series is $x(k)$ $(k = 0,\ 1,\ 2,\ ...)$, we divide this series into M non-intersection segments: $x(mN + k)$ $(k = 0,\ ...,\ N-1;\ m = 0,\ ...,\ M-1)$. Usually, the number N takes the form of the power of 2. The DFT of m-th segment data is

$$X_m(l) = \sum_{k=0}^{N-1} x(mN + k)\exp(-2\pi \mathrm{j}kl / N)\ (l = 0,\ ...,\ N-1) \tag{7.15}$$

The power spectrum value in the l-th bin is $|X_m(l)|^2$. The average power spectrum is defined as

$$|X(l)|^2 = \frac{1}{M}\sum_{m=0}^{M-1} |X_m(l)|^2\ (l = 0,\ ...,\ N-1) \tag{7.16}$$

Eq. (7.16) is one of the methods for calculating the power spectrum value. The other method is first to calculate the auto-correlation function of $x(k)$ and then perform the DFT on auto-correlation. In calculating autocorrelation, whether we adopt segment accumulation or directly calculate for the long length series, the result is the same. But it is necessary to point out that, in either situation, the length of the series should at least be longer than that of the maximum time delay. As an example, suppose the data length is $(M+1)N$. The auto-correlation function is

$$R_x(k) = \frac{1}{NM}\sum_{i=0}^{NM} x(i)x(k+i)\ (k = 0,\ ...,\ N-1) \tag{7.17}$$

Take DFT for $R_x(k)$, we obtain the power spectrum value of $x(k)$:

$$K_x(l) = \sum_{k=0}^{N-1} R_x(k)\exp(-2\pi \mathrm{j}kl / N)\ (l = 0,\ ...,\ N-1) \tag{7.18}$$

It is proved that the value $K_x(l)$ expressed in Eq. (7.18) is close to the value $|X(l)|^2$ in Eq. (7.16).

Fig. 7.11 illustrates the power spectrum of Gaussian random white noise. As we can see, the shape of the power spectrum is close to flat.

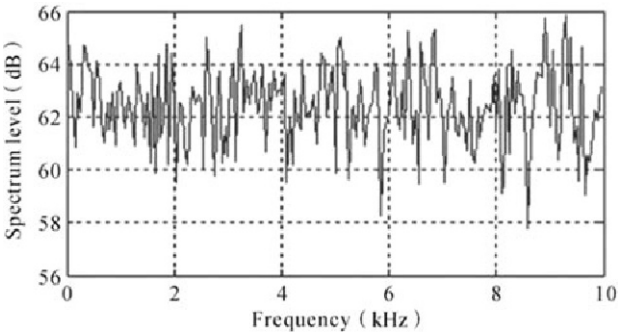


Fig. 7.11. The power spectrum of a Gaussian random signal

7.3 Frequent Use Simulation Method in Underwater Acoustic Signal Processing

Underwater acoustic signal processing theory provides various techniques and algorithms for digital sonar. In order to verify the feasibility of sonar design, it is necessary to establish the virtual sonar environment, including ambient noise, reverberation, interference, Doppler shift, frequency spread and adaptive noise canceling, etc. It also needs a sufficient database of radiated noise and background noise. The combination of a signal processing algorithm and database provides the possibility of completing a system simulation.

In this section, the frequent use algorithm for system simulation and signal processing is introduced. In the simulation process, the time order of the sample is important, but the time duration between samples only has a relative meaning, i.e., each step in the simulation can be considered as being a millisecond or microsecond.

It is worth showing that the simulation is not equal to reality. Our purpose is to make the simulation as close to reality as we can.

7.3.1 Simulation of the Sonar Environment Field

The ideal sonar environment field consists of the following parts:
(1) Isotopic homogeneous ambient noise field;
(2) Plane wave of directional signal;
(3) Plane interference wave;

(4) Reverberation and echo.

Suppose a sonar array consists of N elements. The signal to be detected is $s(k)$. There are M interferences, $r_1(k)$, ..., $r_M(k)$. The signal received by the i-th element is

$$x_i(k) = s(k-\tau_i) + r_1(k-\tau_{1i}) + ... + r_M(k-\tau_{Mi}) + n_i(k) \ (i = 1, \ ..., \ N) \quad (7.19)$$

where, τ_i, τ_{1i}, ..., τ_{Mi} are supposed to be integers. This is defined as the quantization result of analogue time delay value. The reverberation is not included in Eq. (7.19), it will be discussed later.

Signal Simulation

In order to simulate a plane wave signal, it is necessary to generate a Gaussian random signal or sinusoidal signal denoted by $s(k)$. In terms of $s(k)$, the N element signal will be generated. First the maximum time delay $\tau_{\max}$ of the array signal is calculated. Suppose the quantization value of $\tau_{\max}$ is M_0, i.e., $\mathrm{INT}[\tau_{\max} / T_s] = M_0$. All τ_i's are one of numbers in $[0, M_0]$. The methods for generating $r_i(k)$ $(i = 1, \ ..., \ M)$ are the same as $s(k)$.

Example

Consider an $N = 6$ equal spaced line array. The distance of adjacent elements is $d = 20$ cm. If the incidental signal angle is $\theta = 15^\circ$, based on the formula $\tau_i = (i-1)d \sin\theta / c$ $(i = 1, \ ..., \ 6)$, it obtains the relative time delay 0, 34.5, 69, 103, 137.5, 172 (μs)

If the sampling frequency is f_s = 100 kHz, then T_s = 10 μs. Therefore the tapped numbers of the time delay are 0, 3, 7, 10, 14, 17. The signal we simulated should be $s(k)$, $s(k-3)$, $s(k-7)$, $s(k-10)$, $s(k-14)$, $s(k-17)$.

For the simulation of a sinusoidal signal, sometimes it is more convenient to change the time delay to a phase angle. For example, if the frequency of the sinusoidal signal is f_0 and the sampling frequency is f_s,

$$x(k) = A\cos(2\pi k f_0 / f_s + \varphi) \quad (7.20)$$

where A represents amplitude, φ is the phase, uniformly distributed in $(0, \ 2\pi)$. Any time delay for signal $x(k)$ can be transferred as a phase difference. For example, $x(k-3) = A\cos(2\pi k f_0 / f_s + \phi_3)$, and $\varphi_3 = \varphi - 6\pi f_0 / f_s$. The benefit of the time delay to the phase difference is to save the capacity of the shift register.

Multiple Independent Noise Simulation

In chapter 4 we have mentioned that the simplest model of a background noise field is the isotopic, homogeneous field. This model is difficult to simulate completely, since the noise field is the function of four arguments: one time coordinate and three position coordinates. In a different model of the noise field, the time / space correlation function is different. Therefore, in sonar simulation, the noise field is further simplified for independent noise. This is valid when the distance of the elements is large enough. So it is concluded that the problem of noise field simulation is the simulation of multiple independent noises.

Fig. 7.12 illustrates the method for generating N independent noises. Suppose we have random white noise $RN(k)$. It is considered as a "seed" for generating multiple independent noises. Based on the assumption $R(i)$ and $R(j)$ are independent for $i \neq j$, we divide $RN(k)$ into N groups:

Group 1: $RN(0)$, $RN(N)$, ...;

Group 2: $RN(1)$, $RN(N+1)$, ...;

...

Group N: $RN(N-1)$, $RN(2N-1)$, ...

The moving average of the first group gives the signal of the first element

$$n_1(k) = \sum_{i=1}^{K} RN[(k-i)N]w(i) \tag{7.21}$$

The moving average of the N-th group gives the signal of the N-th element

$$n_N(k) = \sum_{i=1}^{K} RN[(k-i)N + N - 1]w(i) \tag{7.22}$$

where the coefficients $w(i)$ $(i = 1, ..., K)$ are used to adjust the weight in the moving average, such that the correlation function or power spectrum satisfies the desired requirement.

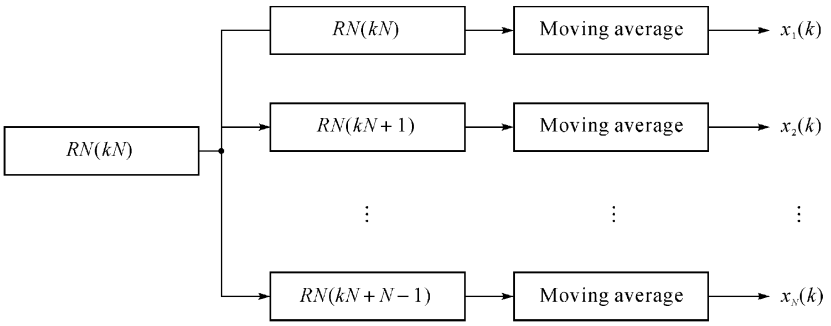


Fig. 7.12. An algorithm for generating N independent random numbers

Simulation of Reverberation

In discussing reverberation modeling, we have shown that since the mechanism of reverberation is very complicated, therefore the simulation technique for reverberation is preliminary. As a matter of fact, in the very short moment after the signal is transmitted, the sonar system allows for the existence of a blind area. It means that for a short time the reverberation simulation can be neglected. It is sufficient to concern the sonar ranging distance in reverberation limitation. Therefore we should study the reverberation simulation a short time after the signal is transmitted. We should simulate not only volume reverberation, but also bottom and surface reverberation. As we have shown in Chapter 4, in this condition the amplitude of reverberation obeys the Rayleigh probability distribution. And the carrier of reverberation is the narrow band signal, which is transmitted like a signal but has some kind of frequency shift or frequency spread. Fig. 7.13 shows a simple model for generating a reverberation simulated signal. Starting from random series $RN(k)$, it is easy to generate two independent random Gaussian signals $x(k)$, $y(k)$. The performance of the correlation function is configured by the algorithm described in Eq. (7.21).

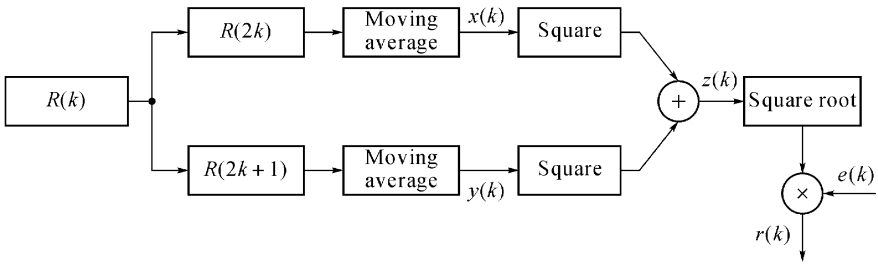


Fig. 7.13. Digital simulation of reverberation

The signal $z(k) = [x^2(k) + y^2(k)]^{1/2}$ is selected to modulate a random series, which is related to the transmitting signal

$$e(k) = A\frac{\exp(-\alpha k)}{k^m} \tag{7.23}$$

Therefore the reverberation is given by

$$r(k) = e(k)z(k) \tag{7.24}$$

The selection of A and α depends on the simulation conditions. Since we are lacking knowledge about the correlation performance of reverberation signals from different elements, so it is hard to give a definite algorithm to show how we can get simulation reverberation for each element from $r(k)$. The only thing we

can do now is to make a little time delay change for $r(k)$ as the representative signal for some element.

7.3.2 Simulation of Signal Processing Module in a Sonar System

The kernel of a signal processor is the beamforming. Once the signal elements are ready, it is necessary to perform digital filtering and beamforming. In order to complete the special function of a sonar system, to include LOFAR, DEMON, ALE and ANC, it is also necessary to carry out suitable algorithms. But these algorithms are easy to complete.

Digital Filtering

The signal from the array element has a very accurate time delay value; therefore the digital filtering must keep the time delay relationship between array elements. That means the FIR filtering cannot add any additional time delay to the signal. Otherwise the beamforming system cannot work properly. Since the FIR filter is one of the linear shift phase filters, therefore the utilization of the FIR filter can guarantee the time delay relationship between array elements and have no effect on beamforming.

Fig. 7.14 illustrates digital filtering for the signals. The discrete weight coefficient window is selected to perform FIR filtering. Usually it is the DFT results for some of the transfer functions of the analogue filter. Suppose the weight coefficients are $w(k)$ $(k = 0, \ ..., \ M-1)$. Then the output signal of this filter for input $x(n)$ is given by

$$y(n) = \sum_{k=0}^{M-1} x(n-k)w(k) \tag{7.25}$$

For each element we should perform the same FIR filtering and get a signal as the input of the beamforming system.

In Fig. 7.14 the FIR filter we selected is the low pass filter, so the output signal is the smoothed results of input signals.

$z(k) = [x^2(k) + y^2(k)]^{1/2}$ is selected as the amplitude to modulate a random series, which is related to the transmitted signal.

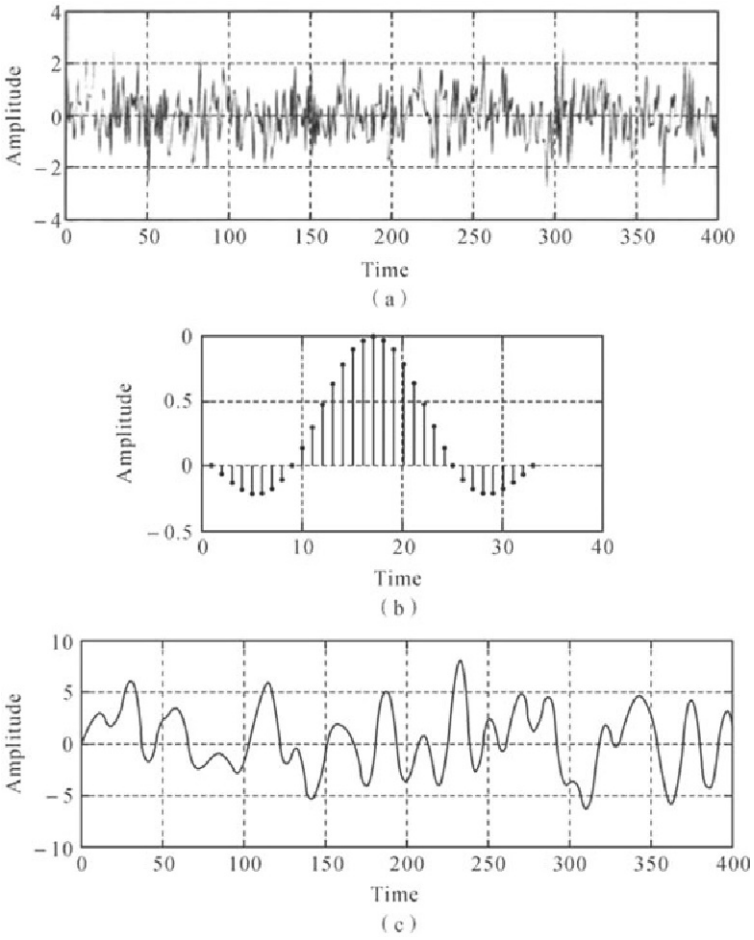


Fig. 7.14. Gaussian band limited noise filtered from Gaussian white noise. (a) Gaussian white noise; (b) Impulsc response; (c) Band limited Gaussian noise

Beamforming

Beamforming is the key technique in a sonar system. It is the basis of target bearing, ranging and classification. We have discussed the directivity function of a sonar system in Chapter 4. The theoretical assumption of directivity calculation is that the received signals from array elements are stationary random processes and the beamforming is the ensemble of the average result of the time delay-sum-square. This means the integration time of a system tends to the infinite. But it is of course impossible in a real situation. So there must have been some differences between the steady state directivity function and practical beamforming results. Fig. 7.15 shows an example of the ideal directivity function

and the system simulation results. Fig. 7.15(a) is the ideal directivity function. Fig. 7.15(b) is the real response for the target. It is easy to see the differences, particularly in the position far from the main lobe direction. The reasons for this are:

(1) Quantization error of time delay. The relative time delay value for the array elements is analogue, but in beamforming it is quantized. However fine the time delay lines are, the differences always existed.

(2) Quantization error of amplitude. The quantization error caused by the A / D converter will result in quantization noise and the beamforming value will be changed.

(3) Effect of finite integration time. Since the integration time of beamforming is finite, therefore the value of the direction function in any bearing angle is actually an average value, which randomly varies around the theoretical steady state response.

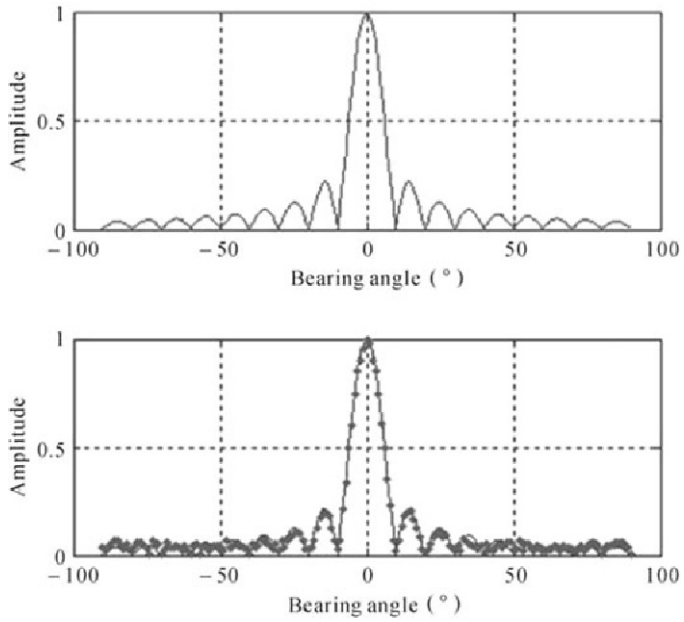


Fig. 7.15. Digital simulation of directivity function of linear array

In digital sonar design, system simulation for beamforming is necessary. The basic procedure is discussed in the following.

The amplitude layered operation is expressed as

$$y = \frac{\Delta}{2}\left[\sum_{i=1}^{N-1}(\operatorname{sgn}(x - i\Delta) + \operatorname{sgn}(x + i\Delta)) + 2\operatorname{sgn}(x)\right] \tag{7.26}$$

where x represents input and y represents output. The step of the layer is expressed

by

$$\Delta = \frac{V_0}{2^{b-1} - 1} \tag{7.27}$$

where V_0 represents the maximum possible value of the input signals and b is the bit number of quantization. The actual layer number is $N = 2^{b-1}$. Eq. (7.26) is convenient for theoretical analysis. But in practical calculation, when the bit number b is large, then the calculation is complicated. Therefore we prefer to use the following expression, which is easy to implement in simulation

$$y = \Delta[\mathrm{INT}(x / \Delta) + (\mathrm{sgn}(x) + 1) / 2] \tag{7.28}$$

Eq. (7.28) only has few singular points, for example when $x = 0$, $y = \Delta / 2$. Since this is a small probability event, we can neglect it. Therefore Eq. (7.28) is recommended in system simulation.

Now suppose the incidental signal is a sinusoidal signal with frequency f_0, the received signal in the i-th element is $s_i(t) = \cos[2\pi f_0(t + \tau_i(\theta_0))]$, where $\tau_i(\theta_0)$ represents the relative time delay of the i-th element. If the beamforming is in the direction θ, then the signal $s_i(t)$ should be delayed $\tau_i(\theta)$. We have

$$u_i(t) = s_i(t - \tau_i(\theta)) = \cos[2\pi f_0(t + \tau_i(\theta_0) - \tau_i(\theta))]$$

In considering the sampling frequency and the quantization error e_i of the time delay, we have

$$x_i(k) = u_i(kT_s + e_i) = \cos[2\pi f_0(kT_s + \tau_i(\theta_0) - \tau_i(\theta) + e_i)] \tag{7.29}$$

where $T_s = 1 / f_s$ represents the sampling duration, e_i represents the compensation error of the time delay. It is usually uniformly distributed in $[-T_s / 2,\ T_s / 2]$. The beamforming output is given by

$$D(\theta) = \left[\frac{1}{K} \sum_{k=1}^{K} \left(\sum_{i=1}^{N} x_i(k) \right)^2 \right]^{1/2} \tag{7.30}$$

When $k \to \infty$ in terms of an ergodic theorem we know that $D(\theta)$ tends to the steady state directivity function. In the practical simulation, we can replace $x_i(k)$

in Eq. (7.30) by $y_i(k)$ calculated from Eq. (7.28)

$$y_i(l) = \Delta[\mathrm{INT}(x_i(k)/\Delta) + (\mathrm{sgn}\, x_i(k) + 1)/2] \tag{7.31}$$

and we have

$$D(\theta) = \left[\frac{1}{K} \sum_{k=1}^{K} \left(\sum_{i=1}^{N} y_i(k) \right)^2 \right]^{1/2} \tag{7.32}$$

Eq. (7.31) is the final expression for system simulation of beamforming.

Iterative Algorithm of Adaptive Noise Canceling (ANC)

The ANC technique is extensively used in sonar signal processing. We have discussed this topic in Chapters 3 and 4. The application areas include interference suppression, line spectrum enhancement and channel equalization. The most popular adaptive algorithm is the steepest descend method. This algorithm is based on the LMS criteria. The iterative equation (see also Eq. (3.124)) is

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \mu(\boldsymbol{x}^{\mathrm{T}}(k)\boldsymbol{w}(k) - d(k))\boldsymbol{x}(k) \tag{7.33}$$

where $\boldsymbol{w}(k) = [w(0), \ldots, w(N-1)]^{\mathrm{T}}$ is the k-th iteration weight vector. And $\boldsymbol{x}(k) = [x(0), \ldots, x(N-1)]^{\mathrm{T}}$ is the signal vector. $d(k)$ is the desired signal. It is a signal correlated with some components in the input signal. μ is a factor to control the convergence of the iteration procedure. The error signal at moment k is

$$\varepsilon^2(k) = [\boldsymbol{x}^{\mathrm{T}}(k)\boldsymbol{w}(k) - d(k)]^2 \tag{7.34}$$

This is an important quantity for evaluating the iterative process, when the value of $\varepsilon(k)$ tends to the stationary. This indicates that $\boldsymbol{w}(k)$ is close to the optimum value $\boldsymbol{w}_{\mathrm{opt}}$.

The relationship between $\varepsilon(k)$ and k is called learning behavior, which describes the convergence behavior of adaptive processing. Fig. 7.16 is an example of a learning curve. From this figure we can see that when the iteration number is close to 8,000, the adaptive process tends to a steady state. If the sampling duration is 50 μs, then it means the learning process will be completed after 0.4 s. This kind of convergent speed is acceptable in underwater acoustic signal processing.

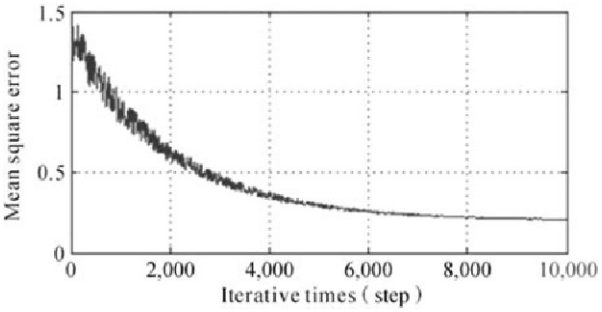


Fig. 7.16. Learning curve of adaptive filtering

7.3.3 Simulation of Post Processing

There are various algorithms in post processing; we will just discuss post beamforming processing. It is necessary to perform time integration on each beam, therefore the information of target bearing can be extracted. There are several methods of time integration that can be adopted in post processing. One is square detection, i.e., energy detection, the other is absolute detection. It is proved that these two methods are equivalent in some cases.

Suppose $x(k)$ is the data series of some beam. The square detection is

$$y(k) = \left[\frac{1}{M} \sum_{i=0}^{M-1} x^2(k-i) \right]^{1/2} \tag{7.35}$$

In the expression of square detection, there is a multiplication operation which costs a lot of time. If we use absolute detection instead of square detection, the arithmetical operation will save time. Let

$$y(k) = \frac{1}{M} \sum_{i=0}^{M-1} | x(k-i) | \tag{7.36}$$

If $x(k)$ is a Gaussian random signal, then the square detection and absolute detection are equivalent. It means $[E(x^2(k))]^{1/2}$ and $E[|x(k)|]$ are equivalent. In fact, if the probability density function of $x(k)$ is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-x^2 / (2\sigma^2))$$

then for square detection

$$y_1 = [E(x^2(k))]^{1/2} = \left[\int_{-\infty}^{\infty} x^2 p(x)\mathrm{d}x\right]^{1/2} = \sigma$$

For absolute detection,

$$y_2 = E(|x(k)|) = 2\int_0^{\infty} xp(x)\mathrm{d}x = \sqrt{\frac{2}{\pi}}\sigma$$

We see that y_1 and y_2 are differing up to a constant factor $\sqrt{2/\pi}$.

The output format of post processing depends on the algorithm of average expression. The frequent use methods are local average and recursive average (see Eqs. (2.89) and (2.90)). If the average result is renewed in each sampling duration, we obtain

$$z(k) = \frac{1}{M}\sum_{i=1}^{M-1} y(k-i) \tag{7.37}$$

where $y(k)$ represents the k-th sample of post average results and $z(k)$ is the output data to time / bearing display. Eq. (7.37) can be rewritten as

$$z(k) = z(k-1) + y(k)/M - y(k-M)/M \tag{7.38}$$

It is easier to implement in hardware.

7.3.4 Specific Software for System Simulation

There is some general purpose software on the signal processing market. It provides many useful signal processing techniques and tools which can be used in system simulation in sonar design, for example MatLab, LabView, etc.

7.4 Design of Sonar Signal Simulator

The system simulation technique described in previous sections is necessary during the period of sonar design. Once the sonar system is manufactured, one should test the sonar system to see if the technical and static performances satisfy the predicted expectations. In the development of modern digital sonar, a new field is fast growing which is the manufacture of a sonar simulator. The main task of a sonar simulator is to provide complete verification tools for sonar testing. The

sonar simulator can also be used to train the sonar operator, since it can provide a virtual sonar environment which is very close to the real situation.

The kernel component of a sonar simulator is the array signal generator. It provides necessary signals received by an array system. Sometimes it is a difficult task, since the analogue time delay of a multiple channel requires considerable cost in terms of hardware and software. In this section, the basic theory and techniques of a general purpose sonar simulator design will be discussed.

7.4.1 General Concept of Sonar Simulator

The main mission of a sonar simulator is to create, as accurately as possible, the sonar environment field and, if necessary, the sonar simulator should offer specific interfaces to the sonar system. Fig. 7.17 illustrates the relationship between a sonar system and a sonar simulator. We can see that the sonar simulator can provide the analogue signal output, digital signal output and the ranging signals. By using this kind of sonar simulators, sonar system testing will be convenient.

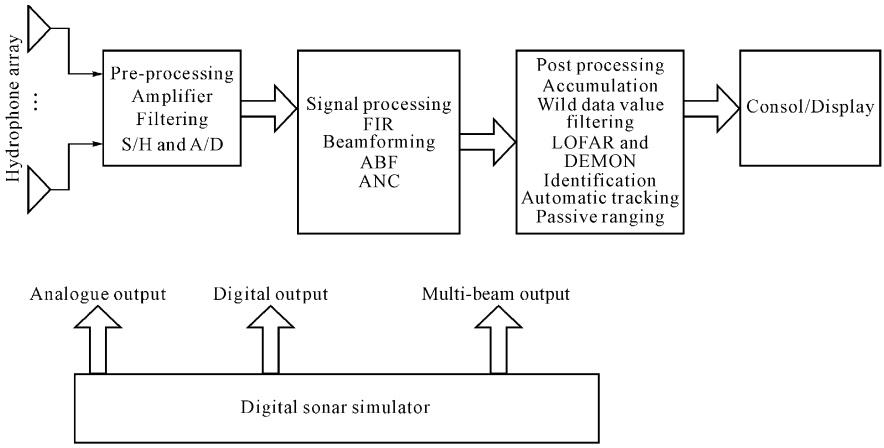


Fig. 7.17. Sonar and sonar simulator

Because of the characteristics of underwater signal propagation, the requirements of time delay accuracy are very high. Therefore the hardware cost is considerable. For example, if the bearing accuracy of a sonar system is 0.1°, then the sonar simulator should offer an array signal with an accuracy much higher than 0.1°. Otherwise it is impossible to test the sonar system if it reaches a bearing accuracy of 0.1°.

Example

Suppose the distance of two element is 2 m. If the incidental angle of the signal is 0°, the time delay between the two elements is 0 μs. If the incidental

angle is 0.1°, then the time delay difference becomes 1.16 μs. Therefore the signals generated by the sonar simulator should at least be in the order of much less than 1 μs.

The array signals generated by the sonar simulator actually form a virtual array, but this array is not always the same as a true array. The finer the accuracy of the time delay, the better the coincidence. Therefore it is worth pointing out that sufficient accuracy in time delay simulation is absolutely the most important aspect in the design of a sonar simulator.

Fig. 7.18 shows the block diagram of sonar simulator functions. This is a general purpose sonar simulator, which is designed to test passive sonar. The basic parameters can be selected and programmed, including the array configuration, the number of channels, the number of targets, frequency band, target power spectrum and target moving trace.

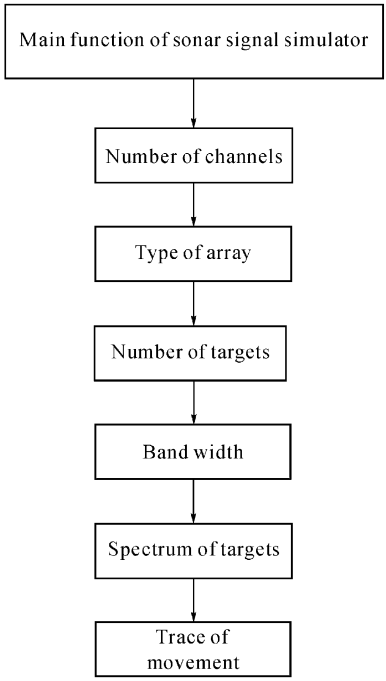


Fig. 7.18. Main function of general purpose sonar signal simulator

7.4.2 Design Principle of General Purpose Sonar Simulator

The key function of a sonar simulator is to generate multiple channel signals. Suppose the number of channels to be simulated is N. The i-th signal to be generated is $x_i(t) = x(t - \tau_i(\theta))$ $(i = 1, \ ..., \ N)$.

The calculation of $\tau_i(\theta)$ is based on the array configuration. Generally speaking, the incidental signal can be considered as a plane wave, but in the case of ranging signal simulation, the sound wave is transmitted from a point source and the acoustic wave is not a plane wave. If the minimum quantization duration of the time delay value is selected as Δ, then the time delay value $\tau_i(\theta)$ should be converted to the integer times of Δ. We have $x(k-k_1)$, ..., $x(k-k_N)$.

In underwater acoustic application, Δ is at least 1 μs, or even less. It is necessary to show that the quantization duration of the time delay for a sonar simulator is totally different from that in beamforming. The former is much less than the latter. Fig. 7.19 shows how to generate the array signal by using a shift register.

Since Δ is very small, for a large array the length of the delay line will be very long and this is a problem with sonar simulator design.

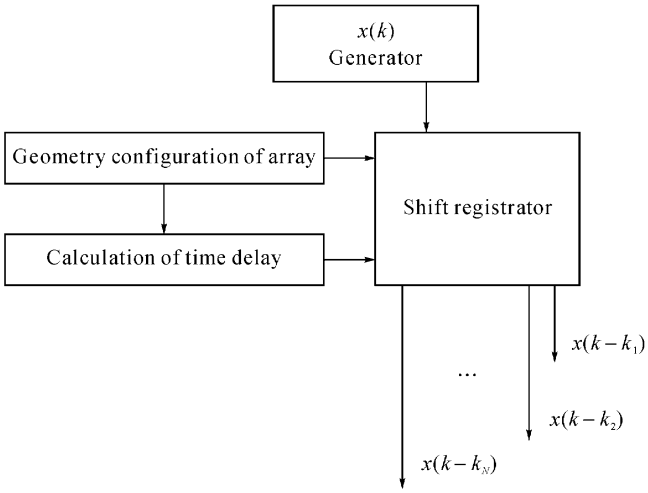


Fig. 7.19. An algorithm for generating an array signal

Fig. 7.20 illustrates the block diagram of function selections of a sonar simulator. This is actually the initialization process in simulator running.

(1) Number of channels N: 2 – 128.
(2) Array type: Equidistant linear array (input the distance of element); Circle array (input radius value); Arc array (input the working sector); Conformal array (input the coordinates of each element); Random array (input the coordinates of each element).
(3) Number of targets TN: up to 16. Frequency band [W_1, W_2] and power spectrum $P_i(f)$ (i=1, …, TN) of targets.
(4) Power spectrum of ambient noise.
(5) Moving trace of targets: Straight line; Circle; Trajectory of torpedo.

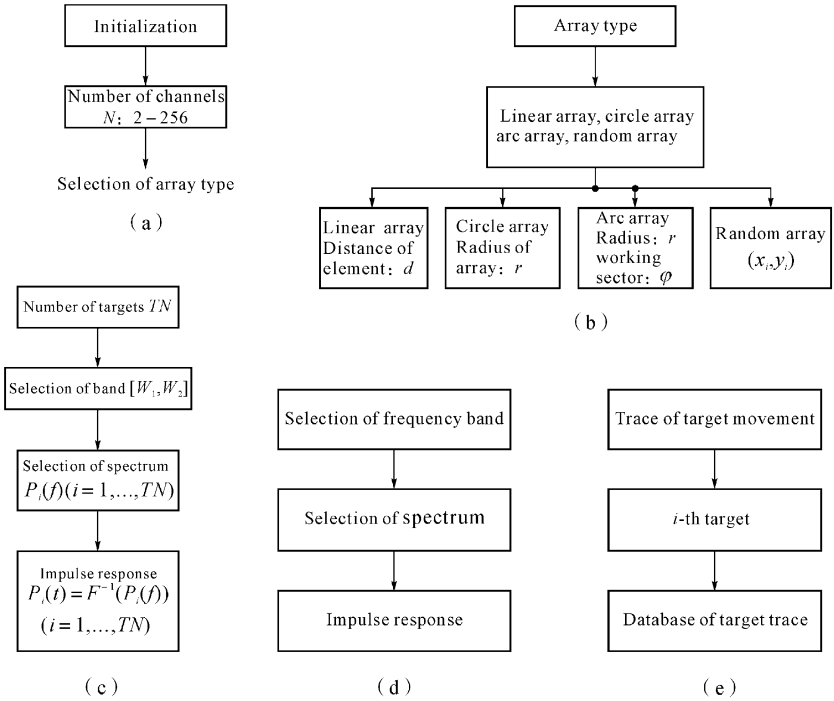


Fig. 7.20. Detailed parameter information of a sonar simulator. (a) Selection of channel number; (b) Selection of array type; (c) Selection of targets; (d) Selection of ambient noise spectrum; (e) Moving trace of targets

7.4.3 Algorithm for Generating Infinite Non-cyclic Random Number

The generating of a random number is the key point in system simulation. It is the basis of generating a good characteristic random number in a sonar simulator. We are going to introduce a simple, effective method for generating an infinite, non-cyclic random number. This method is proposed by Rader and applied in digital signal applications by Oppenheim[13]. The algorithm proposed by Rader is different to the linear shift feedback register; it is a non-linear feedback shift register. The length of the shift register is usually longer than 30 stages. The seeds of the random number are preset in the register, supposed to be uniformly distributed in the segment $(-1/2,\ 1/2)$. The $(k+1)$-th random number is produced by module $1/2$ addition. The algorithm is as follows:

$$x(k+1)=\begin{cases} x(k)+x(1) & -1/2 \le x(k)+x(1) \le 1/2 \\ x(k)+x(1)-1 & x(k)+x(1) > 1/2 \\ x(k)+x(1)+1 & x(k)+x(1) < -1/2 \end{cases} \tag{7.39}$$

The theory and application proves that the random number generated by this algorithm has excellent statistical characteristics: uniform distribution, independent, non-periodic. Fig. 7.21 is the block diagram of this algorithm.

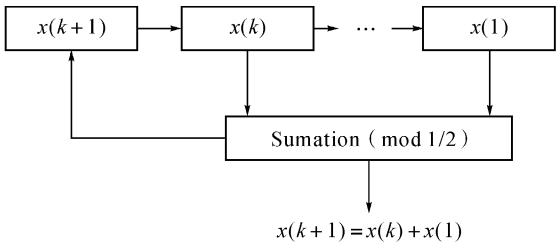


Fig. 7.21. A method for generating an infinite non-cyclic random number

The obvious advantage of this algorithm is that it is easy to realize in hardware. Figs.7.22 and 7.23 illustrate the flowchart of Rader's algorithm. This flow chart is easy to implement using DSP chips. Once we obtain this kind of infinite, non-cyclic random numbers, it is easy to produce another random number with assigned frequency performance. The method adopted here is FIR filtering (see Fig. 7.24). The interpolation technique is used after the FIR filter to get finer sampling.

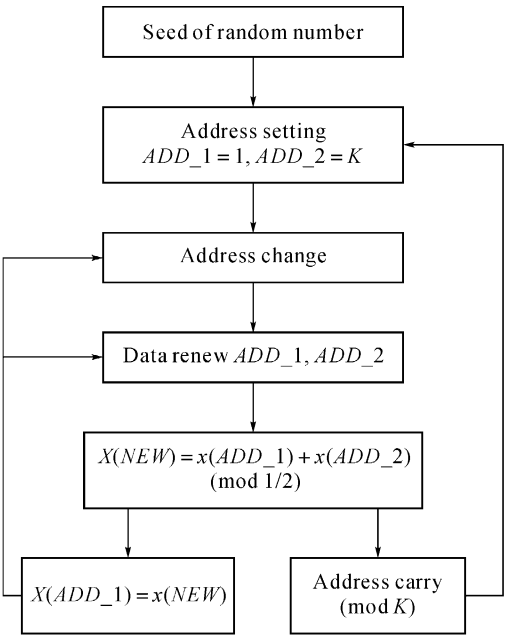


Fig. 7.22. A method for fast generation of white noise by using SRAM

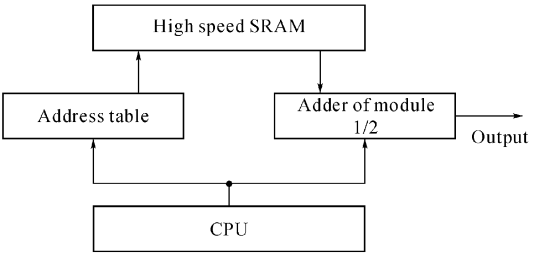


Fig. 7.23. Algorithm for generating multiple channel white noise

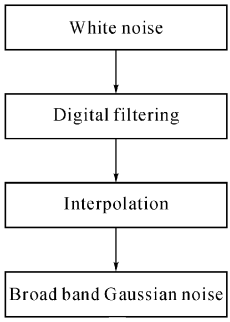


Fig. 7.24. Algorithm for generating broad band Gaussian noise

7.4.4 Generating of Array Signals

Based on the infinite non-cyclic random number, we can obtain the necessary array signals with suitable time delay. As an example, we will discuss a discrete linear array. Suppose the distance between adjacent elements is d. We want to generate the array signals with an incidental angle θ (Fig. 7.25). Hydrophone H_1 is selected as the reference hydrophone. The relative time delay of i-th signal is

$$\tau_i = (i-1)d \sin\theta / c \tag{7.40}$$

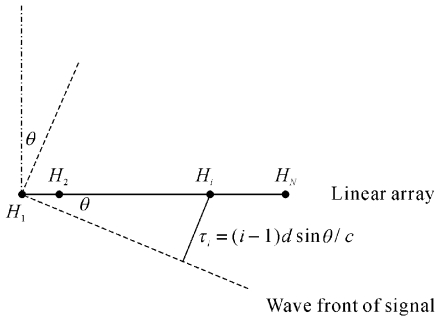


Fig. 7.25. Simulation of line array signal

The time delay accuracy of a simulated signal should be at least in the order of $1\ \mu s$; it is necessary to have a high speed SRAM (Fig. 7.23) or interpolation algorithm (Fig. 7.24).

The method described above is the case of a plane wave simulator. For the ranging signal simulation, the algorithm is different. Fig. 7.26 gives the basic scenario of ranging signal simulation. The signals we want to simulate are three sub-array signals after beamforming. Suppose the reference hydrophone is H_1.

We have

$$\tau_{12} = [d_1 \sin\varphi + d_1^2 \cos^2\varphi / (2R)] / c \tag{7.41}$$

$$\tau_{23} = [d_2 \sin(\varphi+\alpha) + d_2^2 \cos^2(\varphi+\alpha) / (2R)] / c \tag{7.42}$$

In terms of Eqs. (7.41) and (7.42), we can calculate the time delay between three sub-array beams, which steer in the direction of the target.

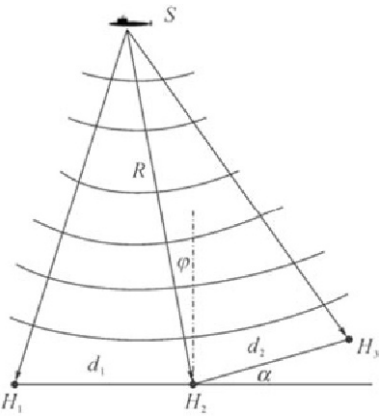


Fig. 7.26. Simulation of passive ranging signal

Now we are going to discuss the problem of the adjustment of the SNR. The signal of the array elements consists of signals, interference and noise. They are separately generated and are added together to obtain synthetic array signals. The SNR has to be adjusted (Fig. 7.27). The method for adjusting the SNR is to change the amplitude of the signal or noise. One thing we have to note is the dynamic of signals. It is necessary to keep the dynamic range unchanged when we adjust the SNR.

Example

In a fixed point arithmetic sonar simulator, suppose the maximum possible amplitude of a signal is 12 bits and the noise is also 12 bits. If we take 2^8 as the average level of noise then, when the signal level is 2^8, the signal-to-noise ratio will be 0 dB. In this time the maximum signal-to-noise ratio we can obtain is 2^4,

about 24 dB. But the minimum SNR is not 2^{-8}, since in this situation the distortion of the signal will be very serious.

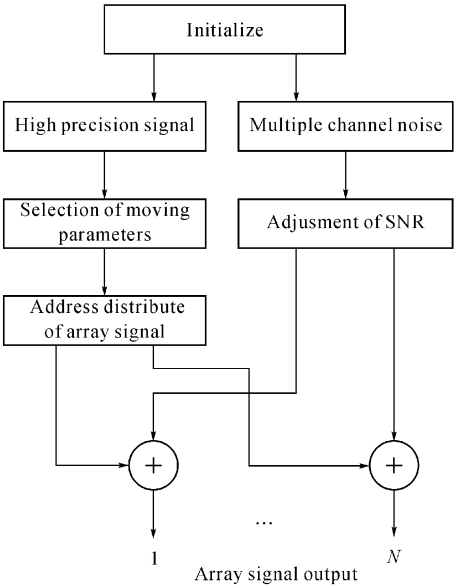


Fig. 7.27. Synthesis of array signals

The method for a passive sonar signal simulator described above can be also used in an active sonar signal simulator, if little changes are carried out. Fig. 7.28 shows the block diagram of an active sonar simulator design.

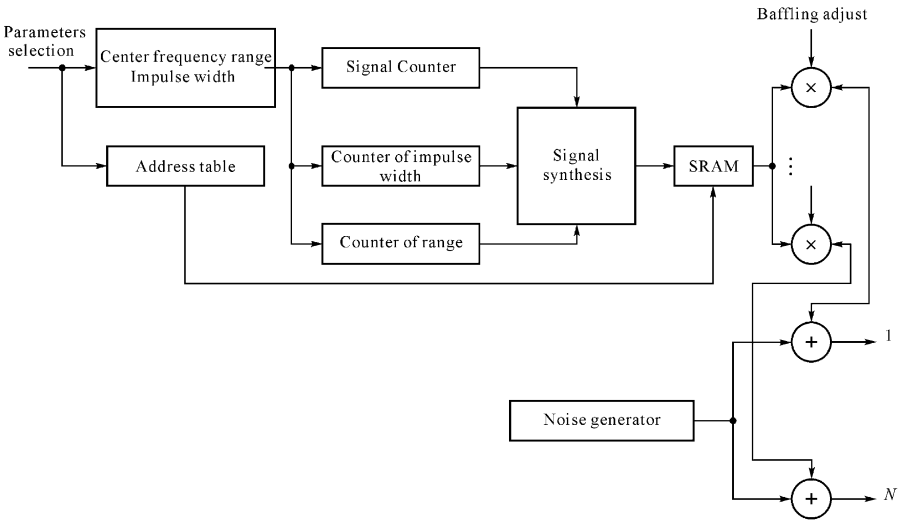


Fig. 7.28. Block diagram of synthesis of active sonar signal simulation

7.4.5 An Application Example of Sonar Simulator Design

We will then briefly introduce an example of sonar simulator design and the test results of this sonar simulator.

The non-linear shift register is used to generate a random number. The length of the register is selected as $k = 50$. The seeds of the random number preset in the register are listed in Table 7.2.

Table 7.2 Example of 50 random number seeds

+0.4073	+0.2862	−0.3895	−0.1755	−0.3782
−0.0760	−0.0536	+0.0574	+0.1664	−0.4867
+0.4161	−0.0994	−0.1997	−0.3704	+0.3282
+0.4280	−0.4350	−0.1780	−0.0114	−0.3127
+0.0304	−0.3399	−0.2882	−0.2533	+0.1074
−0.2893	−0.1791	−0.3286	−0.4233	−0.4741
−0.3752	−0.4880	+0.2864	−0.3249	+0.4290
+0.3514	+0.0116	+0.2939	+0.2494	−0.1603
−0.4054	+0.2226	+0.2939	+0.2494	−0.1603
−0.2663	−0.0135	+0.4238	−0.4401	−0.1571

The random number generated by this register has very good performance (Fig. 7.29). Suppose the sampling duration is 100 μs. Fig. 7.30 illustrates the auto-correlation function of this random number series. With the increase in integration time, the fluctuation of the auto-correlation value is considerably decreased. Based on this random number series, 64 channels of independent random numbers are generated. They also have very good statistical characteristics. The cross-correlation function of any two channels is shown in Fig. 7.31.

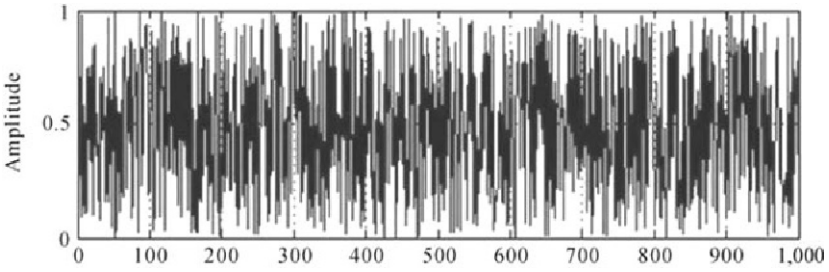


Fig. 7.29. Example of uniformly distributed random number

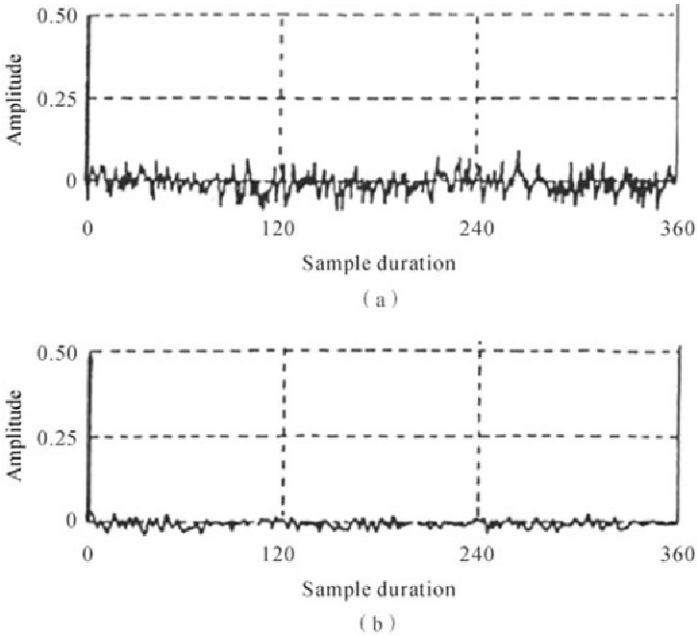


Fig. 7.30. Example of auto-correlation function of random number. (a) Integrationg time: 36 ms; (b) Integrationg time: 1 s

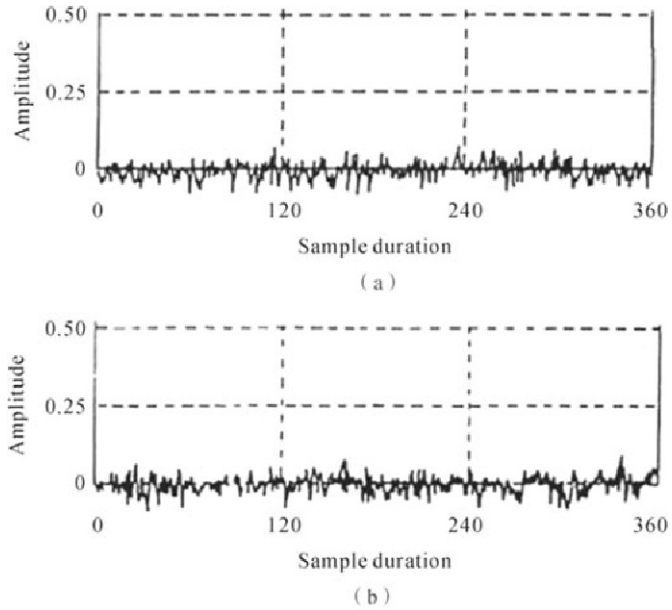


Fig. 7.31. Example of cross correlation function of random number. (a) No.1 and No.2 channels; (b) No.1 and No.25 channels

We then describe how to generate a superfine time delay signal. We have repeatedly stressed that array signals with precise time delay value are the basis of a sonar simulator. If we are unable, due to some limiting conditions, to produce a superfine time delay signal with our hardware, then we can consider using software or a combination of hardware and software instead. Fig. 7.31 illustrates an algorithm for generating a superfine time delay signal using software.

In this method, we use the interpolation algorithm to generate a 0.1 μs accuracy signal based on a 1 μs accuracy signal. In this algorithm the array signals are generated in terms of an address register with 0.1 μs accuracy. De-sampling the generated signal, we obtain the necessary superfine time delay signal.

For example, if we want to get a signal $x(t)$, $x(t+\tau)$ and $\tau = 28.5$ μs, the original signal series we already have is: $x(T_s)$, $x(2T_s)$, …, $T_s = 1$ μs. The interpolation result is $x(T_s')$, $x(2T_s')$, ..., $T_s' = 0.1$ μs .We take two series: $x(T_s)$, $x(20T_s)$, $x(40T_s)$, ... ; $x(285T_s')$, $x(485T_s')$, $x(685T_s')$, ... It has 28.5 μs time delay difference.

The zero padding algorithm described in Chapter 2 can be used for interpolation processing and we will not discuss it again. That means the superfine time delay signal can be generated by a combination of software and hardware. Therefore the hardware cost will be considerably reduced.

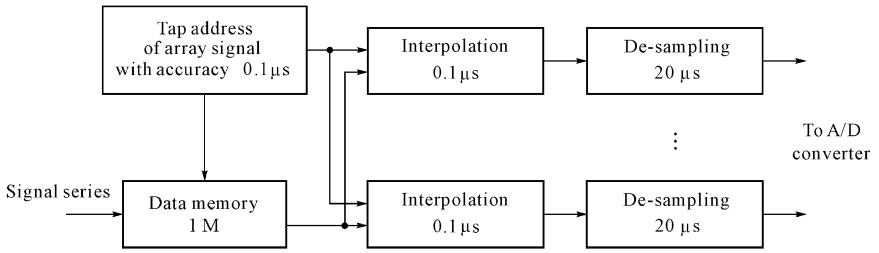


Fig. 7.32. A method for generating an array signal with superfine delay time

References

[1] Naylor, T. H., Balintfy, J. L., Burdick, D. S., et al.: Computer Simulation Techniques. Wiley, New York (1969)

[2] Lewis, T. G., Smith, B. J.: Computer Principles of Modeling and Simulation. Houghton Mifflin Comp., Boston (1979)

[3] Ruckdeschel, F. R.: BASIC Scientific Subroutine, Vol.II. Mc-Graw Hill, New York (1981)

[4] Gordon, G.: System Simulation. Prentice Hall, New Jersey (1978)

[5] Camdy, J. V.: Model based signal processing in the ocean. IEEE Oceanic Eng. Society News Letter, XXXV-3, pp. 199-205 (2000)

[6] Anarld, C. R.: Digital simulation of a conformal DIMUS sonar system. AD

277577 (1962)
[7] Higgins, T. M., Turriff, A. E., Patrone, D. M.: Simulation-based undersea warfare assessment. John Hopkins APL Technical Digest, 23-4, pp. 396-403 (2002)
[8] Mansour, M. A., Smith, B. V., Edwards, J. A.: PC-based real-time active sonar simulator, Radar. Sonar and Navigation IEE Proc., 144-4, pp. 227-233 (1977)
[9] Shreider, A.: The Monte Carlo Method. Pergamon Press, Oxford (1966)
[10] Knight, W. C., Pridham,R. G., Kay, S. M.: Digital signal processing for sonar. In: Proc. of IEEE, 69, pp. 1451-1506 (1981)
[11] Hudson, J. E.: Monte Carlo simulation of an active sonar, Radio Electronic Eng. Vol.40 (1970), 265-271
[12] Gordon, R. L.: A sonar design tool: Multi-channel Monte Carlo simulation. AD / A 060234 (1968)
[13] Oppenheim, A. V. (Eds.): Applications of Digital Signal Processing. Prentice Hall, New Jersey (1978)

8

Introduction of Typical Modern Digital Sonar

In the previous chapters, the basic principles of digital sonar design have been discussed. In applying these theoretical results and techniques to design specific sonar, there are many questions and problems we face. These include unified design of the wet end and the dry end, limitations of the platform for sonar installation and software / hardware testing, etc.

Only when a specific sonar design is completed, is it possible to get a better understanding of the theory and technique. In the sonar design process, some new topics and issues may appear, and the solving of these problems will become a motivation of the new theory and technique. In a word, practice-theory-practice is the most effective technique for developing modern digital sonar.

In this chapter a new area of sonar design will be introduced—model-based sonar[111,120-126]. In the development of this new concept, robust signal processing theory[115-119] and matched filtering theory in the field of ocean acoustics play a very important role [114-118].

Some typical modern digital sonar systems are briefly discussed, including towed array sonar[5-33,42], shore-based sonar[127,128], flank array sonar[73,74] and synthetic aperture sonar[46,47,76-83,99]. From these descriptions of digital sonar we can see the main scope of modern digital sonar.

8.1 Concept of Model-Based Sonar

The performance of digital sonar strongly depends on the ocean environment. The real underwater acoustic channel is spatial and time varying. The performance of digital sonar often degrades due to model mismatch. The system gain, design-based ideal model usually cannot work in the real ocean environment because of model mismatch. Since the 1980's, many scientists and researchers have been interested in studying the effect of the ocean environment. They proposed that in sonar design the ocean environment should be taken into account in the design procedure, and the ocean environment should become a part of the sonar system.

The concept of model-based sonar is proposed in this background. This concept can be understood in different ways. On the one hand, it is necessary to design sonar which can meet the requirements of multiple models, i.e., when the environment of the sonar system changes then the sonar system can adaptively turn to the new model. On the other hand, it is necessary to design the sonar system more robustly, i.e., the performance of a sonar system may not be optimum, but is robust enough. It is not sensitive to model mismatch. Studies in capturing the uncertainties in the ocean environment form one of the interesting topics in this field.

Although the concept of model based sonar is new in sonar design, it already attracts much research attention in underwater acoustics.

8.1.1 Basic Concept of Model-Based Sonar

Target detection and location has been the basic mission of a sonar system for a long time. With the development of stealth technology, the submarine becomes more and more quiet. Therefore, in order to face the new challenge, it is necessary to study new methods for detecting a quiet submarine. One method is to include the ocean environment in the sonar design.

Acoustic wave propagation in the underwater acoustic channel is very complicated. Ambient noise and the conditions of acoustic signal propagation are spatial and time varying. In particular, in shallow water the acoustic channel is bounded by the sea surface and sea bottom and the channel characteristics become more complex.

Traditional sonar is based on optimum detection theory; it has good performance in ideal conditions. But once the real situation differs from the ideal assumption, the performance of a sonar system may be degraded seriously. Fig. 8.1 illustrates the fixed model sonar system and the model-based sonar system.

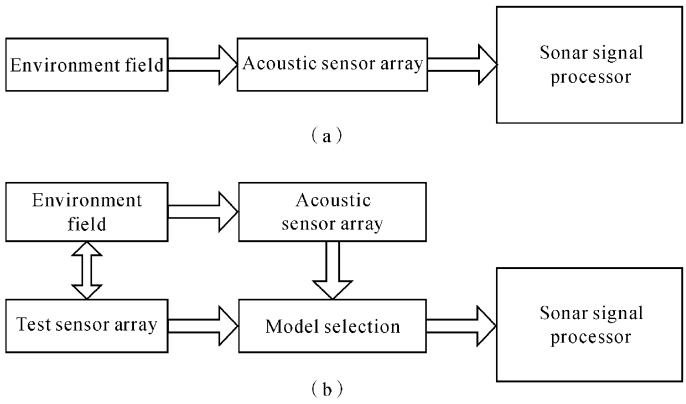


Fig. 8.1. Classification of sonar signal processing model. (a) Traditional sonar system; (b) Model based sonar system

For the fixed model sonar system, the design is based on a certain ocean environment. Modeling is necessary, but the mechanism of sonar signal processing is unchanged once the design is completed. MBP (model-based processing) is different. It receives the environmental information and, based on the changes in the ocean environment, adaptively adjusts the sonar parameter, even the signal processing schema, to meet these changes. Therefore, the real time information of sonar environment can be feedback to sonar system and influence the sonar work.

Scientists in ocean acoustics are interested in the earlier work in model-based signal processing and MFP matched field processing. In 1966, Clay proposed a matched field method for depth locating by using a wave guide theory. In 1977, Hinich proposed that one should include the propagation model in the signal processing framework. The brief history of MFP can be found in the literature[113].

Fig. 8.2 shows the relationship of the signal processor of model-based sonar to the ocean environment. The sonar itself must establish various models, including ambient noise, sea bottom and sound speed profile, etc. Fig. 8.3 gives the descriptions of field parameters. The block diagram explains how to transfer the test results to the model estimates. The sound pressure field $p(z_l)$, sound speed field $c(z_l)$ and sound speed gradient field $\theta(z_l)$ should be transferred to the estimate of mode wave $\hat{\phi}(z_l, \theta)$, the estimate $\hat{p}(z_l, \theta)$ of sound speed and the estimate $\hat{\theta}(z_l, \theta)$ of sound speed gradient. And also, to calculate the predicted error $\varepsilon(z_l, \theta)$, it is necessary to modify it in real time.

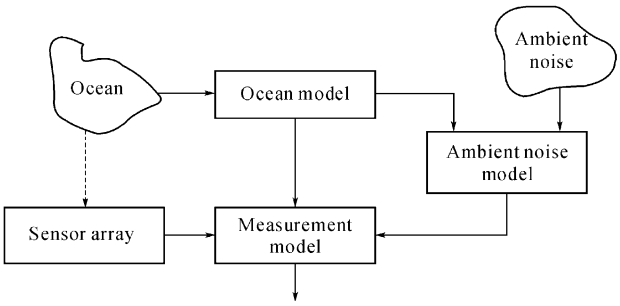


Fig. 8.2. Basic data source of model-based sonar

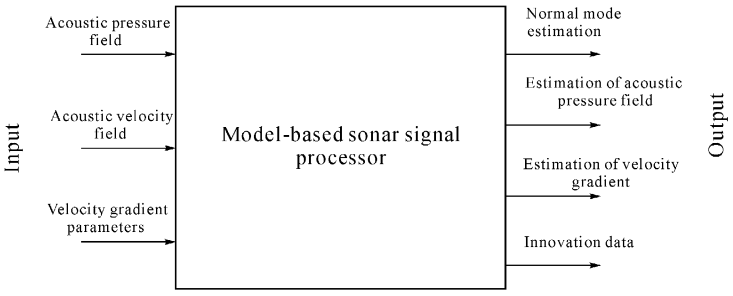


Fig. 8.3. Model-based ocean acoustics processing system

8.1.2 Robust Signal Processing

As the kernel of a sonar system, the signal processing unit plays a very important role in target detection. The sonar detection performance strongly depends on the signal processing. The traditional sonar system is designed based on the assumption that the ambient noise is uniform, homogeneous and isotropic. The signal to be detected is, or at least in a short time duration, a stationary Gaussian random process. The system gain of this kind of sonar is, in a certain sense, proportional to the number of array elements and the integration time.

But the real ambient noise field of the ocean is neither uniform, nor homogeneous and isotropic. Therefore, the value of the system gain will have some differences from the real situation. Sometimes these kinds of differences will produce serious results for the sonar operation. These will not only appear in active sonar detection, but also in passive sonar detection.

The most typical example is the matched filter. Theoretically speaking, the matched filter is the optimum linear filter in the detection of a deterministic signal in the noise background. Since the echo signal is usually distorted in frequency domain and time domain, the matched filter does not even work in a real ocean environment.

Fig. 8.4 is an explanation of the concept of robust signal processing. The horizontal coordinate represents the model mismatch; this is hard to describe in quantity. But we can understand the basic meaning of robust signal processing. That means that in an ideal situation the system gain from robust signal processing is less than that from optimum signal processing. But robust signal processing is not sensitive to model mismatch, i.e., in the case of model mismatch, the system gain of robust signal processing is larger than that of optimum signal processing, which is what we can expect most frequently. This is just the meaning of "robust".

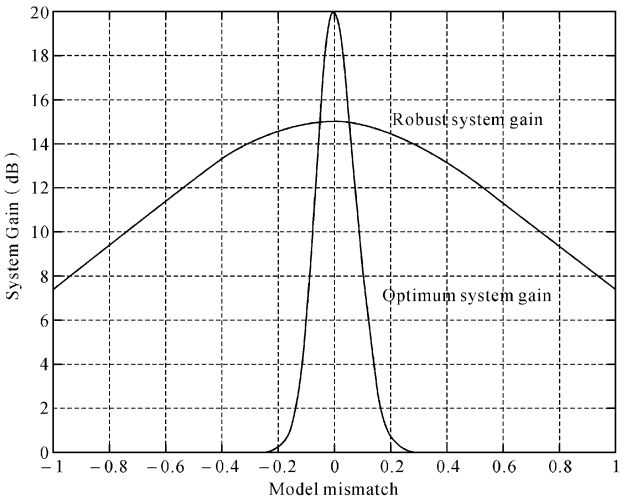


Fig. 8.4. Concept of robust detection

The theoretical description of robust signal processing of model-based sonar can be derived as follows.

Suppose H is a possible sonar design space, the element of this space is the possible signal processing method. Q is a set consisting of various models. Define a metric function $M(h, q)$, $h \in H$, $q \in Q$, which is used to measure the system performance using model q and signal processing method h. The traditional sonar design is for a fixed $q_0 \in Q$. Find $h_0 \in H$ such that

$$M(h_0, q_0) = \max_{h \in H} M(h, q_0) \tag{8.1}$$

That means $h_0 \in H$ is the optimum signal processing method for model $q_0 \in Q$.

If the real ocean environment model q_0 is varying, it can be changed in a subset P of Q, i.e., the element of model $p \in P \subset Q$. In this situation, how to choose the signal processing method? A reasonable strategy is to use the maximum-minimum method in game theory. That means for any $h \in H$ find q, which minimizes $M(h, q)$. And then find h, which maximizes the value of $\min_{q \in P} M(h, q)$, i.e.,

$$\max_{h \in H} \min_{q \in P} M(h, q) \tag{8.2}$$

The robust signal processing method h_R is the signal processing method which satisfies the following equation

$$\min_{q \in P} M(h_R, q) = \max_{h \in H} \min_{q \in P} M(h, q) \tag{8.3}$$

This strategy is easy to understand. In the real ocean environment, whatever signal processing method we have chosen, we should imagine that we may be facing the worst situation (model), and then we must find the optimum signal processing method.

8.1.3 Signal Processing System of Model-Based Sonar

The MBP described above for model-based sonar is based on very preliminary research results. The theory and application is far from being well established as the standard. There is no recognized theoretical framework but, at least, the characteristics of model-based sonar should be:

(1) Recursive;

(2) Uncertainty of the noise field should be considered;

(3) Possibility of linear and non-linear time / spatial variation should be considered;
(4) Necessity to have on-line test ability in the iteration process;
(5) Necessity to have ability of real time monitoring. The results of real time monitoring must be compared with the predicted value;
(6) The system must be able to estimate the metric function, which is used to describe the system performance.

The acoustic modeling for the acoustic field includes the following four models: ray theory of sound field, spectrum integration, mode theory, parabolic approximation of wave propagation.

8.1.4 Matched Field Filtering

MFP is one of the algorithms of MBP. It is also a hot topic which attracts much interest in underwater acoustics. The main idea of MFP is to use the information from the sound pressure field in the signal processing to obtain range and depth information of the target. The calculation results are compared with the results from mode theory. The likelihood criterion is used to modify the final results of the target range and depth.

Fig. 8.5 is a simplified model of MFP. It is natural to choose the likelihood criterion in MFP. Suppose we want to estimate the value of parameter θ and we have the observation data x, which could be a scale quantity or a vector. How can we estimate parameter θ based on the observation x? It is reasonable to consider, in terms of the value x, in what kind of parameter θ it is most possible to get observation data x. Therefore the value θ, which can make $p(x \mid \theta)$ maximum, is the right choice.

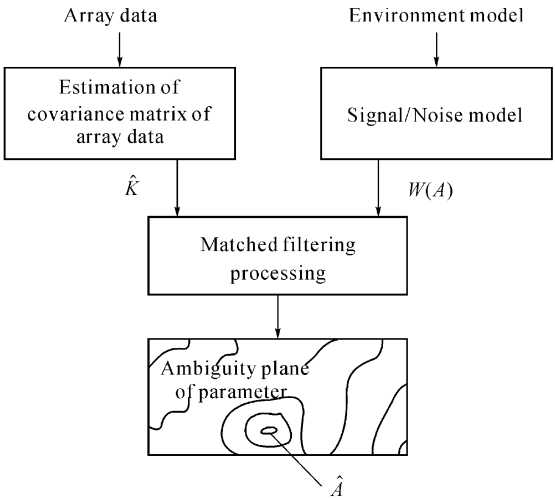


Fig. 8.5. Block diagram of matched filtering processing

Based on this explanation, it is easy to establish the operation framework of MFP. Suppose there is a sound source with unknown range and depth. It is necessary to determine the range and depth based on the signal transmitted from this source. The likelihood criterion shows that we should consider this question in an inverse fashion: since we have already obtained the sound field from this source, what kind of sound source can produce this kind of sound field? The calculation procedure can be divided into three steps.

(1) Collect $N \geq 1$ data of hydrophones;
(2) Choose a known propagation model, calculate the sound field produced by different denominated range and depth;
(3) Based on MFP, find the correlation behavior between the predicted sound field and observed sound field. The range and depth which maximize the correlation function are just the range and depth we are finding.

Example[122]

The depth of the sound source is 100 m, the distance between source and receiver is 45 km. The frequency of the transmitted signal is 10 Hz. There are equally spaced linear arrays with 15 hydrophones and the distance of adjacent elements is 75 m. By using a Kraken mode model, the increment in distance is 0.15 km. From 0 to 150 km there are 1,000 range points; The increment in depth is 10 m, From 10 to 250 m there are 25 points. Therefore the total grids of the sound field are 1,000×25 = 25,000. The MFP calculation results show that the sound source is at distance of 44.5 km and a depth of 100 m.

8.2 Linear Towed Array Sonar

Linear towed array sonar was developed in the 1970s, and the history of the application of a towed array can be traced back to 1917[5].

Since the development of the stealth submarine, the weak signal detection problem has attracted more research attention. Many scientists and engineers have turned to the application of towed array sonar, which has been successfully used in oil / gas exploration in offshore waters. The towed array can be far from the platform; the interference from propellers is considerably reduced. This is really a very brave decision, since the shape of the towed array is not a straight line; beamforming faces serious challenges. On the other hand, the maneuvers of a towed array are affected by the long towed cable and acoustic module; it is necessary to make some kind of innovation in vessel design.

This practice has already proved that the advantages of a towed array are attractive and the disadvantages can be overcome in some ways.

It is reported that the active / passive towed array is the only way to make up for the inability to detect a submarine[123]. In this section we will introduce the basic technique of a towed array sonar.

8.2.1 Brief History of Linear Towed Array Sonar

Differing from most sonar systems, the extensive application of towed array sonar began in the civilian area, mainly in oil / gas exploration in offshore waters. The seismic exploration method requires high resolution oceanographic image sonar and a high efficiency signal processing algorithm. The towed array sonar meets the requirements of seismic exploration. The Western Company of the USA, Plessey Inc. of the UK and the Simrad Corporation of Norway developed seismic towed array systems which use a high pressure air gun to transmit a low frequency signal to the sea floor. The towed array sonar receives the echo signals from the sea floor. The recording signal from multiple line arrays is processed to obtain information about sea floor resources. The frequency used in seismic exploration is lower than 120 Hz.

In the utilization of a linear towed array for petroleum exploration, one knows that these kinds of towed arrays are able to detect the radiated noise of a submarine. Therefore, the technique used in seismic exploration was transplanted to military applications. Since the 1960s, the US navy has begun to use a linear towed array on surface ships. Of course, the technique has been changed to meet military applications. The frequency used on surface ships is much higher than in seismic arrays. The earliest type was AN / SQR-14. It is a long distance surveillance sonar manufactured in 1975. Thereafter, different types of towed array sonar were developed for the US navy, which were installed in surface ships and submarines, including AN / SQR-19, AN / UQQ-2 and AN / TB-27A, etc. The navies of the main western countries developed their own towed array sonar as well, for example ATAS and COMTASS-II in the UK, CAPTAS in France and ACTAS in Germany. Now, the new version of a twin linear towed array sonar, called SURTASS (surveillance towed array sensor system), is installed in SWATH (small water area twin hull) vessel. It is an active / passive low frequency towed array system.

As a participant in towed array development in the USA, Lemon pointed out [5] that the early work on towed array sonar began in 1917 in the US navy. Dr. Hayes invented a twin 12 elements linear array and the detection distance was about 2,000 yards. After this, the development of towed array sonar was very slow until 1940, when Danish scientist Holm proposed a new design for linear towed array sonar. In the late 1950s, Lasky of ONR and Fitzgerald of Chesapeake Instrument Corp. designed a towed array sonar with a diameter of 3 inches. The timeline of the development of towed array sonar in the US Navy for surface ships and submarines can be summarized as follows:

(1) Towed array sonar for surface ships:

1967: AN / SQR-14;
1971: AN / SQR-15;
1980: Active towed array sonar;
1981: SURTASS;
1983: AN / SQR-19;

1989: SWATH vessel, towed array sonar;
1994: Applied Physical Laboratory, twin towed array;
2003: Cardinal towed array.

(2) Towed array sonar for submarines:
1971: Hughes Aircraft Company, TUBA towed array;
1973: AN / BQR-23;
1979:AN / TB-16;
1983: AN / TB-23;
1994: AN / TB-29;
2004: Optical fiber towed array sonar.

From this timeline of towed array sonar development, we can see the history of towed array development in the US Navy, and the technical achievements in towed array research and applications.

8.2.2 General Concept of Towed Array Sonar

Towed array sonar can be categorized as being of two types. One is active towed array sonar and passive towed array sonar. In a submarine only passive sonar is installed. In a surface ship, passive towed array sonar or active towed array sonar is installed.

Fig. 8.6 illustrates the block diagram of towed array sonar. The limit of the wet end and dry end is clear, but for simplification we prefer to include the winch in the wet end.

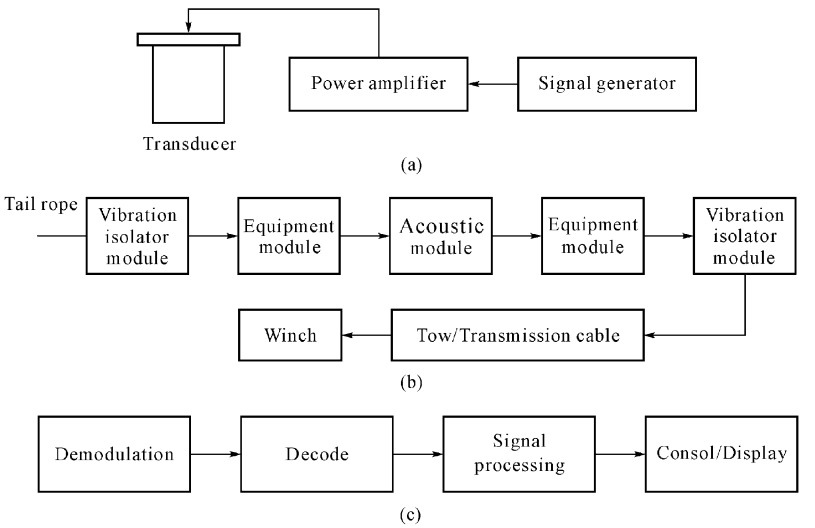


Fig. 8.6. Constituents of active / passive towed array sonar. (a) Transmit part; (b) Receiving part; (c) Signal processing part

The signal processor of towed array sonar is the same as that for any other kind of sonar, but a monitoring system of the towed array status is added. For active towed array sonar, it is necessary to have a transmitting transducer. It is usually towed from the aft platform. The transducer is installed in a tow fish with dome. There are two kinds of transducer arrays, one is a vertical array consisting of several transducers and the other is a horizontal array, consisting of several flexible tension transducers. It is necessary to have special deployment and depth fixing equipment. One can also use the so-called "critical angle tow method", i.e., by adjusting the length of cable and speed of the vessel one can change the depth of the transducer array. Fig. 8.7 shows two kinds of transducer arrays.

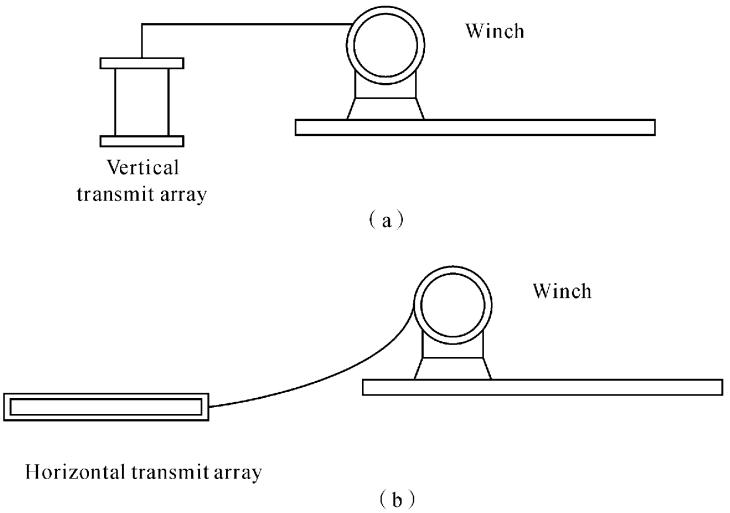


Fig. 8.7. Two kinds of transmit array of towed array sonar. (a) Vertical transmit array; (b) Horizontal transmit array

The receiving part of the wet end of a towed array is quite complicated. Besides the acoustic module, which consists of hydrophones to receive the underwater acoustic signal, there are several auxiliary components, a vibration isolation module and equipment module. The vibration isolation modules are installed on both sides of the acoustic module and their main function is as a mechanical filter. It reduces vibration during the towing; the equipment modules are also installed on both sides of the acoustic modules. Some necessary sensors, including depth sensors, course sensors and temperature sensors are installed too.

The information from equipment modules is transferred to the signal vessel and displayed in the signal processor, so that we can monitor the status of the tow cable and acoustic module.

In the acoustic module, the hydrophones and other necessary components, for example the pre-amplifiers and filters, are installed. If we use digital cable, it is necessary to install multiplexer and coding circuits in the acoustic module or

equipment module.

The tow cable as a mechanical tool for towing the acoustic module also has the function of signal transmission. There are three different types of signal transmission media, analogue cable, digital cable and optical fiber cable. The analogue cable transfers the analogue signal; the digital cable, typically a coaxial cable, transfers the coded digital signal. The optical fiber cable transfers the optical signal.

Fig. 8.8 illustrates the decomposition of main function of towed array sonar. It is supposed that the signal from the wet end is digital. The specific function here is the left / right ambiguity resolution problem, since towed array sonar with a single towed acoustic module cannot identify whether the target comes from port or starboard.

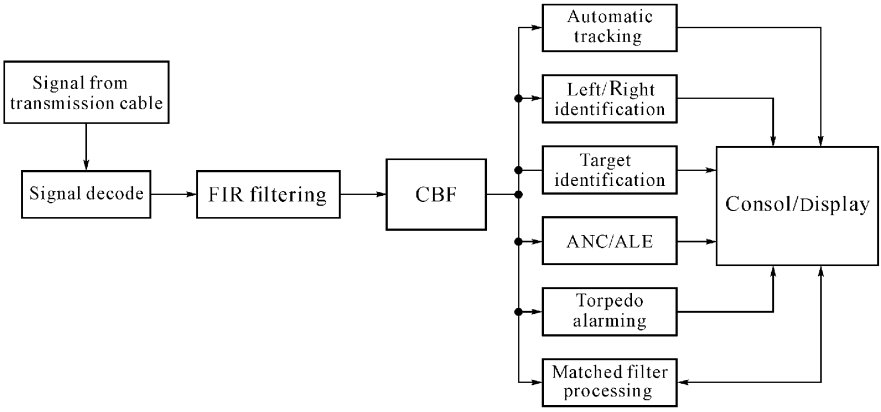


Fig. 8.8. Decomposition of main functions of towed array sonar

8.2.3 Status of a Tow Cable under the Sea

Generally speaking, the underwater tow system for surface ships has two different types, that are both the main components of the wet end of a sonar system (Fig. 8.9). One is a tow cable plus linear array which is a distributed mass system. The other is a tow cable plus tow fish or tow body which is an integrated mass system. For the distributed mass system, sometimes it is necessary to hang a heavy fish on the tow cable to keep the depth of the linear array.

The moving behavior of the underwater tow body or tow array is of concern to the sonar designer. It is actually a theoretical topic in hydro dynamics. There is no well-established theory that can perfectly describe the status of a tow cable under the sea, nor some practical algorithm which can accurately predict the movement of a tow cable and tow body in the sea. Many experiments at sea and empirical test results provide useful reference data[12,20-23,27-29].

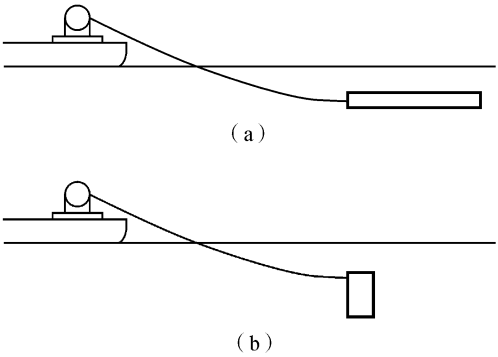


Fig. 8.9. Underwater tow system. (a) Tow cable and linear array; (b) Tow cable and towed body

When we study the moving rule of the entire tow system or tow body, the origin of the coordinate is usually selected as tow point, i.e., the end point of a tow cable or the starting point of the working module. Suppose the plane of the tow system is *OXY* and the axis *OZ* is perpendicular to the *OXY* plane (Fig. 8.10).

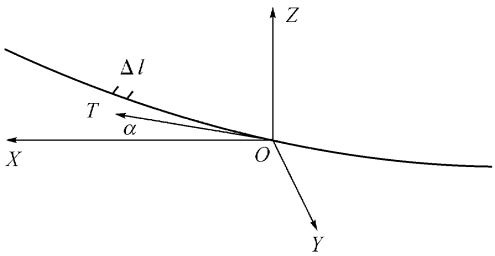


Fig. 8.10. Derivation of dynamic equation of underwater tow cable

In a general case, if the towed system is at a certain depth and the tow speed kept constant, then the neighborhood of the tow cable and towed array does not produce cavitation phenomena. Therefore, it is considered that the tow system is moving in a non-boundary, non-compressible liquid. The operation force in each unit is mainly determined by the viscous force; therefore the main task is to study the combination effects of inertial force and viscous force.

In hydrodynamics there are two constants, which describe tow body and tow cable movement. One is the Reno number and the other is the Fraud number. The Reno number R_e describes the relationship between inertial force and viscous force

$$R_e = vL / \nu \tag{8.4}$$

where v represents the moving speed of two bodies, L is the characteristic length of two cables and v is the viscous coefficient. The Fraud number F_r represents the relationship between inertial force and gravitational force:

$$F_r = v/(gL) \tag{8.5}$$

where g represents the gravitational acceleration speed. The theoretical calculation shows that the positive resistance force F_x, the upward / downward force F_y and the side force F_z can be expressed as

$$F_x = c_x \rho d \Delta l v^2 / 2 \tag{8.6}$$

$$F_y = c_y \rho d \Delta l v^2 / 2 \tag{8.7}$$

$$F_z = c_z \rho d \Delta l v^2 / 2 \tag{8.8}$$

where ρ represents the cable density, Δl is a small unit in the cable, d is the diameter of the cable; c_x, c_y and c_z represent the coefficients of liquid dynamics in directions x, y and z respectively. The kinetic equation of a tow cable is

$$\begin{aligned} &T = \text{constant},\quad l = T(\cot\alpha - \cot\alpha_0)/r,\quad x = T(1/\sin\alpha - 1/\sin\alpha_0)/r \\ &y = T\ln[\cot(\alpha/2)/\cot(\alpha_0/2)]/r,\quad z = 0 \end{aligned} \tag{8.9}$$

where r represents the resistance force of the unit cable and T is the tow force.

The solution of Eq. (8.9) shows that the cable in the OXY plane is close to a catenary. When the density of the cable is correctly selected, the curvature of the catenary is quite large. Therefore the shape of the acoustic module is approximately a straight line.

Based on Eqs. (8.6) – (8.8), the kinetic equilibrium equation is

$$T\sin\alpha = c_y \rho S v^2 / 2, \qquad T\cos\alpha = c_x \rho S v^2 / 2 \tag{8.10}$$

and we obtain

$$T = \rho S v^2 \sqrt{c_x^2 + c_y^2}\,/\,2 \tag{8.11}$$

The viscous coefficients c_x and c_y can be only measured by experiment. The typical value of sea water is $c_x \approx 1$ and $c_y \approx 0$. S is the section area of cable.

Example

If the cable diameter $d = 5$ cm, $v = 10$ kn and $\rho = 2.2$ t/m^3, then $T \approx 215$ kg.

8.2.4 Shape Distortion of Towed Array and Its Effect on Beamforming

In the two processes, the acoustic module of a towed array cannot be kept as a straight line. Any kind of shape distortion will result in the degradation of beamforming performance. The distortion of the array shape has some effect on the beamforming, but it is by no means necessary that we need to know the exact position of each element of the array.

Suppose the equal spaced linear array consists of N elements, the signal received by the i-th element is $s[t+\tau_i(\theta)]$ where $\tau_i(\theta)$ denotes the time delay with respect to some reference elements in the ideal case. If the position of the i-th element is biased in some way, then the received signal from the i-th element becomes $s[t+\tau_i(\theta)+\delta_i]$. Therefore the output of the beam, which steers to θ, is

$$D_0 = E\left(\sum_{i=1}^{N} s(t+\delta_i)\right)^2 \tag{8.12}$$

when all $\delta_i = 0,\ i = 1,\ ...,\ N$, then $D_0 = N^2\sigma_s^2$.

In the case of array shape distortion, some δ_i are not equal to zero and Eq. (8.12) becomes

$$D_0 = \sum_i \sum_j R_s(\delta_i - \delta_j) \tag{8.13}$$

Because the frequency for towed array sonar is usually lower, comparison with the correlation radius of signal δ_i is therefore small. We can take Taylor expansion at zero point. We obtain

$$R_s(\tau) \approx \sigma_s^2[1-(\alpha\tau)^2/2] \tag{8.14}$$

where α represents a parameter which describes the behavior of $R_s(\tau)$.

Therefore Eq. (8.13) can be rewritten as

$$D_0 = N^2\sigma_s^2 - \sum_i \sum_j [\alpha(\delta_i - \delta_j)]^2/2 \tag{8.15}$$

Eq. (8.15) shows the response reduction of the main lobe due to array shape distortion.

For example, suppose the frequency f = 500Hz, then the approximation expression for the correlation function of $\cos(2\pi f\tau)$ is $\cos(2\pi\tau) \approx 1-(2\pi f\tau)^2/2$. In this case the value of main lobe reduction is about 5% and it can be neglected.

The above discussion focuses on the main lobe beam, but for the entire directivity function, the calculation is quite complicated. It is necessary to know the exact position of each element. We will discuss the calculation of the directivity function in its simplest distortion form, i.e., the sinusoidal array shape distortion[23-27]. We will derive a method to find the coordinates of each element. Suppose the distortion only occurs on the horizontal plane and the acoustic module of the towed array is supposed to be non-flexible. Fig. 8.11 illustrates the shape distortion of towed array. The length of the acoustic module is Z_0, the coordinates of N hydrophones are (x_i, y_i) $(i = 1, ..., N)$. Before distortion

$$x_i = (i-1)d, \quad y_i = 0 \ (i = 1, ..., N) \tag{8.16}$$

where d is the distance between adjacent elements.

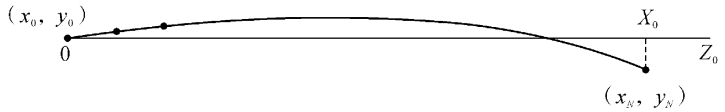


Fig. 8.11. Shape distortion of towed array

Suppose the incidental angle of the signal is θ, the bearing angle of the towed array sonar is θ_0. The output of the beamforming system at angle θ_0 is

$$D^2(\theta, \theta_0) = \left[\sum_{i=1}^{N} \sin(\xi_i(\theta) - \tau_i(\theta))\right]^2 + \left[\sum_{i=1}^{N} \cos(\xi_i(\theta) - \tau_i(\theta))\right]^2 \tag{8.17}$$

where

$$\tau_i(\theta) = (i-1)d \sin\theta / c \tag{8.18}$$

$$\xi_i(\theta) = [x_i \sin\theta + y_i \cos\theta] / c \tag{8.19}$$

For the sinusoidal array shape distortion

$$y_i = A \sin(px_i) \tag{8.20}$$

where A and P are a pair of parameters for describing shape distortion. In order to find coordinates (x_i, y_i) based on the assumption of $y = f(x) = A\sin(px)$, it is necessary to use integration along the arc, that is

$$(i-1)d = \int_0^{x_i} \sqrt{1 + f'^2(x)}\,\mathrm{d}x \tag{8.21}$$

For $i = N$, we have

$$Z = (N-1)d = \int_0^{x_N} \sqrt{1 + f'^2(x)}\,\mathrm{d}x \tag{8.22}$$

Because $f(x)$ is a nonlinear function, $x_i - x_{i-1}$ depends on i. Eq. (8.21) is actually an ellipse function. Define

$$E(X,\ A,\ p) = \int_0^X \sqrt{1 + f'^2(x)}\,\mathrm{d}x = \int_0^X \sqrt{1 + A^2 p^2 \cos^2(px)}\ \mathrm{d}x$$

It can be rewritten as standard ellipse function form, i.e.

$$E(X,\ A,\ p) = (Q_1 / p) \int_0^{pX} \sqrt{1 + Q_2^2 \sin^2 x}\ \mathrm{d}x \tag{8.23a}$$

where

$$Q_1 = \sqrt{1 + A^2 p^2},\quad Q_2 = Ap / Q_1 \tag{8.23b}$$

Based on Eqs. (8.23a) and (8.23b), we obtain

$$E(X,\ A,\ p) = (i-1)d,\ x_i = E^{-1}[(i-1)d,\ A,\ p],\ y_i = A\sin(px_i)\ (i = 1,\ ...,\ N) \tag{8.23c}$$

Fig. 8.12 illustrates the relationship of A and X_0. For $A = 2$ m, X_0 changes from 49.9 m to 45.9 m.

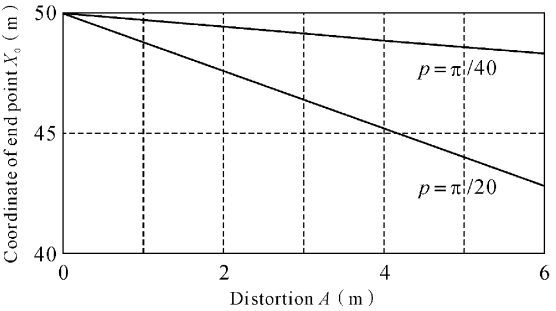


Fig. 8.12. The relationship of p, X_0 and A

Figs. 8.13 and 8.14 show the maximum time delay resulting from array shape distortion. We can see that if the distortion satisfies the sinusoidal shape assumption, then it is possible to predict the exact position of each element. Therefore, it can provide an estimation of beamforming. If the value of

A and p is large enough, then it is impossible to have normal beamforming. The sonar operator has to wait until recovery of the array shape to normal.

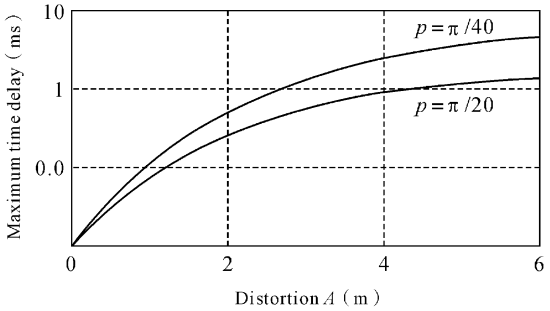


Fig. 8.13. Time delay resulting from distortion of array shape

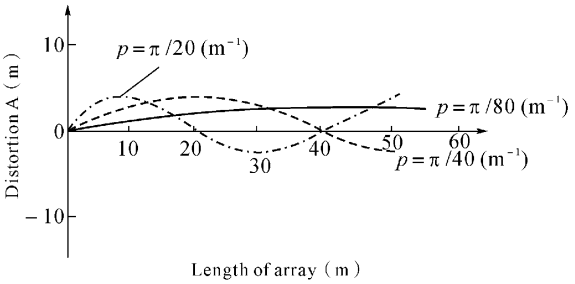


Fig. 8.14. Example of array shape distortion

8.2.5 *Some Special Problems with Towed Array Sonar Design*

There are some special problems we have to pay attention to in towed array sonar design. These problems only occur in towed array sonar.

8.2.5.1 Bearing Ambiguity Problem

Since the hydrophone in a towed linear array is omni-directional, therefore it is impossible to identify whether the target is on the left or right side of the array. This inherent problem for a single linear towed array is called left / right ambiguity. An effective method for solving this problem is to make a maneuver when a contact is confirmed. Fig. 8.15 illustrates this method. If the signal presents itself at α, then the target may be on the port side or starboard side. To make a vessel maneuver the θ angle, if the target is on the port side, then the maximum value

will appear in $\beta_1 = \alpha - \theta$; if the target is on the starboard side, then $\beta_2 = \alpha + \theta$. The disadvantage of this method is that the maneuvering time is sometimes not neglected.

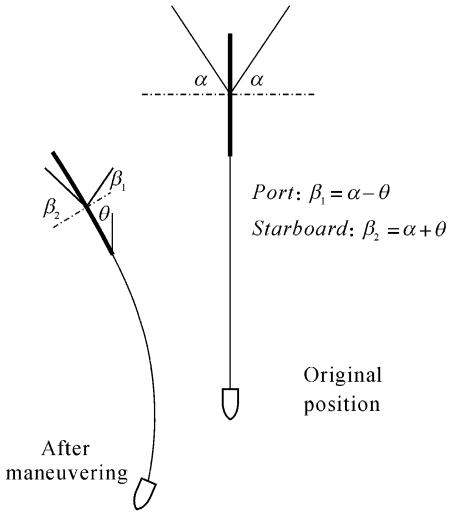


Fig. 8.15. Left / Right ambiguity solving by using vessel maneuver

There are also other kinds of methods for solving left / right ambiguity problems. For example, the utilization of a dipole or triple hydrophone combination can distinguish the target incidental angle, since the dipole or triple hydrophone has a directional beam pattern which can be used to identify the target incidental angle. Of course it works on a certain frequency band. But the diameter of the acoustic module will increase due to the utilization of dipole-element or triple-element hydrophone groups.

If the sonar system is allowed to use a twin linear array system, then the left / right ambiguity problems can be easily solved, since the twin linear array system has a good horizontal directivity function which can identify a left / right target. Fig. 8.16(a) illustrates the directivity function of a dipole-element hydrophone in twin linear array and Fig. 8.16(b) illustrates the calculating the directivity function of twin-array sonar.

In Fig 8.16(a), for $d / \lambda = 0.25$, it is easy to identify the target from the left side ($\varphi = \pi / 2$) or right side ($\varphi = 3\pi / 2$).

The directivity function of the twin array shown in Fig. 8.16(b) is

$$E(\theta, \theta_0) = \left| \frac{\sin[N\pi \mathrm{d}(\sin\theta - \sin\theta_0) / \lambda]}{\sin[\pi \mathrm{d}(\sin\theta - \sin\theta_0) / \lambda]} \right| \times \sqrt{2} \left\{ 1 + \cos\left[\frac{2\pi d_0}{\lambda} (\cos\theta - \cos\theta_0) \right] \right\}^{1/2} \quad (8.24)$$

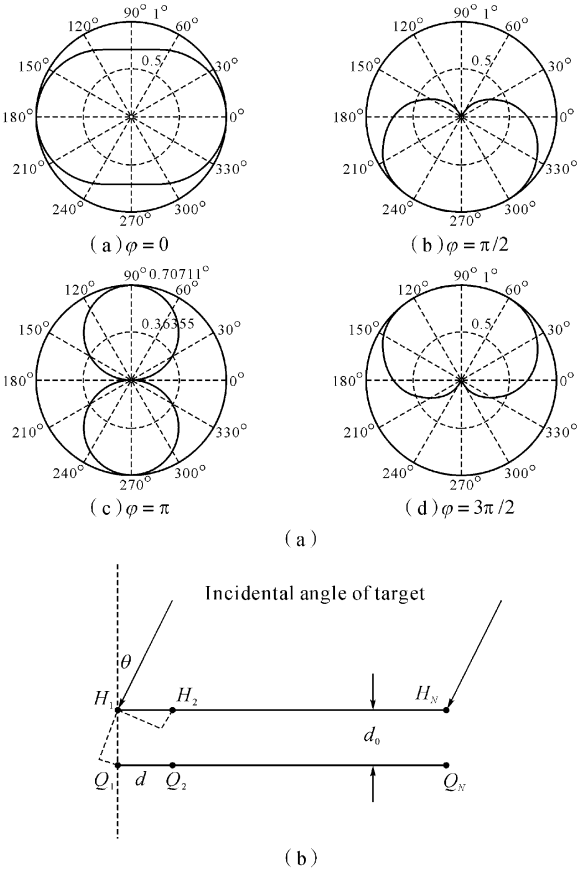


Fig. 8.16. (a) Beam pattern of dipole; (b) Beam pattern of twin-line array of equal-spaced elements

8.2.5.2 Flow Noise of Towed Array

The acoustic module of a towed array sonar is towed under the sea; the flow noise is one interference source for the hydrophone. Since hydrophones are installed in a special material-made jacket, therefore the performance of flow noise strongly depends on the material, the smoothness of the skin, diameter of the acoustic module and towed speed. It is easy to understand that the flow noise is proportional to the towed speed and inversely proportional to the diameter of the acoustic module.

In Chapter 4, we introduced the flow noise of a seismic array. The experimental test shows that the flow noise for a seismic array is in the frequency band of 10 – 250 Hz. The diameter of the seismic array is usually 8 – 10 cm. For a seismic array the frequency band of flow noise is just in the operational frequency band of seismic exploration. This is the main interference source for a seismic towed array. Fortunately, the tow speed of the seismic array is usually very low

and therefore the effect of flow noise problems is not serious. For military sonar, since the working frequency is usually above 100 Hz, the flow noise is often neglected. Fig. 8.17 shows the test results of towed array sonar from WTD (Bundeswehr Technical Center for Ships and Naval Weapons), Germany[6].

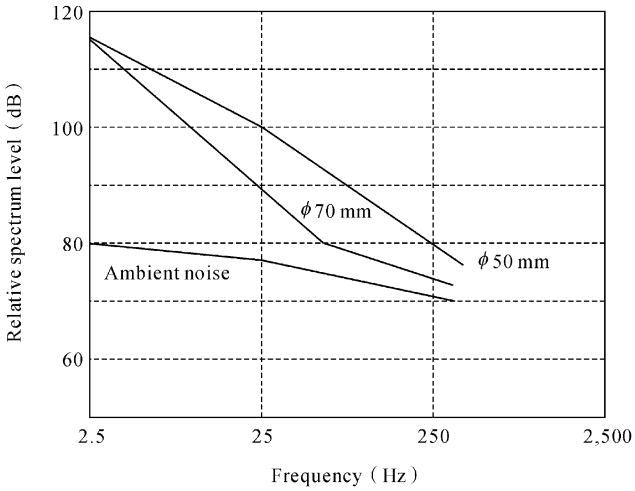


Fig. 8.17. Example of flow noise of two cables

The frequency range of measurement is 2.5 – 250 Hz. The name of the measurement system is called WTD 71. The results show that at a speed of 8 kn, for a diameter of 50 mm and 70 mm, the flow noise resulting from the tow cable is negligible when the frequency is above 250 Hz.

As for the relationship between flow noise and the diameter of the acoustic module, there is no theoretical closed form expression yet.

8.2.5.3 Relationship between Cable Length, Vessel Speed and Cable Depth

The outstanding advantage of towed array sonar is the flexible depth choice for the acoustic module. It can be adjusted to the desired depth to meet the requirements of sound speed profile, i.e., the favorite depth for sonar operation. There are mainly two methods for adjusting the depth of the acoustic module. One is the so-called "sea bird" which is attached to the jacket in a certain place on the acoustic module, and it is remotely controlled by cable. The sonar operator can send commands to the sea bird if necessary. The sea bird goes up or down to adjust the depth of the acoustic module. But the range it can manage is very limited. It is mainly used in seismic towed array sonar. It is rarely used in military towed array sonar.

The other method used to adjust cable depth is called the "critical angle towing" method. The acoustic module of towed array sonar is designed to have zero buoyancy floating force. The depth of the tow cable is the function of tow speed and length. The relationship between tow cable, length of tow cable and tow

speed can be measured in sea trials. Actually, one can obtain the table in which the sonar operator and commander can find the necessary data to deploy the acoustic module at optimum depth.

Figs. 8.18 – 8.20 show the relationship of cable depth, cable length and tow speed. From these figures, we can get a general idea of the main parameters of tow cable movement.

Figs. 8.21 and 8.22 give some experimental results measured at sea. It is necessary to keep the entire acoustic module in a horizontal position. It is not easy when the acoustic module is long. Just as we have shown before, generally speaking, the acoustic module in the sea is near a catenary. If the density is close to the sea water, then the curvature of the catenary will tend to be infinite. In order to measure the flatness of the acoustic module, two depth sensors are installed at the end of the acoustic module. The measured results are illustrated. The length of the acoustic module is 100 meters. The test tow speed is 6 kn and 8 kn respectively. In this figure, the end A is closer to the vessel than B.

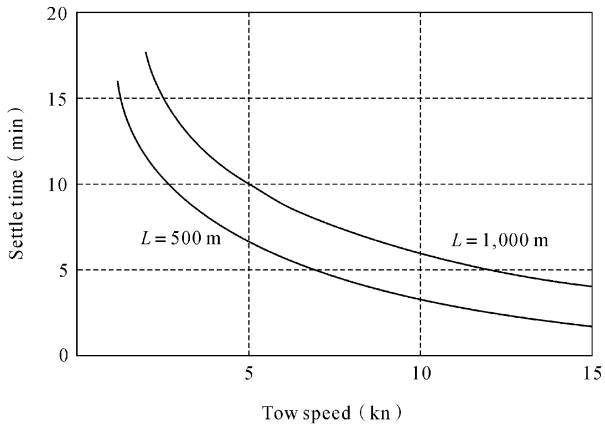


Fig. 8.18. Example of relationship between the tow speed and the settle time

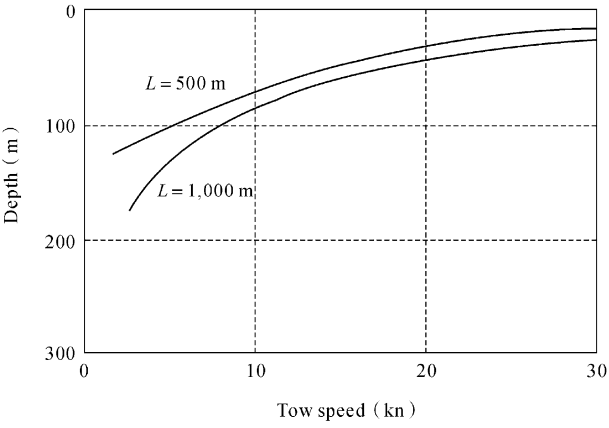


Fig. 8.19. Example of the relationship between tow speed and the depth of acoustic module

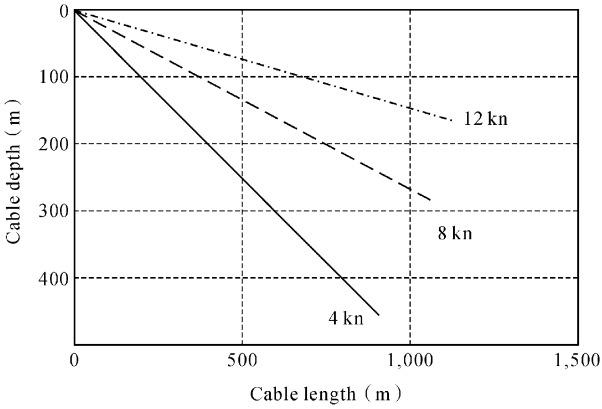


Fig. 8.20. Relationship of cable length and cable depth in critical angle towing

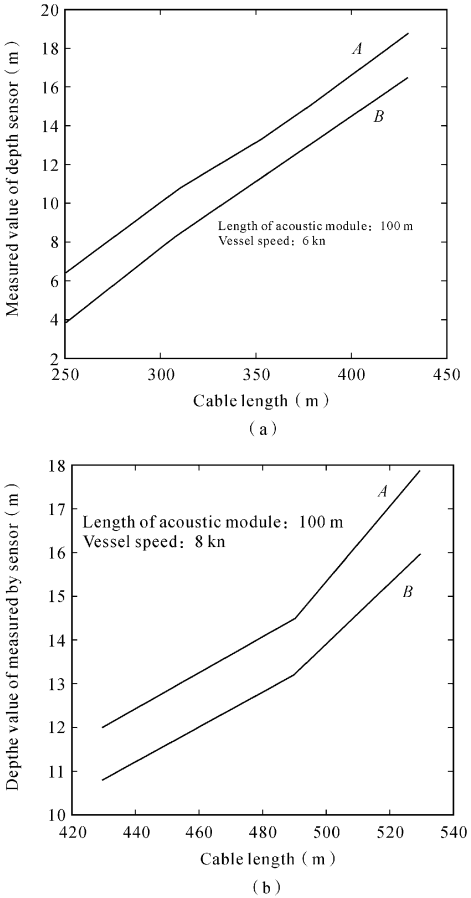


Fig. 8.21. (a) Experimental results of towed array; (b) Experiment results of tow cable

8.2.5.4 Acceleration Sensitivity of Hydrophone

When the hydrophone of a towed array sonar receiving signal is in motion, a sudden change in speed (positive or negative acceleration) and vessel maneuvering will result in additional noise. We hope the acceleration noise will not affect the signal received from the target, i.e., the acceleration sensitivity should be as small as possible. The typical value of voltage sensitivity of a hydrophone is about –195 dB (ref. 0 dB = 1 V / μPa), and the typical value of acceleration sensitivity is about –60 dB (ref. 0 dB=1 V / g). In order to reduce acceleration sensitivity, it is necessary to pay attention to the design of anti-acceleration. One measure is to combine two poles placed in opposite directions, such that the acceleration response can be cancelled out. Another measure is to install the hydrophone in a specific anti-vibration rack, which amortizes the sudden change in tow speed.

8.2.5.5 Tow Ship Noise Canceling

Directional noise originating from the tow ship, mainly machinery and hydrodynamic noise, often limits the sonar performance. For active / passive towed array sonar, the interference generated from the tow body and propeller will seriously result in a performance decrease of the sonar system. Fig. 8.22 illustrates the effects of the tow body and propeller noise.

In a shallow water environment, these kinds of interferences cannot be neglected. We will briefly analyze the method of calculating the DOA of interference.

Fig. 8.22(a) illustrates an active/passive towed array operated in shallow water, where A is the tow body, the direct distance from A to the passive receiving array is r_0. The length of the passive receiving array BC is l_0. Generally speaking, A and BC are located at the same depth of h_1. The distance from BC to the sea bottom is h_2.

In the towing process, the flow noise of the dome will be one of the main noise sources of the system. The interference can be roughly divided in to two parts: the direct wave and the reflected wave from the sea surface and sea bottom. In this paper we only calculate the reflected wave from the sea bottom.

Suppose we choose the normal line of BC as 0° direction, then the direct wave is in a direction of –90°, which is the endfire direction. Since the width of the main lobe of the beam pattern in the endfire direction is quite wide, interference from the direct wave will result in a blind area in target detection. The reflected wave from the sea bottom is another source of interference. Seriously speaking, if the sea bottom is flat, then the incidental angle in the receiving arrays B and C of the reflected wave resulting from tow body A will be α and β, respectively. But in practice, the sea bottom is not flat so the reflected wave will extend outside of the region $[\alpha, \beta]$. If there are some objects on the sea bottom, the size of which can be compared with the wavelength of tow body noise, then the reflected wave can also appear in a direction other than α and β. This will make the situation more

complicated. Of course, the strength of the reflected wave depends on the material on the sea bottom and also depends on the sound velocity profile and sound propagation path.

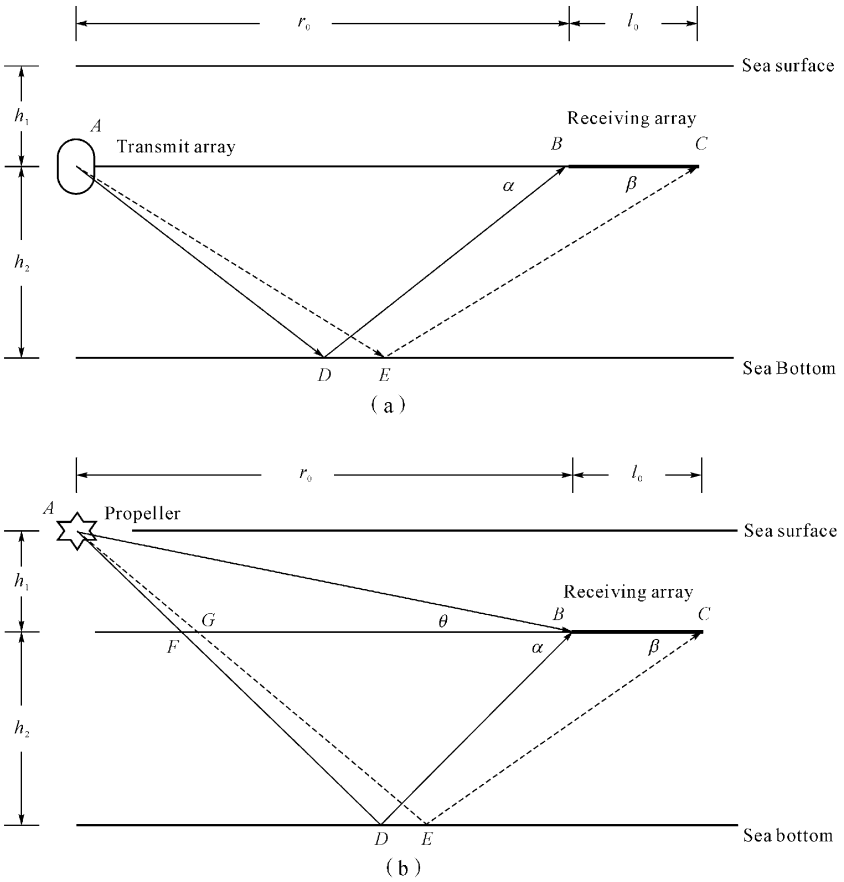


Fig. 8.22. (a) Interference resulting from towed body; (b) Interference resulting from propeller

In the assumption of the equal sound velocity profile, it is proved that the incidental angle of α, β can be given as

$$\alpha(\mathrm{tg}^{-1})2h_2 / r_0,\ \beta = \mathrm{tg}^{-1}[2h_2 / (r_0,\ l_0)] \tag{8.25a}$$

For cable length r_0=400 m, length of receiving array l_0= 60 m, if h_2= 75 m, then $\alpha \approx 20.5°$, $\beta \approx 18°$.

Fig. 8.22(b) illustrates the interference resulting from the propeller of a towed platform. Denote the length of receiving array as l_0, the distance between the propeller A to receiving array BC is r_0. The angle between the direct wave of

propeller noise and the horizontal line is θ. The incidental angle of the reflected wave of propeller noise from the sea bottom is shown as α and β. It can be shown that the expressions for calculating θ, α, β are as follows:

$$\theta = \mathrm{tg}^{-1}(h_1 / r_0) \tag{8.25b}$$

$$\alpha = \mathrm{tg}^{-1}[(2h_2(h_1) / r_0] \tag{8.25c}$$

$$\beta = \mathrm{tg}^{-1}[(2h_2 + h_1) / (r_0 + l_0)] \tag{8.25d}$$

Eqs. (8.25b) – (8.25d) are slightly different from Eq. (8.25a). This is because the propeller of the towed platform and receiving array are not on the same horizontal plane. The perpendicular distance h_1 is shown in Eqs. (2) – (4).

For example, if $h_1 = 25$ m, $r_0 = 400$ m, then $\theta = 3.5°$; if $h_2 = 75$ m, $l_0 = 60$ m, then $\alpha = 23.6°$, $\beta = 20.8°$.

Compared with the interference of the tow body, the incidental angle of interference which results from the propeller of the platform, is much closer to a normal line. Therefore it will have a more serious effect in signal detection. Of course, the strength and distribution of the signal spectrum of interference must be considered.

In order to suppress the tow ship noise, an effective algorithm is called adaptive IBF (inverse beamforming)[163]. In a conventional beamforming system, the tow ship noise component in each beam is subtracted by the signal which steered in the direction of tow ship noise.

For a towed array sonar with N elements, for a particular direction θ, the beam output is

$$p(f,\theta) = \frac{1}{N}\sum_{n=1}^{N} w_n X_n(f) \exp[-2\pi \mathrm{j} nf \cos\theta / c] \tag{8.26a}$$

where $X_n(f)$ represents the Fourier transform of the n-th signal, d is the hydrophone spacing, w_n is the weighting coefficient and c is the sound velocity. The IBF algorithm is realized by setting

$$X_{\mathrm{tow}}(f,n) = p_{\max}(f,\theta_{\max}) \exp[-2\pi \mathrm{j} nd f \cos\theta_{\max} / c] \tag{8.26b}$$

and

$$X_{\mathrm{new}} = X - X_{\mathrm{tow}} \tag{8.26c}$$

where $\theta_{\max}$ is the detected tow ship noise direction. The corresponding beam output $p_{\max}(f, \theta_{\max})$ is then estimated for each frequency bin to enable the reconstruction

and suppression of the interference component, as shown in Eq. (8.26).

8.2.6 Examples of Towed Array Sonar

In the previous section, basic knowledge of towed array sonar has been discussed. In the following, we will introduce two examples of passive and active/passive towed array sonar. The materials introduced here can be found in the literature and on the Internet[129-134].

Fig. 8.23 shows a typical configuration of passive towed array sonar. The wet end of this sonar consists of two vibration isolated modules, two equipment modules and one acoustic module. There are 64 – 96 hydrophones installed in the acoustic module which can be configured as a middle frequency array and as a low frequency array. The dry end consists of a signal processor and display/control system. It has narrowband signal processing and wideband signal processing. LOFAR and DEMON can be used for target detection and identification.

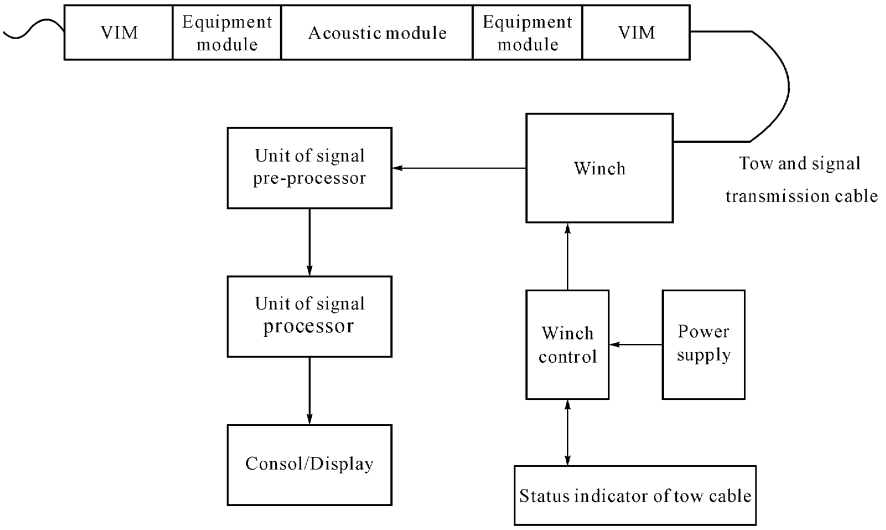


Fig. 8.23. Typical configuration of passive towed array sonar

Fig. 8.24 shows a typical configuration of active/passive towed array sonar. The vertical transmitter array is installed in the front of the receiving array, which is a twin array system. The vibration isolated modules are installed on both sides of the receiving array. The dry end consists of a signal processor, display/control, transmitter and other auxiliary equipment, including an array status monitoring cabinet.

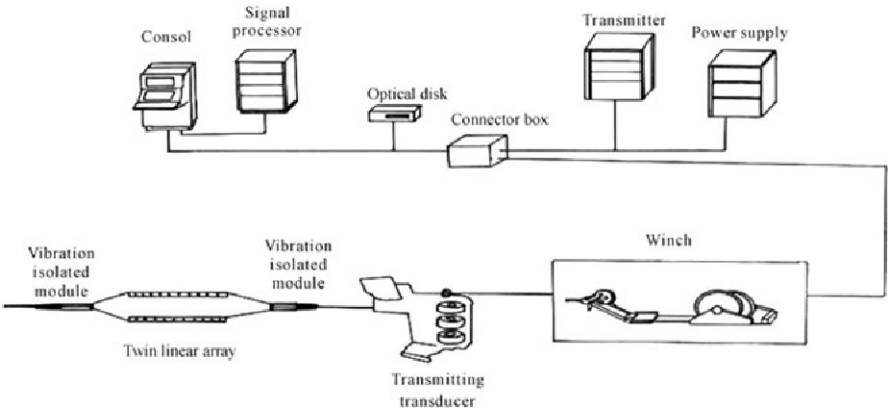


Fig. 8.24. Typical active/passive towed array sonar

8.3 Flank Array Sonar for Submarine

The advantage of towed array sonar is that the acoustic module is far away from the propeller noise; therefore this will considerably improve the detection performance of the sonar system. But for the submarine, the installation of a winch and long tow cable and the maneuvering itself results in some problems. Therefore, the sonar designer seeks a new type of sonar array; it keeps the large aperture of a towed array, but has a relatively small effect on submarine maneuvering. This is the flank array. It is sufficient to handle the size of a submarine but the disadvantage, compared with the towed array, is the effect of interference from submarine noise.

In this section, we will briefly introduce the basic design principle of flank array sonar and focus on self-noise cancellation, array baffling and beamforming issues.

8.3.1 General Concept of Flank Array Sonar

The main type of sonar system installed in a submarine is bow array sonar; the forms of array are spherical or cylindrical. The aperture of this kind sonar array is usually less than 4 m for the space limitations. Therefore the frequency to be used in target detection is higher than 1,000 Hz. If the receiving hydrophones are installed along the body of submarine, the array aperture will increase. The operation frequency of the sonar system can descend to a low frequency and thus the detection range can increase. The earliest extended array sonar was adopted in US submarines of the Los Angeles class. It is called a “conformal array”. The

hydrophone of conformal array sonar extends from the bow of the submarine to the flank, in a vertical direction, and in three layers. But the aperture of conformal array is not large enough compared to the flank array. The hydrophone arrays of flank sonar are symmetrically distributed on the port and starboard sides.

Fig. 8.25 illustrates the configuration of a flank array in a submarine. The aperture of the sonar array is usually about 30 – 40 m. The operation frequency band is about 100 – 1,500 Hz. The utilization of a low frequency can considerably increase the detection range of the sonar system. The other specification of a flank array sonar is its passive ranging function. This is because the three-point array, which is necessary for passive ranging, can be easily accessed from a flank array. In this sense, flank array sonar may be considered as multi-functional passive sonar.

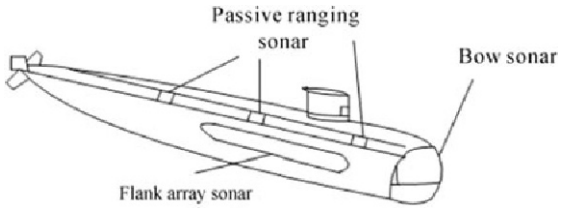


Fig. 8.25. Position of flank array sonar in a submarine

8.3.2 Adaptive Canceling of Self Noise

The hydrophones of a flank array are installed in the hull of a submarine. The signal received by the hydrophone consists of two parts: target radiated noise and ambient noise; the flow noise around the hydrophone and the vibration from the submarine itself. The latter is harmful, and we should suppress it as much as we can. Fig. 8.26 shows a hydrophone installed in the hull of a submarine. In order to isolate the vibration from a submarine, it is necessary to cover this up with a layer of de-coupling materials, which can reduce the effect of self vibration. When the submarine is moving under the sea, the flow noise resulting from the cavitation will become a main source of interference.

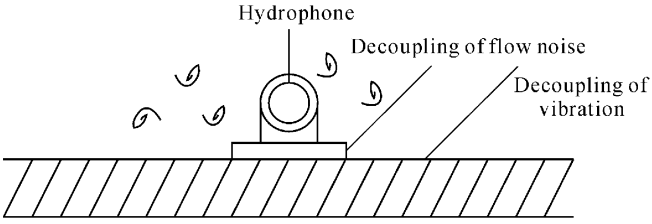


Fig. 8.26. Environmental noise of flank array hydrophone

As is well known, the piezoelectric material of a hydrophone is sensitive to the change in sound pressure in the acoustic field. If the sound pressure in location (x, y, z) is $p(x, y, z; t)$, then the hydrophone in this location can detect the sound pressure and reflect the linear variation. But if the hydrophone is moving from a static state with a certain velocity, or the linearly uniform moving hydrophone suddenly accelerates, then the hydrophone will have some response to the change in acceleration speed. The sensitivity of acceleration velocity depends on the structure of the hydrophone and the direction of acceleration velocity.

The wave equation of a resonance sound field is

$$\nabla^2 p(x, y, z; t) - \frac{1}{c}\frac{\partial^2 p(x, y, z; t)}{\partial t^2} = 0 \tag{8.27}$$

where c represents sound velocity and ∇ denotes the operator

$$\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$$

Based on Eq. (8.27), the initial condition and boundary condition, the sound pressure can be solved as:

$$p(x, y, z; t) = A\exp[\mathrm{j}(kx\cos\theta\sin\varphi + ky\sin\theta\sin\varphi + kz\cos\varphi)]\exp(\omega t) \tag{8.28}$$

where $k = \omega / c$ represents the wave number and ω is the resonant frequency. The propagation direction of the sound wave is (θ, φ) and θ is the wave direction on the horizontal plane, φ is the angle between the wave path vector and z axis. In the homogeneous sound field, the kinetic equation of the sound wave is

$$\frac{\partial v(t)}{\partial t} + \frac{1}{\rho_0}\nabla p(x, y, z; t) = 0 \tag{8.29}$$

where $v(t)$ represents the particle vibration velocity and ρ_0 is the density of the medium. Substituting Eq. (8.29) into Eq. (8.28), we can calculate the projector of $v(t)$ on three axes

$$v_x = p\cos\theta\sin\varphi / (\rho_0 c) \tag{8.30}$$

$$v_y = p\sin\theta\sin\varphi / (\rho_0 c) \tag{8.31}$$

$$v_z = p\cos\varphi / (\rho_0 c) \tag{8.32}$$

From Eqs. (8.30) to (8.32), we know that the vibration velocity of the particle depends on the sound pressure and density of the medium, and is also related to the movement status $(\theta,\ \varphi)$.

Since the flow noise and vibration noise resulting from the submarine itself are harmful to sonar detection, it is necessary to suppress them as much as possible. A very efficient method is to use an adaptive noise canceling technique. We have to install a sensor to pick up the self vibration from the submarine just at the location of the hydrophone. The output of this sensor will be used as a reference signal to cancel background noise in the received signal of the hydrophone. The system configuration is illustrated in Fig. 8.27.

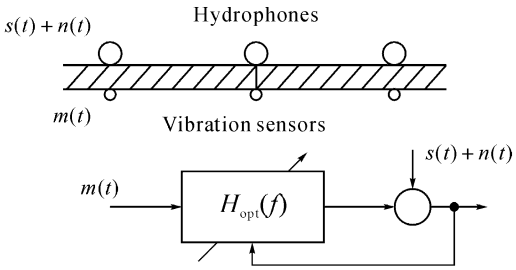


Fig. 8.27. Adaptive noise cancellation of hull vibration

Suppose the noise from hull vibration is $m(t)$ and the signal received by the hydrophone is $s(t)+n(t)$, where $s(t)$ represents the signal of target radiated noise and ambient noise, and $n(t)$ is the hull vibration noise. If the location of the vibration sensor is very close to the hydrophone, then $m(t)$ and $n(t)$ are strongly correlated. The signal $m(t)$ is selected as the reference input of the adaptive filter and $s(t)+n(t)$ is selected as primary input. The result of noise cancellation is to minimize the mean square value of error signal $\varepsilon(t) = \tilde{m}(t) - s(t) - n(t)$. We have

$$E[\varepsilon^2(t)] = E[(\tilde{m}(t) - n(t))^2] + E[s^2(t)] \tag{8.33}$$

where $\tilde{m}(t)$ represents the filtered result of $m(t)$. To adjust $H(f)$, make $E[\varepsilon^2(t)]$ minimum and we will obtain the minimum value of $E[(\tilde{m}(t) - n(t))^2]$. This is equivalent to maximizing the output SNR

$$(SNR)_{\text{out}} = \frac{E[s^2(t)]}{E[(\tilde{m}(t) - n(t))^2]} \tag{8.34}$$

Therefore, this kind of noise cancellation is very effective. Fig. 8.28 is an example of a noise cancellation result[136]. In this example, the hull vibration has a

very strong line spectrum component. It is distributed in the frequency band 0 – 350 Hz. The spectrum level is higher than the continuous spectrum by about 10 – 25 dB.

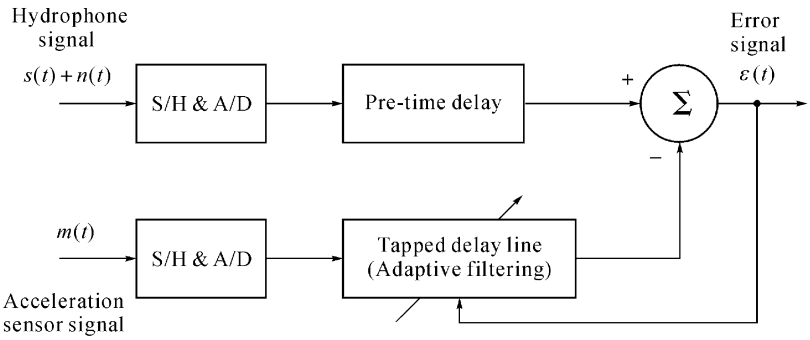


Fig. 8.28. Realization of adaptive self noise cancellation for flank array sonar

After noise cancellation, the line spectrum components almost disappear. The other example is given in Fig. 8.29. In order to verify that the adaptive noise canceling filter only suppresses the vibration resulting from the submarine and does not affect the single frequency detection of the underwater acoustic signal, the signal of f_0 = 825 Hz is transmitted. As we can see in this figure, the line spectrum components of hull vibration are cancelled, but the single frequency signal remains.

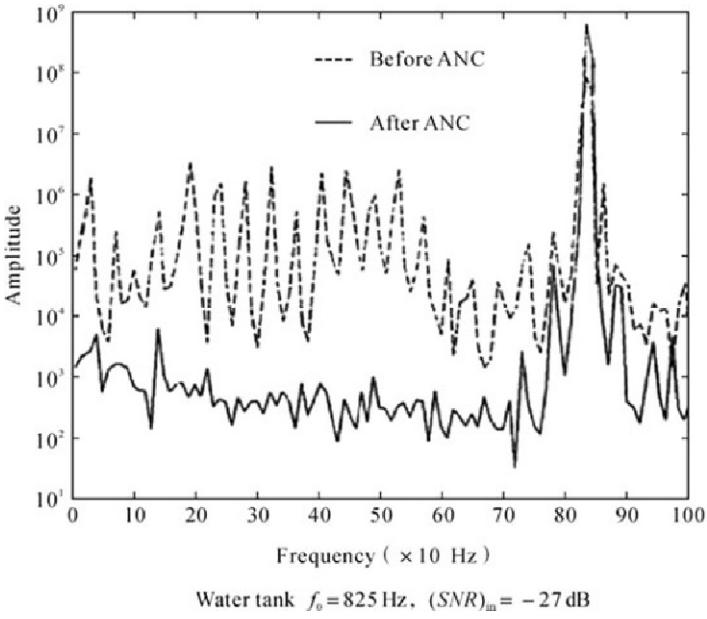


Fig. 8.29. Result of self noise ANC

It is necessary to show that, in principle, for each hydrophone we need a sensor to pick up the hull vibration. This is a big cost, since the number of signal channels will be double. Some researchers show that the main interference form of hull vibration is the line spectrum component. In this situation, we should then focus on the suppression of line spectrum components. Since the correlation function of the single frequency signal only depends on the phase difference, the number of vibration sensors can be reduced.

8.3.3 *Vibration Isolation of Hydrophone and Acoustic Damping Materials*

There are two different types of hydrophones that can be installed in a flank array: the conventional point type hydrophone and the plane hydrophone made of PVDF (polyvinylidene fluoride) material. The latter is better for suppressing low noise. Whatever kinds of hydrophones are adopted in a flank array, it is necessary to pay attention to the problem of vibration isolation. The base of the hydrophone should be made of special acoustic damping materials.

The basic requirements of acoustic absorbing materials are resistance match and a big absorption rate. That means the acoustic wave should be non-reflection transmitted in a material with large absorption, say 99%. The sound absorption mechanism can be roughly divided into two categories. One type of absorption results from movement (friction), and the dissipation of material generates the acoustic absorption; the other type of sound absorption comes from the relaxation effect of the material. Now the main materials used in sound absorption are rubber and plastics. Of course, some special treatment is necessary for the materials, for example the filling of plastic enhancement material, changing the material structure by mechanical techniques, etc.

When the de-coupled material is used in the base of the hydrophone, the acoustic resistance of the materials should much less than the materials (metal, ceramics) that the hydrophone is made of. A good vibration isolation material plays a very important role in flank array installation. It is a necessary pre-treatment for an adaptive noise filter. Fig. 8.30 is an example of the practical experiment. The vibration isolation material used in this experiment has quite good performance above 200 Hz. But when the frequency is lower than 200 Hz, the isolation effect is decreased.

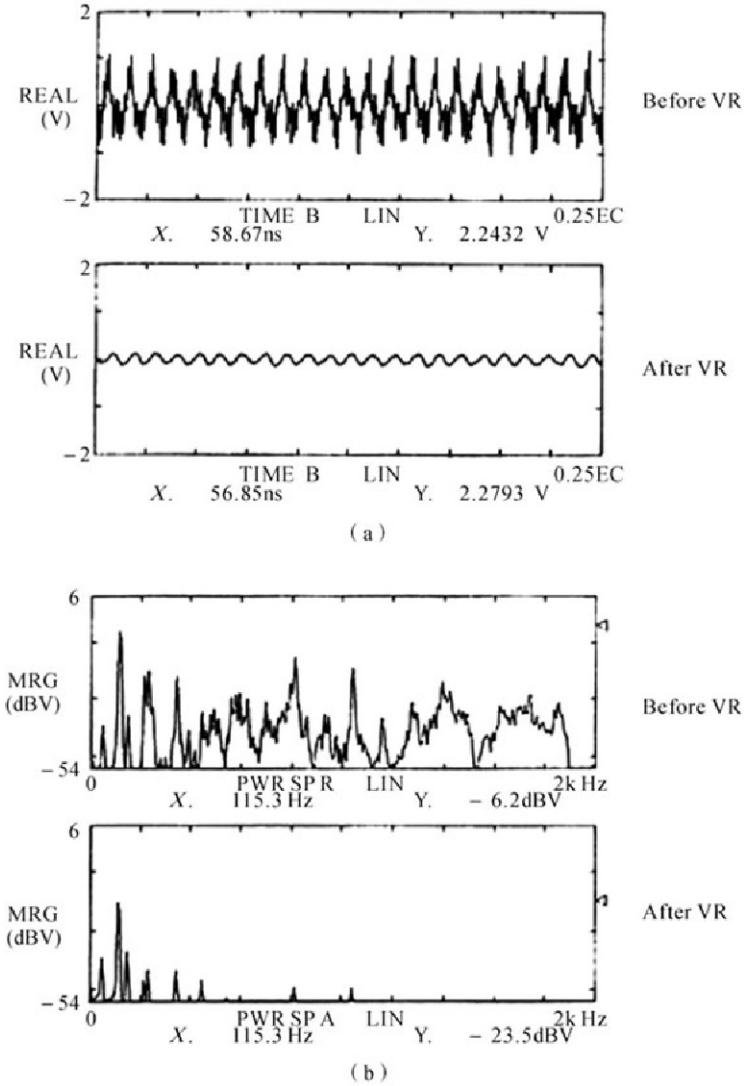


Fig. 8.30. Experimental results of vibration reducing (VR) of flank array. (a) Time domain; (b) Frequency domain

8.3.4 Beamforming of Multiple Elements with Different SNR

The hydrophones of a flank array are installed on both sides of the submarine. On account of the different locations of the hydrophones, the platform noise received by these hydrophones also varies. An important fact is that the noise level of the submarine gradually decreases along the hull, starting with the propeller. This

means that the flank array is operated in the case where the SNR of elements is different. This situation has not happened before.

Suppose the i-th signal is

$$x_i(t) = s[t + \tau_i(\theta_0)] + a_i n_i(t) \tag{8.35}$$

where $a_i \geq 0$ represents the magnitude of i-th noise $n_i(t)$ The noise $n_i(t)$ ($i = 1, \ldots, N$) cannot be considered as independent. Let

$$E[n_i^2(t)] = \sigma_n^2 \ (i = 1, \ ..., \ N) \tag{8.36}$$

The signal, which steers in the target direction, is

$$y(t) = \sum_{i=1}^{N} x_i(t - \tau_i(\theta_0)) = Ns(t) + \sum_{i=1}^{N} a_i n_i(t - \tau_i(\theta_0)) \tag{8.37}$$

It follows that the output SNR will be different from a conventional beamforming system. The straightforward calculation shows that

$$(SNR)_{\text{out}} = \frac{N^2 \sigma_s^2}{\sigma_n^2 \sum_i \sum_j a_i a_j \rho_n[\tau_i(\theta_0) - \tau_j(\theta_0)]} \tag{8.38}$$

where $\rho_n(\tau)$ represents the correlation coefficient of submarine self-noise. In calculating Eq. (8.38), the ambient noise is neglected.

8.3.5 System Configuration of Flank Array Sonar

Fig. 8.31 shows the block diagram of a typical flank array sonar. The flank array sonar is quite different from the other sonar system we discussed before, which is in the signal pre-processing unit. There are two completely different input signals. One is the underwater acoustic signal from hydrophones and the other is the signal from the vibration sensor; it is usually completed by the acceleration sensor.

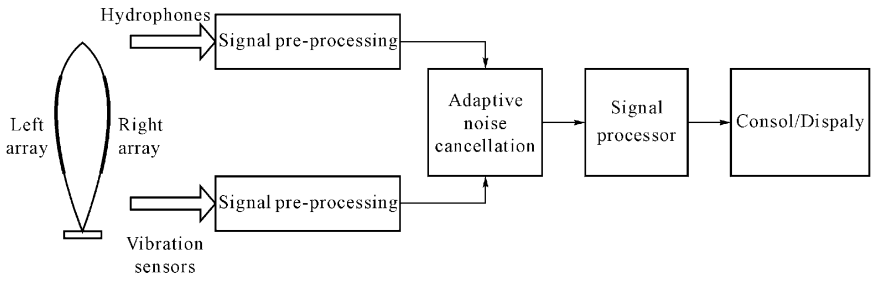


Fig. 8.31. Basic configuration of flank array sonar of submarine

For installation of a flank array, the submarine hull can be divided into two categories: a single hull and a double hull. The rule for installation of the wet end of a flank array and signal processing method is completely different. For a double hull submarine, since the hydrophones are installed in a non-pressurized shell, the array baffle becomes a important issue in array design. Inside the pressurized hull, which is filled full of air, it is a good sound reflector. The distance between the non-pressurized hull and pressurized hull will play a very important role in time delay calculation. It is necessary to adopt measures to prevent the effect of time delay between the double hulls. A reasonable measure is to set up some kind of baffle system in the non-pressurized hull. Fig. 8.32 explains how to use a suitable baffle for a surface ship and a double hull submarine.

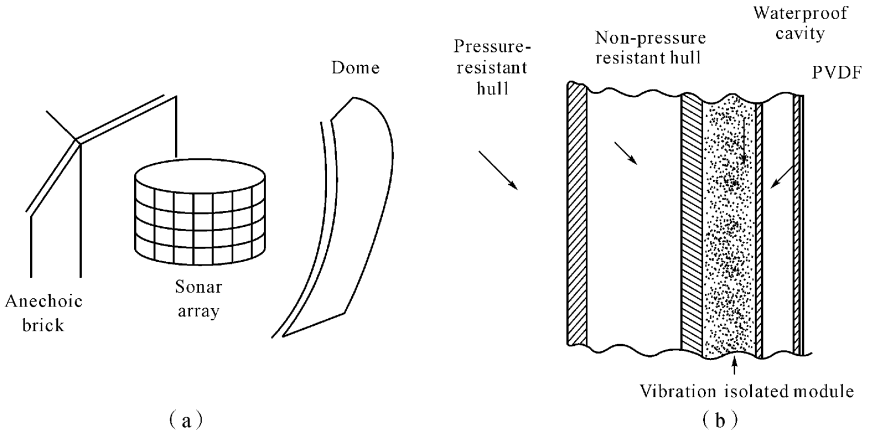


Fig. 8.32. Baffling of sonar array. (a) Surface ship; (b) Double hull submarine

Fig. 8.32(a) shows the baffle of a cylindrical array for a surface ship and Fig. 8.32(b) illustrates the structure of an air cavity baffle for a double hull submarine. The PVDF hydrophone is covered in the steel air cavity, and the latter is attached to absorption rubber. This structure will guarantee the signal is synchronized in target detection.

8.3.6 *Example of Typical Flank Array Sonar*

There are many different types of flank sonar that have been installed in submarines. Two typical flank array sonar systems are introduced here[136,137]. One is FAS 3-1, produced by STN Atlas. The wet end structure is illustrated in Fig. 8.33. All the hydrophones are omni-directional. Each four hydrophones have one acceleration sensor, which is used to pick up hull vibration. On both sides there are 192 hydrophones and the number of beams is 192, too. The tracking target number is up to 16. The system configuration is shown in Fig. 8.34.

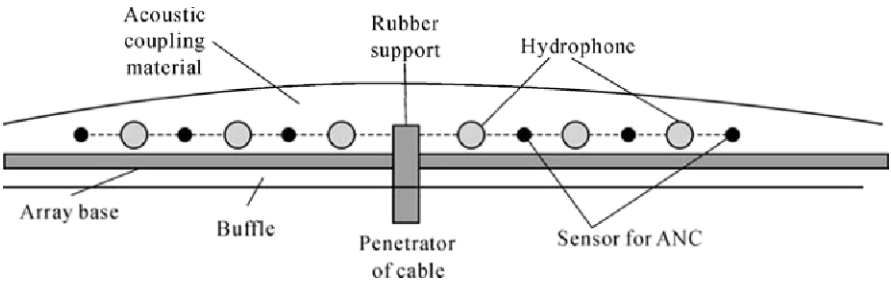


Fig. 8.33. Example of flank array installation (side view)

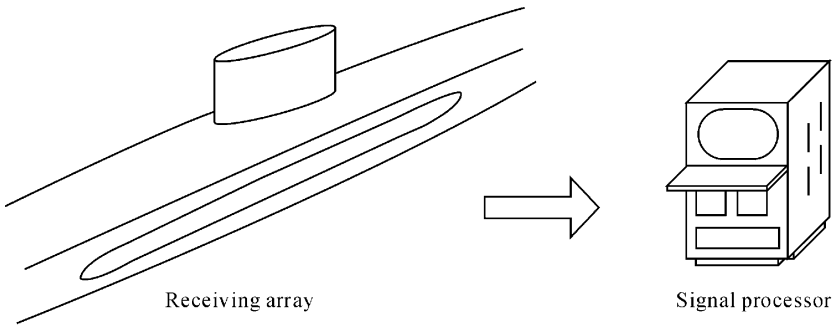


Fig. 8.34. Configuration of flank array sonar

The flank array sonar CORIS-5 developed by the RAFAEL Company in Israel is the other example. The wet end of this sonar is original in a towed array sonar and a necessary modification has been made to install it in a submarine. On each side there are 48 hydrophones, each group with 8 elements. The distance between adjacent elements is 0.5 m. The operation frequency is 10 – 1,500 Hz, sampling frequency is 5,120 Hz. The adaptive noise canceling filter has 32 weight coefficients. The beam number is 128. The frequency resolution ability is 0.375 Hz, with 1,280-point FFT. The nearest neighborhood criterion is adopted in the target recognition algorithm; the correct recognition probability is about 85%.

8.4 Fixed Array Sonar

Fixed array sonar is a special member of the sonar family. This sonar was developed a relatively long time ago. For example, the USSR BOJIXOB type sonar and US navy AN/FQQ-6 shore-based stations were developed in the 1950s[127,128]. The main characteristic of fixed array sonar is that the wet end is fixed, i.e., mounted on the sea bottom. This is the disadvantage of fixed array sonar. Once it is exposed, it is easy to destroy. But the advantage of fixed array sonar is obvious, for example

the relatively low level of background noise and the flexibility of the array aperture. Recently, the concept of old fashion fixed array sonar has changed. A new type of ADS (advanced deployment sonar) has been developed. The flexible sonar array of ADS can be easily deployed, which is critical in some cases.

8.4.1 Brief History of Shore-Based Sonar System

In the 1950s, the US navy began to install fixed array sonar on both the east and west coasts and the earliest system established was called the Caesar system. The detection range of this system is about 16 – 27 nautical miles. It was reported in the 1960s that 36 fixed array stations were mounted on the east coast of the USSR and on the sea of Japan. At the same time, the US navy developed a new SOSUS (sound surveillance system) in addition to the Caesar system. The maximum detection range is about 50 nautical miles. This is a distributed sonar array system; the signal is transmitted via optical fiber cable.

The original project in the development of an early alarming system for the US navy was to develop active/passive fixed array sonar. Since the 1970s a large amount of sea experiments have been carried out at Bermuda in the Atlantic Ocean. The project is named Artemus. The receiving array of an Artemus system is 3,000 feet long, and the transmitter array weighs a hundred tons. But the project was interrupted after 1.2 billion dollars had been spent.

There were some similar systems in the Soviet Union. The earliest type was the BOJIXOB shore-based station, which is called SOSS (soviet ocean surveillance system) in western countries literature. It was reported[139] in 1968 that a Soviet nuclear Golf Class submarine sank 1200 km north west of Hawaii; the SOSS did not detect this disaster, but SOSUS located it at a depth of 4,877 m.

The signal processing unit of fixed array sonar is at the shore station. The function of the sonar signal processor has been developing quickly. Now it is a combination of underwater fixed array with sonobuoy, satellite and surveillance vessel.

8.4.2 Outline of Shore-Based Sonar Station

Among all sonar systems, the shore-based sonar system has the longest cable connection between the dry end and the wet end. Because the signal receiving system is mounted on the sea bottom and the signal processing unit is at the shore station, the signal transmission cable is usually a coaxial cable or optical fiber cable. If the sonar system is active, it also has a bronze metal cable for transmitting a high power signal. Generally speaking, the wet end of a shore-based sonar system, including sonar array, transmission cable and signal repeater, will cost 70% – 80% of the entire system.

Fig. 8.35 illustrates the configuration of a shore-based sonar system. This is an active/passive fixed array sonar. The cylindrical transmitter and receiver array are installed on the sea bottom. The base of the array system should be mounted firmly on the sea bottom. The transmitter array is placed on the top of the array system. A special cable box is at the bottom of the base. This is a waterproof junction box used to install a pre-amplifier, filter and A/D converter, if necessary.

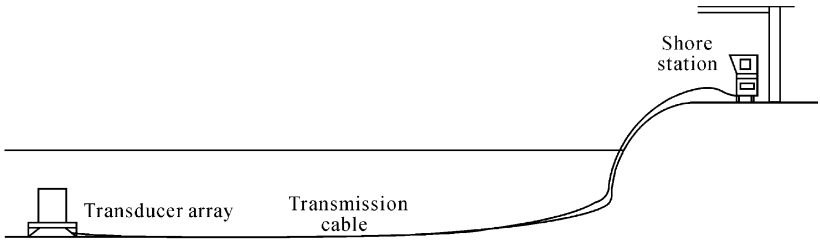


Fig. 8.35. Basic configuration of fixed shore sonar

The geophysical characteristics of the sea bottom, where one is supposed to mount the sonar array, have to be surveyed very carefully. The items are:

(1) Characteristics of sea bottom material;
(2) Ambient noise level according to season;
(3) Wave, tide and wind according to season;
(4) Sound speed profile according to season.

In other words, the sonar array deployment point has to satisfy the following conditions:

(1) Over a wide area, the sea bottom must be smooth, with no hollows or small hills.
(2) The sea bottom should preferably be sandy rather than consisting of soft sludge.
(3) The sea depth should be less than 100 m, so it is easy to carry out ocean engineering.
(4) The sea flow at the array deployment point should be less than 2 kn over the year, such that the force resulting from sea flow is limited.
(5) It should be far away from fisheries.
(6) There should be a relatively low ambient noise level.

A normal survey of the array deployment point is necessary each season, so we can make a precise prediction of sonar performance.

The most popular array form is cylindrical or linear array. The cylindrical array is often used in an active/passive array; the aperture of the receiver array can be 10 m. The mass is around several ten tons. The single linear array is used in passive shore-based sonar. It usually consists of 32 – 100 elements. The aperture of this kind of array may vary from 10 – 300 m. The double linear array system can be used in passive ranging. The array aperture is the same as that in a single array. The distance between two arrays can be from several kilometers to several tens of kilometers.

8.4.3 Some Specific Topics Concerning a Shore-Based Sonar Station

The wet end of fixed sonar is mounted on the sea bottom; it is supposed to have a long working life and very long MTBF. Therefore, the reliability of a sonar array is a very important issue in sonar design and the multiple channels and long distance signal transmission are also hot topics in ocean engineering.

8.4.3.1 Underwater Wired Signal Transmission

The underwater signal transmission via cable can be divided into two classes: analogue and digital. The media for a transmitting signal can be electrical or optical.

Fig. 8.36 illustrates three different block diagrams for signal transmission. In the analogue signal transmission mode (Fig. 8.36(a)), the signal from the hydrophone should be amplified and filtered and then connected to the multiple wires cable. If the distance from the array to the shore station is far, then it is necessary to install a repeater at some intervals in the sea.

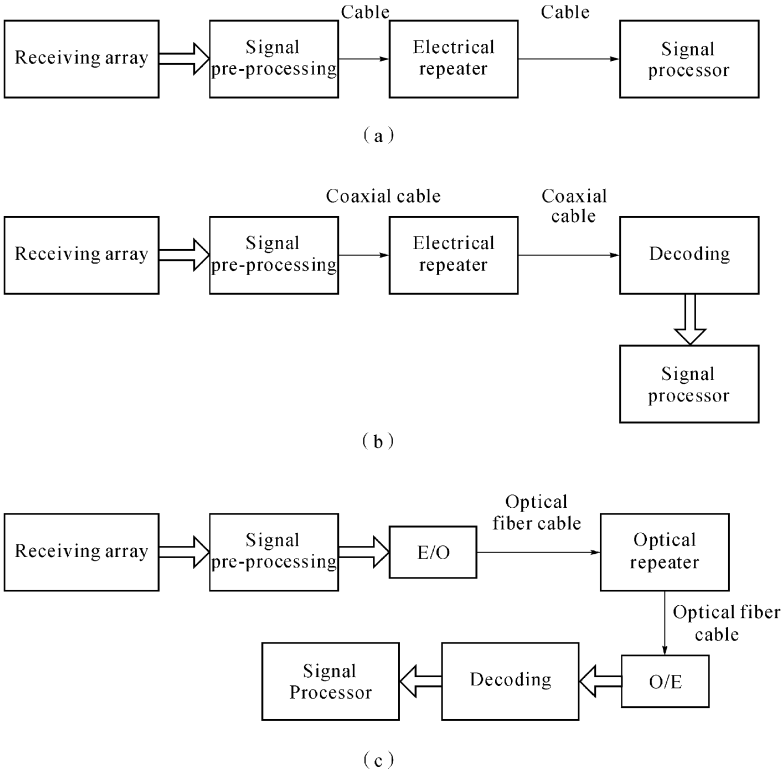


Fig. 8.36. Three types of signal transmission in fixed shore sonar. (a) Analogue transmission; (b) Coaxial digital transmission; (c) Optica fiber transmission

In digital signal transmission mode, the signal is converted from analogue to digital underwater and then transmitted by the digital coaxial cable. For the optical fiber transmission, it is necessary to transfer the electrical signal to the optical signal which is then transmitted by the optical fiber cable. (See Figs.8.36 (b) and 8.36(c))

As we have shown in Chapter 4, for a standard hydrophone the broadband ambient noise may result in about 10 – 20 μV, for the weak signal embedded in the background noise cannot produce a voltage of more than 20 μV. This kind of signal is not enough for long distance transmission. Therefore, it is necessary to have a pre-amplifier following in the hydrophone. The pre-amplifier should be low noise, broadband and high gain. Since the pre-amplifier is put underwater, it must be simple but reliable. Generally speaking, the amplifier gain can be selected as anything from 10 to 100. The attenuation of the multiple wire cable depends on the resistance of the cable. The typical value is about 1 dB/km. That means, for a 20 km transmission, the signal will decrease to 1/10 of the original signal. It is easy to understand that in some cases an underwater repeater is necessary. Otherwise, the weak signal transmitted by the analogue cable will be interfered with. The typical amplifier gain is about 100. But we have to say that the utilization of an underwater repeater will further decrease the reliability of the array system.

The block diagram of a fixed array sonar system with analogue signal transmission mode is shown in Fig. 8.37.

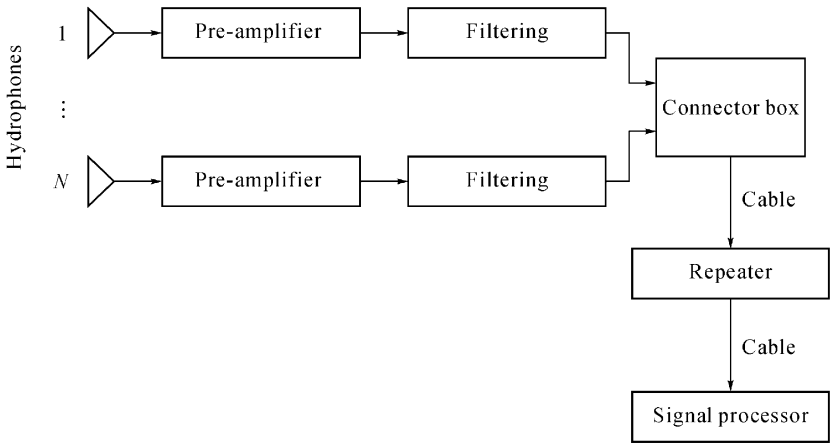


Fig. 8.37. Shore-based sonar with analogue signal transmission cable

Fig. 8.38 shows the block diagram of a shore-based sonar station with digital coaxial cable transmission mode. In this transmission mode, the signal should be amplified underwater, then multiplexed and converted to a 0 – 1 digital symbol flow. The coaxial cable transmits time/space hybrid data flow to the shore station. The decoding function is completed in the front of the receiver. Therefore the

signal processor obtains the multiple channels signal again.

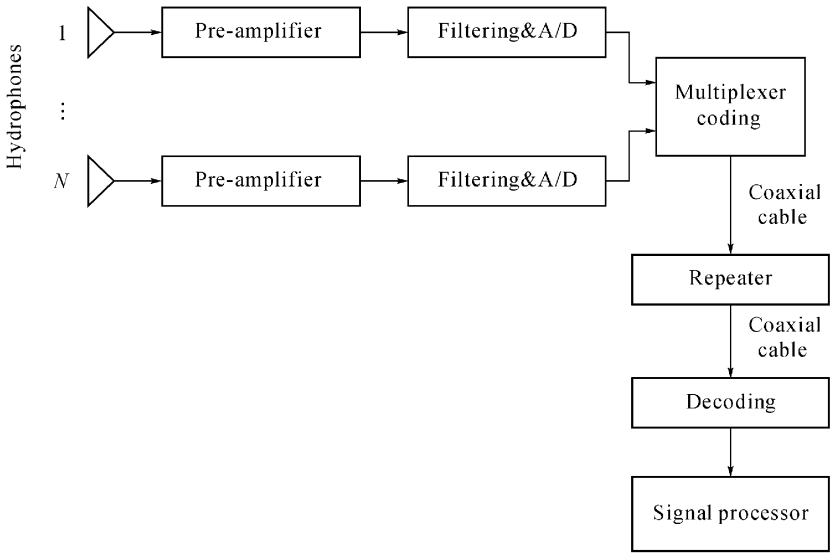


Fig. 8.38. Shore-based sonar with digital (coaxial) signal transmission cable

Example

Suppose the channel number is $N = 64$, the sampling frequency is $f_s = 20$ kHz. The A/D converter is 16 bits. Then the signal transmission rate should be 64×20 kHz/s $\times 16$ bit =20.48 Mbps.

The typical attenuation value of coaxial cable is about 3 dB/km (transmission rate = 10 Mbps, inner cable conductor $\varphi = 2.6$ mm, outer cable conductor $\varphi = 9.6$ mm). The attenuation value depends on the transmission rate and cable specification. But it is worth showing that a digital repeater is necessary in this transmission mode since the attenuation is considerable.

Fig. 8.39 shows the block diagram of a shore-based sonar station with digital optical fiber cable transmission mode. This is the main signal transmission mode in a modern fixed array sonar system. When the optical fiber cable is less than 80 km long, it is not necessary to have a repeater. Research shows that, for a special kind of crystal material optical fiber, there are 13 absorption peak values in a wavelength of 0.6 – 2.72 μm. If one chooses the right optical wavelength, the attenuation of the optical signal will be limited to 0.2 dB/km. The transmission rate can be as much as thousands of bits. It is necessary to have an optical terminal when we use optical fiber transmission. Now on the market, the most popular optical fiber is *ZS* multiple mode optical fiber and 1.55 μm single mode optical fiber.

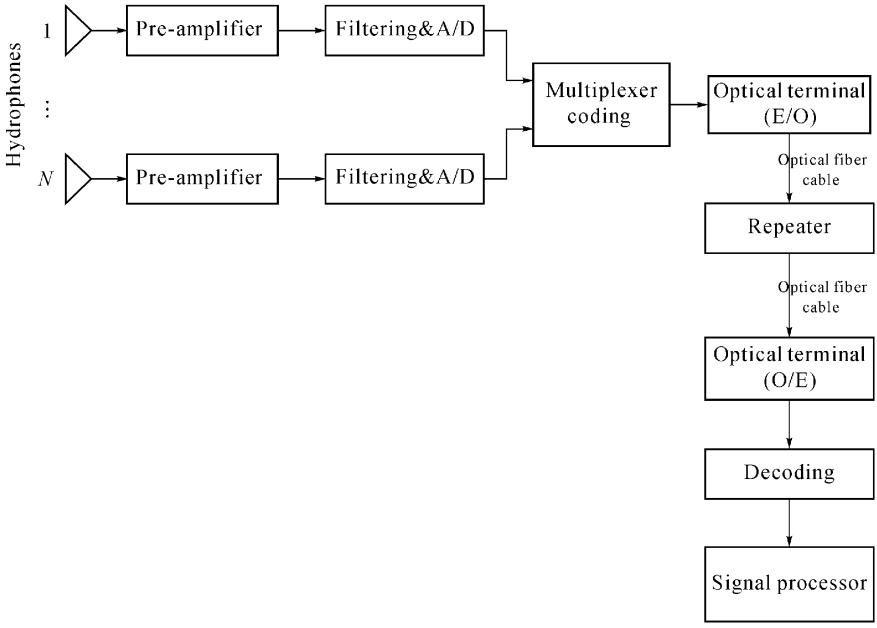


Fig. 8.39. Shore-based sonar with digital (optical fiber) signal transmission cable

The main part of the signal processor of a shore-based station is independent of the signal transmission mode. For the single array system, the sonar operates in passive detection mode. For the two array systems, the sonar has the function of passive ranging. And the signal processing techniques are different.

8.4.3.2 Reliability of Underwater Equipment

Much equipment in fixed array sonar is installed underwater, including the cable box, repeater, hydrophones, cable and optical terminal. Therefore the reliability of underwater equipment becomes a very important issue. It is desirable that once underwater equipment is deployed, it will work well and will have no failure over time. But this is almost impossible since high reliability means greater cost. Eq. (5.122) shows that the mean time between failures is related to the system failure rate, i.e., $MTBF = 1/\lambda$, and the reliability function is $R(t) = \exp(-\lambda t)$.

It is necessary to point out that the concept of MTBF is not equivalent to asking "how long will it be before we need to salvage the equipment?". For example, if we wish the probability that we do not need to salvage equipment within 5 years to be 0.8, then we can calculate that $MTBF = 22.72$ years. If we wish the probability to be 0.08 within 10 years, then $MTBF = 45.45$ years. So we can understand the difficulty of the reliability problem for underwater equipment. It is necessary to decrease the number of components as much as possible. We have to use the most reliable components in underwater equipment. In the 1970s,

the Naval Research Laboratory in the USA established an expert group for testing the reliability of underwater equipment, including underwater cables, transducers and connector, etc. A proposal for ALT (accelerated life testing) has been reported, which can speed up the test procedure and reduce the test period. The standard presented by the ALT group can reduce the test period to 1/15[139]. The factor which affects the lifetime of a shore-based sonar system is damaged to the underwater cable. Since the cable of a shore-based system is installed in shallow water, and is usually exposed on the sea bottom, damage is inevitable, including from artificial and other causes. It is reported by AT&T Labs. in the USA that an optical fiber cable supplied by direct current will produce a special electrical field, which can result in a shark attack. Only armored FBP SL cable can resist shark bites[140]. The optical fiber cable named OPTICAN-1, deployed in September of 1985 by AT&T from Gran Canaria Island and Tenerife Island was bitten by a shark on October 17, 1985. When they adopted the right technique to maintain the cable, this did not happen any more.

8.4.4 Example of Typical Fixed Array Sonar

The shore-based sonar system is extensively used in port protection. The SONAC PFA (passive fixed array), developed by Patria Oy, is a typical example of a shore-based sonar station [128]. This is a double linear array system. The warning area of this system may cover 1,000 – 1,500 km^2. The transmission cable is optical fiber and the distance between the array and the shore station is limited to 40 km. The electronic cabinet of a sonar system has four display screens.

There are also many shore-based sonar stations distributed around the world's oceans. Some are used in ocean monitoring and most of them are used for military tasks. We will discuss this later in the section on "network centric warfare".

8.5 Synthetic Aperture Sonar

SAS (synthetic aperture sonar) is one of the hottest topics of the past 30 years in the field of underwater engineering. As a matter of fact, research studies on SAS not only concern SAS itself, but also the synthetic aperture technique in underwater signal processing[47,49,76-83]. This means the technology in terms of the relatively small physical size of the aperture of the array to obtain the rather higher resolution ability of a large array. The synthetic aperture technique can be used not only in radar and sonar areas, but also in areas such as separation of target radiated noise, in highlighting spot analysis of a submarine echo signal and in inverse beamforming[99-102].

The basic theoretical principles and technical design of SAS[142,143] and the

synthetic aperture technique are introduced in this section. A prototype of SAS and the results of system simulation and experiments are illustrated.

8.5.1 Brief History of SAS

Radar and sonar look like twin brothers and the development history is dramatic. The invention of sonar came before radar, but the development of sonar techniques was slower than radar. The main reason is that the medium of sonar, sea water, is much more complicated than the medium of radar, the atmosphere or a vacuum.

In 1935, based on the reflection principle of an electric/magnetic wave, Walter tested a radar system for detecting an aeroplane. The range he reached was 12 km. After that, he quickly improved his equipment and realized a far greater detection range of up to 80 km. Therefore, the invention of radar came later than sonar by about 20 years, but the subsequent development has been faster than sonar. A typical example is SAR (synthetic aperture radar).

The research work on SAR started in the early 1950s. A practical system was developed by Wiley of the Goodyear Aircraft Corp. in 1951. However, its first experimental validation was carried out in 1953 by a group of scientists at the Univ. of Illinois. The first operational SAR system is believed to be X-band (3 cm wavelength), built in 1957 by Willow Run Lab. of the Univ. of Michigan for the US Department of Defense. The first generation digital SAR was born in 1970s [144-148,162].

Because of the similarity in radar and sonar design principles, after the appearance of SAR then SAS attracted much research interest. But the advance in SAS research work was very limited and the main interest in SAS studies was centered on the passive mode. The big difference in the propagation medium between radar and sonar is the main reason that SAS research encountered difficulties. For example, the velocity of an electrical/magnetic wave is about 3×10^8 m/s. In order to detect a target within a range of 300 km, the repeat impulse period of SAR is about 2 ms, i.e., in 10 ms duration it is possible to achieve 5 times integration of the echo signal. But in underwater signal detection, the sound velocity in sea water is about 1,500 m/s. In order to detect the target within a range of 1 km, the repeat period of impulse must be at least 1.3 s. Therefore, in using SAS we are facing a very critical problem: how to transfer the time gain to spatial gain?

Research work on SAS started in USA and Japan. Dr. Higgins of NCSL reported that in 1975 a special signal processing algorithm was developed to reconstruct an SAS image. In Japan, earlier work on SAS was carried out in Sato's group. It reported in 1979 that they had found an efficient algorithm which could overcome the effect of medium turbulence in SAS imaging[149].

In 1980s, many new techniques were developed to design an SAS prototype. And a series of experimental results have shown potential application prospects for SAS.

8.5.2 Basic Principle of SAS

In order to better understand the basic principle of SAS, we will review how conventional sonar detects the target on the sea bottom. Fig. 8.40 illustrates the target detection of side scan sonar. The sonar consists of a transmitting transducer and a receiving linear array. They are both towed by the platform. In the towing process, the transducer transmits a signal to the sea bottom continuously, and the receiving array receives the signal from the sea bottom and the target, if it exists.

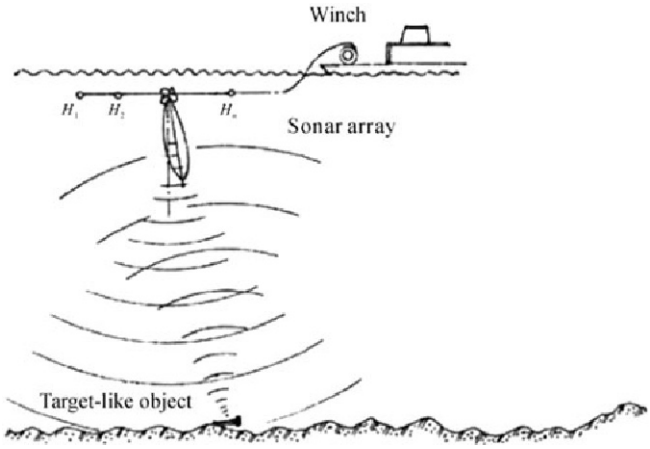


Fig. 8.40. Conventional sonar scans sea bed

When the target-like echo does appear, the signal processor makes a decision to confirm whether a target is present. The performance of this kind of side scan sonar depends on the movement of the platform. The detection efficiency is related to the vessel's speed. It also depends on other factors, such as the depth of sea water, repeat period of impulse, beam width of transmitted signal. The receiving array is always a multi-beam system since the multi-beam system has higher system gain and detection efficiency. Fig. 8.41 is an example of multi-beam side scan sonar. In Fig. 8.41, the beam width of the transmitted signal is φ. It usually covers a wide area of the sea bottom. But the beam width of the receiving system is much narrower than φ.

The bearing resolution ability of a linear array is proportional to λ / L, where λ represents the signal wave length and L is the length of the array. Therefore

$$\Delta\theta = \lambda / L \tag{8.39}$$

If the distance between the array and target is R, then the range resolution on the sea bottom plane is

$$\Delta x = \lambda R / L \tag{8.40}$$

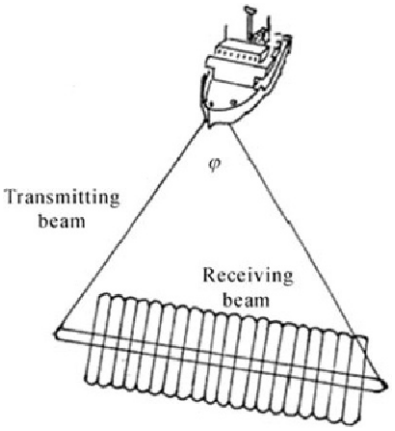


Fig. 8.41. Illustration of multi-beam sonar scanning sea bed

We can see from Eq. (8.40) that in order to obtain higher range resolution ability on the sea bottom plane, it is necessary to decrease the value of zs.

Example

Suppose $L = 10$ m, $\lambda = 0.3$ m (i.e., $f = 5$ kHz), $R = 1{,}000$ m, then $\Delta x = 30$ m.

There are two ways we can increase the resolution ability on the sea bottom plane: to increase the array aperture and to increase the signal frequency. These two methods are constrained in some ways. The former is related to the platform size and the latter is related to the transmission loss of the underwater acoustic signal.

The key point in the synthetic aperture technique is to use a relatively small size array and get a higher resolution ability, which a large array has. SAS is the equipment for realizing this technique. Fig. 8.42 is a simple illustration of the SAS principle. The array length is L. The original position of the array is AB. The signal received at point A is $x(t)$. The signal received at point B is $x(t+\tau)$ and

$$\tau = L\sin\theta / c \tag{8.41}$$

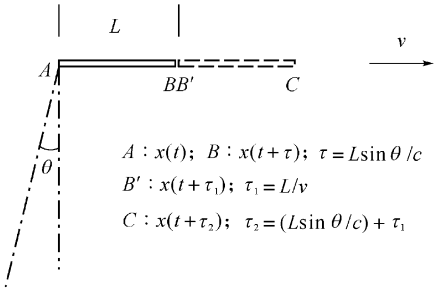


Fig. 8.42. Principle illustration of SAS

Suppose the array is moving uniformly from AB to $B'C$ from moment t. The moving speed of the array is v. It follows that the signals received in B' and C are $x(t+\tau_1)$ and $x(t+\tau_2)$, respectively

$$\tau_1 = L/v, \quad \tau_2 = L\sin\theta/c + \tau_1 \tag{8.42}$$

Since we can achieve beamforming for the physical array AB, therefore it is possible to achieve beamforming for the virtual array AC, if we can calculate the necessary time delay value. If we know the moving speed of the array, then it is not difficult to do this. The block diagram of the beamforming algorithm of SAS is shown in Fig. 8.43.

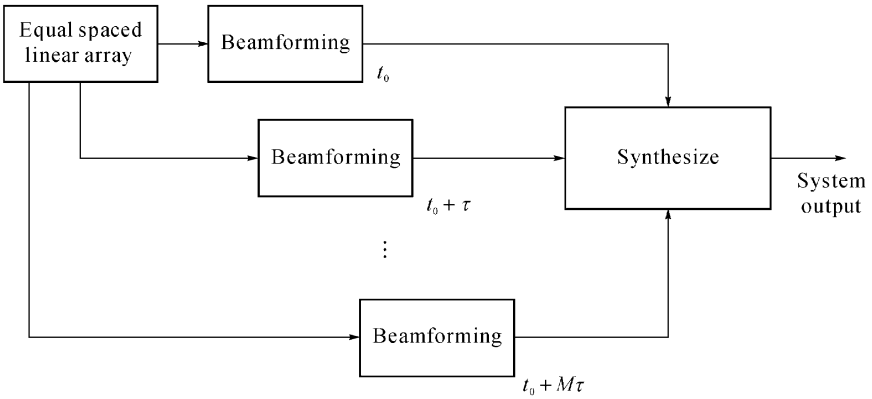


Fig. 8.43. Beamforming of SAS

As we can see in this figure, compared with a conventional beamforming system, it is necessary to store received data in SAS beamforming. At each moment t_0, $t_0+\tau$, ..., $t_0+M\tau$ the beamforming data are stored and synthesized to obtain the entire beam output of SAS.

In the case where the movement status is unknown, or the moving speed is not uniformly straight linear, then the beamforming operation of SAS will fail, or the performance of SAS will be degraded seriously.

Now we will introduce some basic concepts in SAS detection (Fig. 8.44).

8.5.2.1 Resolution Ability of Along-Track and Cross-Track

If the sea bottom is considered to be a plane and the line on which the SAS array is placed is the X direction, the array is moving along the X axis. In the detection sense, there are two kinds of resolution ability on the sea bottom. One concerns the moving direction of the array and it is called the resolution of the along-track. The other is called cross-track resolution ability, which indicates the resolution

ability of SAS in the perpendicular direction to the array moving direction. It is easy to understand that the resolution ability of along-track is usually much better than that of the resolution ability of cross-track. In Fig. 8.44(a), Δx represents the resolution ability of along-track and Δy represents the resolution ability of cross- track. Sometimes Δy is called the side view resolution ability of SAS.

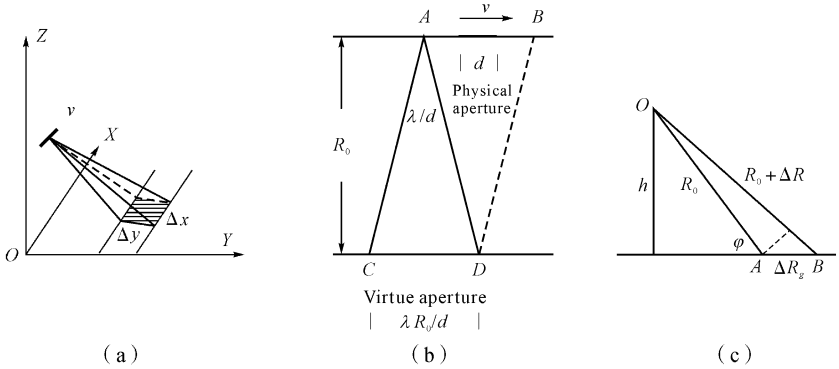


Fig. 8.44 Illustration of basic concept of SAS

8.5.2.2 Resolution Invariant Principle of SAS

In Eq. (8.40) we know that the resolution ability of conventional sonar for the sea bottom target depends on the length of the array and the distance between the array and the target. In order to increase the resolution ability, it is necessary to reduce the distance between the receiving array and the target. But it is not always possible. A very important advantage is that the resolution ability of SAS is independent of the distance between the array and the target; it is even independent of the signal frequency (see Fig. 8.44(b)). Suppose the length of the array is d and it is moving at a uniform speed v. We know that the bearing resolution ability of this array is λ / d and λ is the wavelength of the signal. If the vertical distance between the array and the sea bottom is R_0, therefore on the sea bottom the range highlighted by the transmitted signal is $\lambda R_0 / d$. When the array moves from A to B, the target in position C is not highlighted. Therefore the maximum aperture of the virtual array is

$$AB = L_{\text{SAS}} = \lambda R_0 / d \tag{8.43}$$

For the array of length L_{SAS}, the bearing resolution ability of the along-track, taking account of the double travel time delay, is given by

$$\Delta\theta_{\text{SAS}} = \lambda / 2L_{\text{SAS}} = d / (2R_0) \tag{8.44}$$

The range resolution ability of the along-track is

$$\Delta R_{\mathrm{SAS}} = R_0 \Delta\theta_{\mathrm{SAS}} = d/2 \tag{8.45}$$

From Eq. (8.45) we can see that ΔR_{SAS} is independent of the distance and frequency. It only depends on the physical aperture of the array. Of course, d represents the physical size of a single transmitter and it is definitely related to the frequency. It is proved that in the case of a multiple elements array, Eq. (8.45) is also approximately valid.

This fact is called the resolution invariant principle of SAS.

8.5.2.3 Minimum Separable Range on Sea Bottom

In Eq. (8.40), based on the view angle of the directivity function, the resolution range ability of the along-track of conventional sonar is $\Delta x = R\lambda / L$.

Generally speaking, this ability is far from the application requirements. In order to distinguish two targets on the sea bottom, what kind of resolution ability is necessary? In Fig. 8.44(c), suppose O is the transmitting point and A, B are two targets: $OA = R_0$ and the angle between AB and the incidental direction to A is φ. The duration between signals transmitted from O to receive the echo from A can be expressed as

$$t = 2R_0 / c \tag{8.46}$$

where c represents the sound velocity. The time duration of the echo from target B should be

$$t + \Delta\tau = 2(R_0 + \Delta\tau)/c \tag{8.47}$$

where $R_0 + \Delta R$ is supposed to be the distance from O to B and

$$\Delta\tau = 2\Delta R / c \tag{8.48}$$

In order to distinguish the targets A and B, it is necessary to separate the echoes from A and B. This means the time difference $\Delta\tau$ can be estimated. It is proved that the distance between A and B is

$$\Delta R_g \approx \frac{\Delta R}{\cos\varphi} = \frac{c\Delta\tau}{2\cos\varphi} \tag{8.49}$$

This is the expression of minimum separable range of two targets on the sea bottom plane.

Example

If the minimum time delay estimation value is $\Delta\tau = 500\ \mu s$ and $\varphi = 0^\circ$, then

$$\Delta R_g \approx \frac{1.5 \times 0.5}{2} = 0.375\ \text{m}.$$

8.5.2.4 The Repeat Frequency of Transmit Impulses

From Eq. (8.45) we know that the separable time duration of two echoes is $\Delta\tau = 2\Delta R / c$. Therefore the repeat frequency f_r of the transmitted impulse must satisfy the inequality shown as follows

$$f_r \le 1/\Delta\tau = c/(2\Delta R) \tag{8.50}$$

On the other hand, based on the Eq. (8.49), in each $d/2$ distance (sea bottom resolution range), the target should be highlighted at least once, i.e.,

$$f_r \ge 2v/d \tag{8.51}$$

Combining Eqs. (8.50) and (8.51), we obtain the requirement for repeat frequency of the impulse

$$\frac{2v}{d} \le f_r \le \frac{c}{2\Delta R} \tag{8.52}$$

This equation can be considered as the basic constraint expression for the parameters d, f_r, v and ΔR.

8.5.3 Calculation of Directivity Function of Passive SAS

Passive SAS does not transmit a signal; it only receives the radiated signal from the target. Based on the uniform speed linear movement, passive SAS collects the received time/spatial signal and stores it. The simulated large aperture signal is produced and processed to obtain the beam pattern of SAS (Fig. 8.45). Suppose in time moment t, the physical array is in position OA. The distance of adjacent elements is d. The signal incidental angle is θ. The signal of the i-th element is

$$x_i(t) = x(t - (i-1)d\sin\theta / c) \tag{8.53}$$

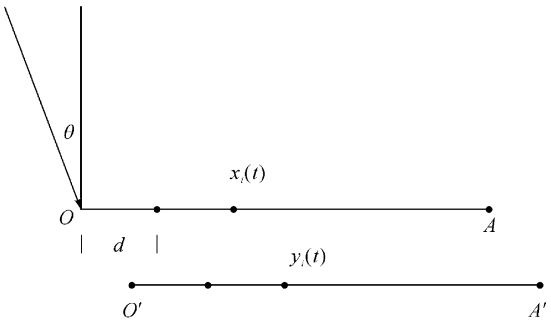


Fig. 8.45. Illustration of time/space sampling equivalence

The time sampling data can be expressed as

$$x_i(kT_s) = x(kT_s - (i-1)d\sin\theta / c) \quad (k = 0,\ 1,\ 2,\ ...) \tag{8.54}$$

where T_s represents the sampling duration.

If the array moves to position $O'A'$ with uniform speed along the straight line, then in time moment $t+\tau$, the *i*-th signal will be

$$y_i(t) = x(t - \tau - [(i-1)d + \tau v]\sin\theta / c) \tag{8.55}$$

The sampling value is $y_i(kT_s) = x(kT_s - \tau - [(i-1)d + \tau v]\sin\theta / c)$. Suppose $\tau = k_0 T_s$, then

$$y_i(kT_s) = x((k-k_0)T_s - [(i-1)d + \tau v]\sin\theta / c) \tag{8.56}$$

When we carefully study Eqs. (8.54) and (8.56), we find that if we choose the right value of k_0, such that $k_0 T_s v = d,\ 2d,\ 3d,\ ...$, then the series of $y_i(kT_s)$ and $x_i(kT_s)$ are the same. In this case, the spatial samples are already transferred to time samples and the aperture of the virtual array has been enhanced. It is proved that the directivity function of passive SAS for *M* times sampling is given by

$$D(\theta,\theta_0) = \left| \frac{\sin[N\pi d(\sin\theta - \sin\theta_0)/\lambda]}{N\sin[\pi d(\sin\theta - \sin\theta_0)/\lambda]} \times \frac{\sin[M v\tau\pi(\sin\theta - \sin\theta_0)/\lambda]}{N\sin[\pi v\tau(\sin\theta - \sin\theta_0)/\lambda]} \right| \tag{8.57}$$

This is actually the product of two directivity functions of linear arrays. One is an equal-spaced linear array with *N* elements and duration *d*; one is an equal-spaced linear array with *M* elements and duration $v\tau$. Fig. 8.46 illustrates the simulation results for $N = 20$ and $M = 8$.

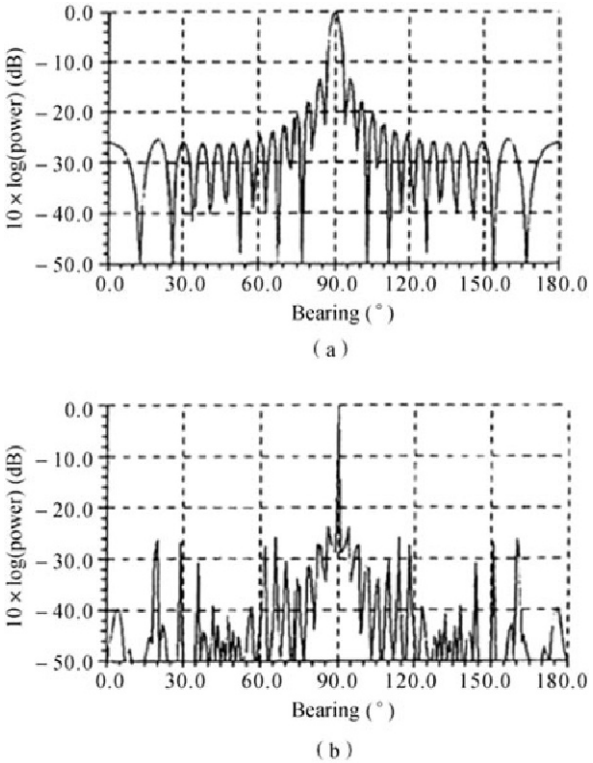


Fig. 8.46. Simulation results of passive SAS (Dec. 1989, IACAS). (a) Physical aperture: 20 m; (b) Synthesize aperture: 8 m×20 m

8.5.4 Basic Algorithm of Active SAS

The image reconstruction algorithm is the kernel technique of active SAS. It is also the basis for other synthetic aperture techniques, including error compensation of kinetics, self-focus and interference SAS, etc. The basic contents of the image reconstruction algorithm contain the technique for reconstructing the target image based on the echo signal and the digital signal processing algorithms.

Suppose the transmitted signal is $s(t)$, the received signal can be expressed as

$$x(t) = s(t - 2R(\tau,\ r)/c) \tag{8.58}$$

where $R(\tau, r)$ represents the distance between the target and array, r is the distance between the target and array in the time moment t, τ is the moving time of the array. When $t = 0$, we have $R(\tau, r) = r_0$. It follows that

$$x(t) = s(t - 2r_0/c) \tag{8.59}$$

The various algorithms of active SAS were developed based on signal processing for $x(t) = s(t–2R(\tau, r)/c$. The frequently used methods are briefly introduced in the following[77-83]. It is worth stating that the methods and algorithms described here are very preliminary, and far from well established, since the active SAS technique is at a fast growing stage.

Point-Wise Imaging Algorithm

For each received signal from hydrophones, it is necessary to calculate the time delay value of $2R(\tau, r)$ / c and then take time compensation and focusing to reconstruct the image. The calculation is relatively large, but since the time delay can be compensated accurately in the range and bearing dimension, the quality of focusing is very good.

Range-Doppler (R-D) Algorithm

For some kinds of transmitted signals, for example the linear frequency modulation signal, the argument of range r and argument of bearing τ can be separated. Therefore, it is possible to perform FFT for one argument and keep another fixed. Finally, the image is reconstructed by taking two folds FFT.

***ω-k* Algorithm in Wave Number Domain**

This is also a two dimensional Fourier transform method; the received data are transformed to the wave number domain. An algorithm, called Stolz mapping, is used to reconstruct the image and this algorithm provides a better resolution of multi-targets.

Chirp-Scaling Algorithm

This algorithm is suitable for the signal of a large time bandwidth product. The range to be searched is divided into many range bins. In each range bin a special factor is multiplied. Thereafter, FFT is performed to get a good quality image.

8.5.5 An Experimental SAS Prototype[150,151]

The Institute of Acoustics, Chinese Academy of Sciences and Hangzhou Applied Acoustic Institute of the China Ship Industry Corp. jointly developed an experimental SAS prototype in 1998. Successful lake experiments were carried out in 1999 and 2001. Thereafter, a subsequent sea experiment was performed.

This is an active SAS system. The block diagram of this system is shown in Fig. 8.47. The wet end consists of a transmitter array and receiver array. It is installed

in a tow fish. Fig. 8.48 shows the configuration of the wet end of this prototype.

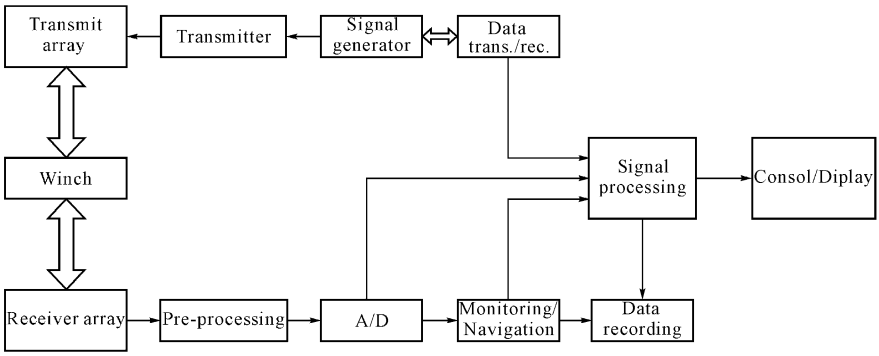


Fig. 8.47. Block diagram of an experimental SAS

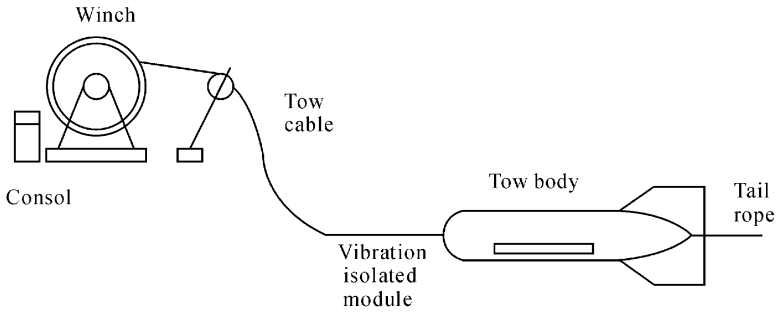


Fig. 8.48. Consists of a towed system of SAS

In order to study some special physical problems in underwater acoustics which appear in SAS research work, some necessary experiments have been undertaken. In particular, the horizontal and vertical correlation performance of the underwater acoustic signal was measured and the steadiness of signal transmission was also studied. Fig. 8.49 illustrates one of these experiments, which was carried out on Xin'an Jiang Lake in Southeast China. Some results are given in Figs. 8.50 and 8.51.

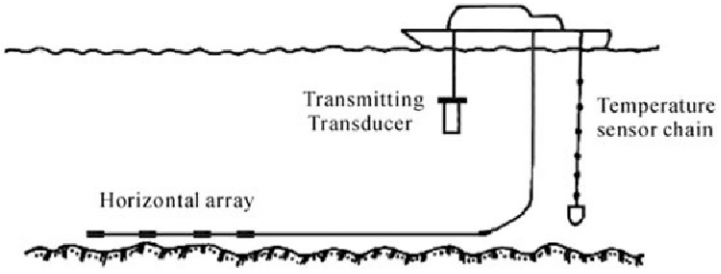


Fig. 8.49. Configuration of SAS experiment on lake (April 1998, IACAS and HAAI)

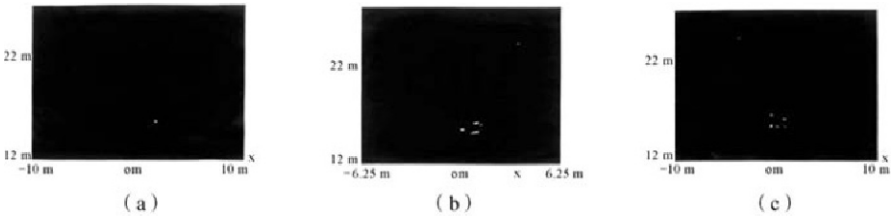


Fig. 8.50. Experimental results of SAS at lake (April 1998, IACAS and HAAI). (a) One target; (b) Three targets; (c) Five targets

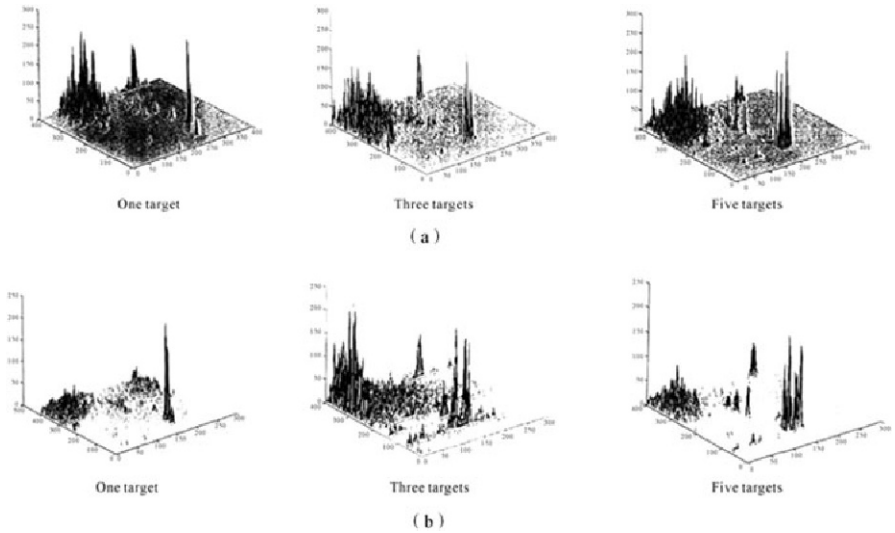


Fig. 8.51. Comparison of CBF and ω-κ algorithm. (a) Conventional beam forming algorithm; (b) ω-k algorithm

In Fig. 8.50, the images of one target, three targets and five targets are reconstructed. The resolution ability of the range is about 20 cm. Fig. 8.51 is a comparison of some different algorithms. Among these algorithms, the ω-κ algorithm shows better performance than others.

The research on SAS has attracted much interest in the underwater acoustics community. Many companies, institutes and universities are involved in the development of the SAS system. Fig. 8.52 is an example of basic research on SAS, which was carried out in the APL (applied physical laboratory) in Washington University[152]. Two 14 m high towers were used to transmit and receive the signal. Fig. 8.53 shows the experiment results of the NATO SACLANT anti-submarine center[110]. It is reported that some mines, embedded in the sands and mud, were explored. A successful experimental result and clear image were published by DTI (dynamic technology Inc.) in September 2000. It was reported, in April 1997, that a wreck of a PBY-4 aeroplane, which sank in 1956 in Washington Lake, was

found by an SAS system developed by DTI. They focus the image at a distance of 350 m and 1,000 m. The resolution ability in the along-track and cross-track is 3 cm and 11 cm, respectfully. The reconstructed image is very clear, even the ribcage of the aeroplane can be identified. Although this wreck of an aeroplane was found by side scan sonar, the image quality is far from that of SAS. This is an example of a successful application of active SAS.

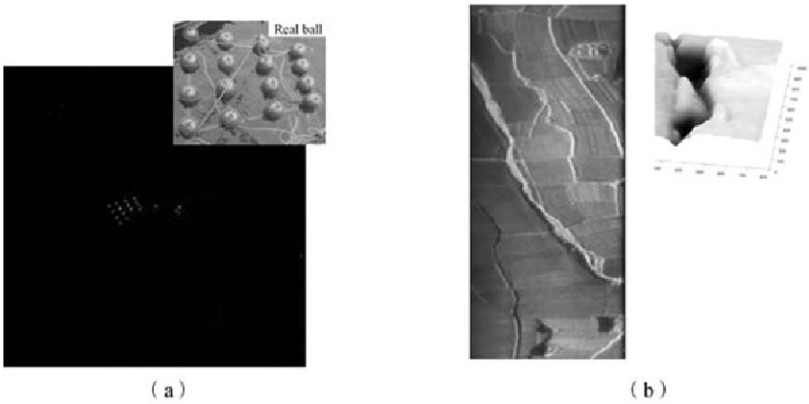


Fig. 8.52. Experimental results of SAS (August 2001, IACAS and HAAI). (a) 16 balls and its images; (b) Image of lake bottom and its zoom image

The main specifications of the SAS system are as follows:

(1) Frequency: 12.5 kHz
(2) Signal format: CW signal, LFM and M-series
(3) Swath range: 800 m
(4) Resolution ability: along-track 10 cm
(5) Size of tow body: 2 m

Table 8.1 lists some typical SAS systems.

Table 8.1 Examples of SAS system

Name	SAMI	Kiwi	TOPAS	MCM	DARPA	CEROS
Organization	ECU	New Zealand	Simrad	North Gruman	DTI	Univ. of Hawaii
Length of Array (m)	2.0	0.3		0.5	1.5	1.5
Center frequency (kHz)	8.0	30		180	50	12.5
Tow speed (m/s)	1.5	0.25 – 0.5	0.1 – 0.25	2.0	1 – 5	1 – 5
Detection range (m)	<2,500	<100	<100	<100	<1,000	<1,000
Resolution ability (m) (cross-track/along track)	0.5	0.15	0.4	0.075/ 0.025	0.11/ 0.03	0.20

8.6 Concept of Network Centric Warfare and Its Application in Sonar

The concept of NWC (network-centric warfare) was proposed by the Chief of Naval Operations, Admiral Jay Johnson in 1997[51]. He said "a fundamental shift from what we called platform-centric warfare to something we call network-centric warfare". The vice-admiral of the US navy, Cebrowski, thinks " it will prove to be the most important revolutionary military affair of the past 200 years"[91]. The concept and practice of NWC was quickly acknowledged by military officers and scientists. The first meeting of NCW was held in October 2000. It seems that NCW has become the most important issue in the research area, which is related to strategy, tactics and armament development.

Of course, the concept of NCW involves the army, navy and air force. But it was first proposed by the navy, and has already been extensively studied and developed by the navy. In this section, we will introduce some basic material about NWC and consider the basic applications in digital sonar design.

8.6.1 Basic Concept of NCW

NCW is the best term for describing warfare organization and operation in the information age. We define NCW as a warfare concept for obtaining information superiority. Based on a network connection of sensors, decision makers and armaments, it is possible to realize knowledge sharing, therefore increasing the command speed, the operational rhythm and enhancing the power to kill and wound, as well as the ability to survive. Therefore, the concept of NCW is in the transfer of information superiority to the ability to make real war.

The concept of NCW was born a short time ago, and the theory and practice is just in its early stages. Therefore it is far from complete. Just as Moffat[85] shows the definitive book about NCW could be written 30 years from now.

It is obvious that war in the information age is different. In the past, military organizations were usually in the leading position regarding research and technology. But in the information age, the advances in information technologies considerably depend on industries in the civilian, commercial area. The characteristics of the information age include the following:

(1) The fashion of wealth creation has changed;
(2) The distribution of global relative to local power has changed;
(3) The complexity of society and the participants of warfare have changed;
(4) The world is a smaller place;
(5) Decision times are compressed and the rhythm of life is faster.

We would like to say that information technology is the DNA of the information age. Information and information technology are providing ever greater opportunities

for companies to create value, which was impossible before the information age.

The development of network technology provides the most powerful support for NCW. Metcalfe's law, says the value or the efficiency of a network is proportional to the square of the node number (Fig. 8.53). Therefore the speed of NCW development is surprising.

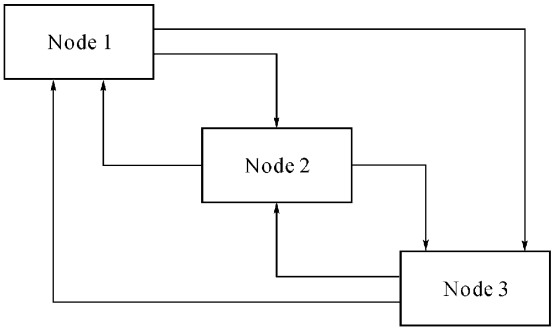


Fig. 8.53. Explanation of Metcalfe's Law

The primary task of NCW is to get superior information by using various sensors and information processing methods. Roughly speaking, superior information includes three aspects: timeliness, relevance and accuracy. Fig. 8.54 shows the concept of information superiority, which we must pay attention to in NCW.

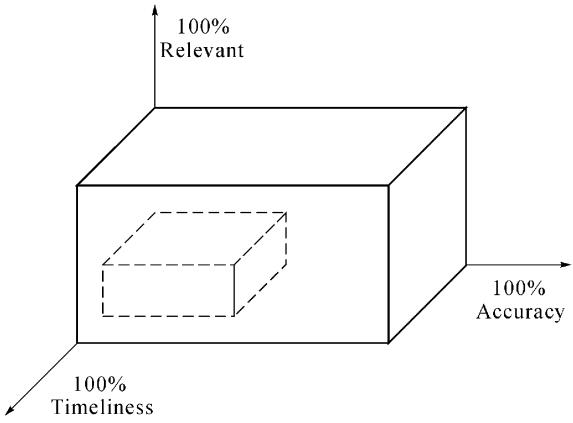


Fig. 8.54. Illustration of superior information

In the enterprise community, successful examples based on providing superior information include, for example, the Dell Computer Corp., Cisco Systems, Federal Express, American Airlines and Wal Mart, etc. But it is worth pointing out that the theory and practice of NCW is not only related to the technology but also, in a wide sense, to military operations. We should understand that NCW is the

natural response of the information age to warfare. What we are concerned with, of course, is the application of NCW in sonar design.

The basic idea of NCW can be understood from Fig. 8.55. We can see that the infrastructure of information technology and the various sensor networks are the basic support for NCW.

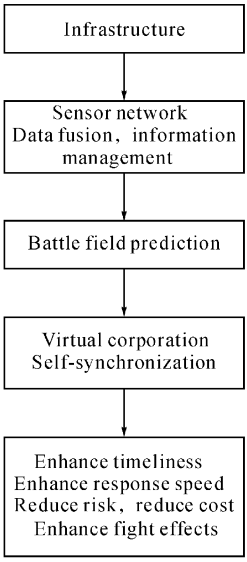


Fig. 8.55. The military as a network centric enterprise

8.6.2 Theory of Complexity System

The research object of NCW is military warfare. The exact description of military warfare in technological terms is very difficult. A possible theoretical way describing a military operation is by complexity system theory. Most research results show that complexity system theory can be used to explain warfare operations, the relationship between warfare participants and the events which may occur in warfare procedures.

A system is called complex if it consists of many relevant entries and many hierarchical layers. The entire characteristics of a complexity system cannot be simplified by the characteristics of some single entry or some of the entries which belong to this system.

An ecosystem is a typical complexity system. Any individual characteristic of a creature is far from being enough to describe the entire characteristics of an ecosystem.

Another feature of a complexity system is that the system is very sensitive to small disturbances. Therefore, behavior prediction becomes very difficult. We cannot expect that an event that happened in the last moment will reappear in the

next moment.

It is reasonable to consider NCW as a complexity system. Sonar design is only a small but important entry in NCW.

Military warfare in the information age has many characteristics that are similar to that of a complexity system. Table 8.2 shows some corresponding items.

Table 8.2 Corresponding concepts of NCW and a complexity system

Complexity system theory	Warfare in information age
Non linear cross-operation	Warfare consists of large amount entries, which have non-linear relationship
Distributed control	Impossible to predict the result of individual battle
Self-organization	Local operation usually is "chaotic"
Non-equilibrium	The essence of military conflict is equilibrium
Adaptability	The participants in warfare have to adapt to the change of environment

NCW derives its power from the strong networking of a well-informed but geographically disposed force. The enabling elements are a high-performance information grid, access to all appropriate information sources, weapons reach and maneuvers with precision and speed of response.

8.6.3 US Navy Research Project in NCW

Since the concept of NCW was proposed in 1997, the US government and military actively push forward research work and application of NCW. In September 2004, 28 major western industrial companies, which manufacture military equipment for NATO, established the NCOIC (network center operation industry consortium). The members of this organization include BaE, Boeing, EADS, Ericsson, HP, Microsoft, Saab, IBM, Thales and Lockheed Martin.

The relevant projects of NCW are briefly introduced in the following.

IT-21 (Information Technology for the 21th Century Initiative)

The goal of this project is to establish a global wideband network for the US navy and to reconstruct the existing C4I (command, communication, computer, control and intelligence) system.

The main network consists of 150,000 computer operators and supports 270,000 users distributed worldwide. This network will connect 300 commands, bases and about 450,000 workstations. It has been running well since 2001.

Aegis

The early studies and research work on the Aegis system, which is a fighting

system for surface ships, started in December 1969. The entire system consists of 6 sub-systems. Over a range of 400 km, the Aegis system has the ability to monitor up to 400 targets.

ADS (Advanced Deployable System)

This is an ocean surveillance system that can be rapidly deployed in the ocean (usually shallow water) in which we are interested. The system is used to monitor and detect underwater targets, particularly submarines.

Dual Use Network for Military and Civilian Use[155-157]

(1) AERONET (Aerosol Robotic Network)
(2) BATS (Bermuda Atlantic Time Series)
(3) DEOS (Dynamics of Earth and Ocean Systems)
(4) H2O (Hawaii-2 Observatory)
(5) NEPTUNE: A project conducted by the USA and Canada in the Pacific. The network covers an area of 500×100 km. A 3,000 km electrical/optical cable is used to connect 30 – 50 nodes.
(6) SeaWeb FRONT (Front-Resolving Observational Network with Telemetry)
(7) MARS (Monterey Accelerated Research System)
(8) Venus (Victoria Experimental Network Under the Sea)

A-RCI (the Acoustic Rapid cOTS Insertion) Program[158]

This is a program for enhancing the performance of the existing sonar system of US submarines. COTS technology is used to reconstruct the submarine system and to considerably increase the system performance. It is reported that this program is not requirement-based but is a capability-based plan. The results of capability enhancing are listed in Table 8.3.

Table 8.3 The effect of A-RCI program

Item	Before reform	ARCI
Average detection probability	23%	87%
Average false alarming number per run	1	0.58
Average target recognition time	Reference 0	27 minutes more
Average continuously tracking time	Reference 0	25 minutes more

T-AGOS Surveillance Vessel[159]

A T-AGOS surface ship is equipped with the advanced SURTASS (Surveillance Towed Array Sonar System) and some classes of T-AGOS vessel have LFA (low frequency array) which can be used to detect quiet submarines. It is reported that

the twin line system is the best operational shallow water towed array and the only multi-line towed array in the US navy. It can be towed in water as shallow as 180 feet, providing significant directional noise rejection, and it resolves bearing ambiguities without turning.

For the submarine, NCW is a two-edged sword. On the one hand, the submarine obviously benefits from the network infrastructure. On the other hand, in order to perform NCW, some new requirements may result in difficulties for the submarine.

From theory to practice, it is necessary to establish a basic national infrastructure. The establishment of an infrastructure is subject to the constraints of the following three basic laws in the information world[160]. These are:

(1) Moore's law—2×every 18 months. In the micro-chip industry, the number of integrated circuits will double every 18 months and the price will be reduced by a half.
(2) Transmission capacity—2×every 12 months. The transmission capability of an optical fiber cable will double every 12 months; this is based on DWDM (dense wave division multiplexing) technology.
(3) Metcalfe's law—2×every 7.5 months. The quantity of data transmitted on the Web will double every 7.5 months.

8.6.4 Applications of the NCW Concept in Underwater Acoustics

The concept of NCW was first proposed by the US navy command. Although it is a concept about warfare in the information age, it is particularly concerned with underwater warfare. In order to achieve superior ability in information warfare, it is necessary to introduce the theory and techniques of communications and networks in the sonar field. Therefore, it will obviously result in a revolution in sonar design theory and applications.

The establishment of an underwater network is related to many different theoretical and applicable areas, including acoustic, wireless radio, optic and computers, etc. At the same time, it is necessary to develop new equipment which meets the requirements of NCW. For example, a new type of sonobuoy, underwater robot and underwater glider, etc. In the following, we will briefly discuss some technical issues concerning NCW in sonar applications.

Transmission Format of Information[161]

There are several methods that can be used to transmit an underwater signal by wire. The most popular data transmission format is ATM (asynchronous transfer mode). This is an international standard transmission. Each data package consists of a 53 B cell, including 48 B data and 5 B head code. The maximum transmission rate is 622 Mbps.

Ethernet is a data transmission method used in local networks. The bandwidth is about 100 Mbps. It can be used in point-to-point connection or star type, local bus connection. The transmission range is limited to 200 m.

The optical fiber is a very good transmission medium and the transmission rate can be up to 1.062 Gbps. When the transmission range is less than 80 km, it is not necessary to use a repeater.

Underwater Monitoring Network[53,127]

An underwater monitoring network is a dual use system for military and civilian applications. The main mission of this system is to monitor the ocean environment and underwater military activities. The most famous underwater monitoring system is SOSUS (sound surveillance system). The history of SOSUS development can be traced back to the 1950s. In 1950, the Hartwell project started, which included the establishment of a prototype of SOSUS. In 1951, the first system with 6 elements was installed in Eleuthera. In 1952, the formal name of SOSUS was coined and it is called the Caesar project in open literature. Up until 1954, 10 Caesar stations had been established, including 3 stations in the Atlantic Ocean, 6 stations in the Pacific Ocean and 1 station on Hawaii. In 1960, the new type of SOSUS, called the King shallow water system was installed and there are 10 stations with 8 elements.

It was reported, on June 26 and July 6 of 1962, that SOSUS detected a USSR diesel/electric submarine and a nuclear submarine.

After the cold war, in 1991, SOSUS was declassified and become an important facility for civil oceanographic studies.

Fig. 8.56 shows the relevant technology of NCW sonar design.

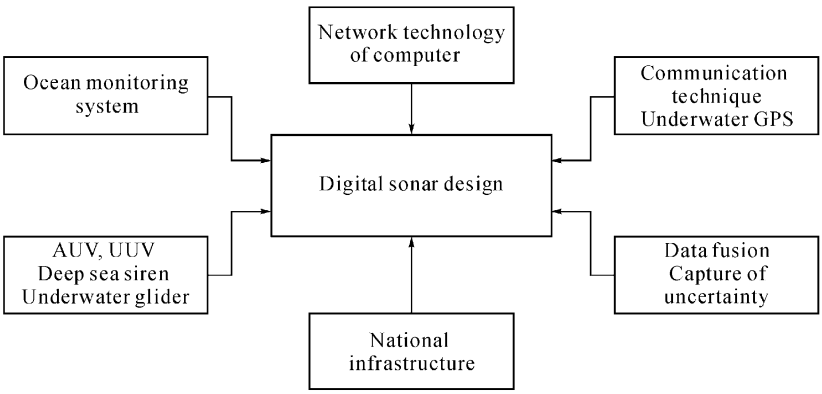


Fig. 8.56. Digital sonar design in the NCW era

In Fig. 8.56, under the conditions of NCW, sonar designers have to absorb much new knowledge from modern information theory, communication technology and computer networks. Of course, it is very important to merge sonar design

theory with NCW strategy. NCW is fast evolving in theory and practice and sonar designers have to meet the fast growing changes in underwater warfare.

8.7 A New Topic in Sonar Design: Diver Detection Sonar

Before the events of September 11, 2001, a shock terrorist attack in New York City, the destroyer Cole of the US navy was attacked by a terrorist boat loaded with a bomb in the Gulf of Aden, on October 12, 2000. The attack resulted in 17 deaths and more than 30 wounded. Thereafter, the anti-terrorism fight became a new challenge in warfare operations. The protection of ports, naval bases and any important sites close to the sea are the concern of both military and civil authorities.

The only effective method for diver detection is a sonar system, since there are no other means that can be used to detect a diver at long distance. DDS (diver detection sonar) is a new member of the sonar family. It usually is active/passive sonar since, in a port or naval base, it is not necessary to expose a sonar system to active operation. But passive-only DDS is also an option for developing a diver detection system[48,70,153-154]. Generally speaking, the diver detection problem can be included in the wider issue of waterside security. The first international conference on waterside security was held in Stockholm[161].

In this section, we will introduce some basic principles of DDS design. Because the object that DDS wants to detect is much smaller than a submarine, the target strength appearing in the sonar equation is very small. This is the main difficulty for DDS design.

8.7.1 Basic Design Principle of DDS

The main underwater attack activities at a naval base, port, VIP sea beach, oil/gas exploration platform, are performed by swimmers. Since terrorist attacks above water are relatively easy to detect by radar and other means, the main concern of waterside security is to detect swimmers. Beside DDS there are also other choices for port protection. For example, if the port facilities are located closely together and limited to a relatively small area and the exit of the port is narrow, then a mechanical fence can be considered as a means of protection. But the use of an underwater fence is subject to the geography of the sea bottom. The opening and closing of an underwater fence is also a difficult problem; a lot of underwater engineering issues have to be solved. Of course, the underwater fence itself needs to be protected, therefore DDS is actually the unique equipment for effectively detecting swimmers.

DDS can be considered as a part of the waterside security problem[161]. Fig. 8.57 illustrates a diver intruding, and Fig. 8.58 shows the main concept of the waterside security problem.

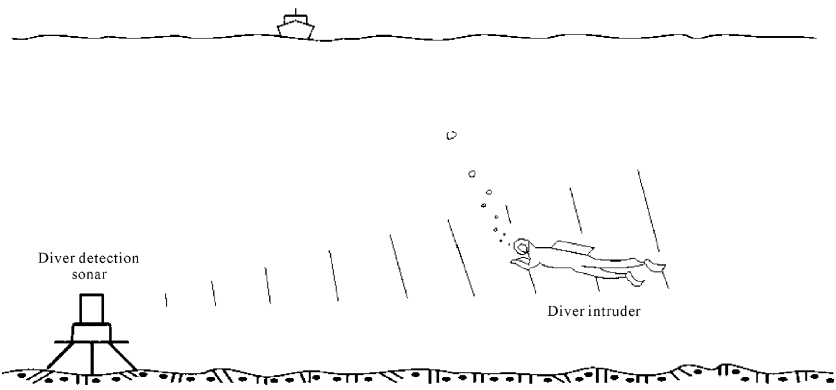


Fig. 8.57. Waterside security: diver intrusion detection

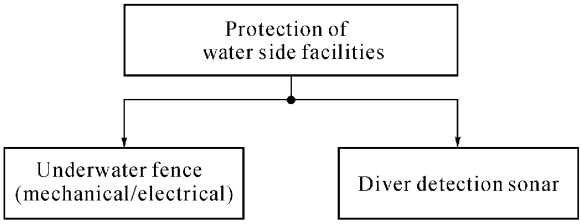


Fig. 8.58. Concept of waterside sonar

There are two kinds of breathing systems for a swimmer: a closed breathing system and an open breathing system. The radiated noise of a closed breathing system is much smaller than that of an open breathing system. Since the swimmer with an open breathing system will exhaust carbon dioxide, therefore this results in noise. The radiated noise of a breathing system is the target of passive sonar detection. But regardless of any kind of breathing system, the swimmer's body is the target of active sonar detection. Since the size of a swimmer is small, therefore the value of the target strength is usually smaller than – 10 dB. It is necessary to have large source level of transmitted signal for DDS.

8.7.2 Design Outline of DDS

Operating Frequency

The detection object of DDS is the swimmer; the body size of an adult can be simplified as 170×20 cm cylindrical. The resolution of DDS should be at least

1/10 of body size, i.e., 2 cm. Therefore the signal frequency of the transmitter should be $f = c / \lambda =$ 75 kHz. Generally speaking, the operating frequency of DDS should be greater than 50 kHz.

Source Level of Transmitted Signal

The requirement of the source level of the transmitted signal depends on the detection range of the active sonar. Suppose the detection range is $R \geq 1000$ m. Based on the spherical attenuation, the transmission loss is $2TL = 120$ dB. The absorption of the signal at high frequency cannot be neglected; we can suppose the absorption value to be 10 dB. Therefore $2FOM = 2TL + 10$ dB. The sonar equation for DDS becomes

$$FOM = (SL + TS - RL + GS + GT - DT) / 2 \tag{8.60}$$

and

$$SL = 2 \times FOM - (TS - RL + GS + GT - DT) \tag{8.61}$$

Take the following value: $2 \times FOM$ = 130 dB, TS = –10 dB, RL = 80 dB, GS = 15 dB, and GT = 5 dB, DT = 5 dB, we have SL = 219 dB.

Beamwidth of Receiver

The beamwidth of the receiver depends on the array aperture. As we have shown, the relationship of broadside beamwidth $\Delta\theta$ and array aperture L satisfy

$$\Delta\theta \approx 0.44\lambda / L \tag{8.62}$$

For example, if $\lambda = 2$ cm, $L = 80$ cm , then $\Delta\theta = 0.011$ arc ≈ 0.63 °.

When the target is biased at an angle of φ from the broadside direction, the value of the beamwidth should decrease by a factor of $1/\cos\varphi$. Therefore, the beamwidth of the receiver for DDS is usually in the order of 1°.

Recently, several diver detection sonars have been developed and introduced on the market, including Sentinel IDS (intrude detection sonar), Kongsberg DDS, QinetiQ Cerberus SDS (swimmer detection sonar) and DDSI Corp. DDS[153]. The main specifications of these sonar systems are similar, but the configuration of the wet end and the signal processing methods have some different features.

Fig. 8.59 is the block diagram of typical DDS. The main specifications of DDS are listed in Table 8.4[153].

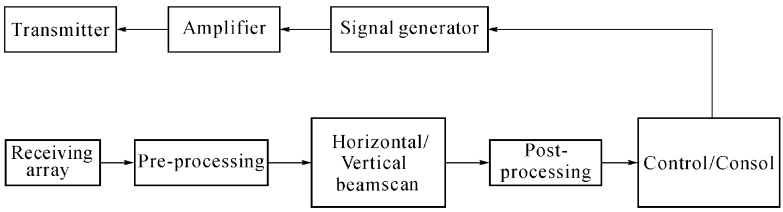


Fig. 8.59. Block diagram of diver detection sonar

Table 8.4 The main specification of Diver Detection Sonar (Israel, DDSI corp.)

Item	Specification
Transmit array	2×64elements
Length	80 cm
Beamwidth	1.6° – 10°
Beam tilt	20°
Receive array	2×96elements
Warning range	360°
Beamwidth	1.06°
Frequency	60 kHz
Wave form	CW, LFM
Source level	226 dB
Range	700 m
Target of tracking	20

References

[1] Robinson, A., Lermusiaux, P.: Prediction systems with data association for coupled ocean science and ocean acoustics. In: Proc. of ICTCA, Honolulu, www.deas.harad.edu (2003)

[2] Haffa, R. P., Patton, J. H. Jr.: Analogues of stealth. Comparative Strategy, 10-3, pp. 257-271 (1991)

[3] Stephanic, T.: Strategic Anti-Submarine Warfare and Naval Strategy. Lexington, New York (1987)

[4] Miasnikov, E.: Can a Russian strategic submarine survive at sea? The fundamental limits of passive acoustics. Science & Global Security, 4, pp. 213-251 (1994)

[5] Lemon, S. G.: Towed array history, 1917-2003. IEEE J. of Oceanic Eng., 29-2, pp. 365-373 (2004)

[6] Zachow, H.: System for noise measurements of towed arrays. In: Proc. of UDT'1998, pp. 262-267 (1998)

[7] Knight, A.: Flow noise calculations for extended hydrophones in fluid and solid-filled towed arrays. J. Acoust. Soc. Amer., 100-1, pp. 245 (1996)

[8] Francis, S. H., Slazak, M., Berryman, J. G.: Response of elastic cylinders to convective flow noise in homogeneous, layered cylinders. J. Acoust. Soc.

Amer., 75-1, pp. 166 (1984)
[9] Lasky, M., Doolittle, R. D., Simmons, B. D., Lemon, S. G.: Recent progress in towed hydrophone array research. IEEE J. of Oceanic Eng., 29-2, pp. 374-387 (2004)
[10] Technical Report of Defense R&D Canada: Towed integrated active sonar, www.drac-rddc.ge.ca
[11] Zeskind, R. M., Feuilet, J. P., Allensworth, W. S.: Acoustic performance of a multi-line system towed in several ocean environments. IEEE J. of Oceanic Eng., 23-1, pp. 124-128 (1998)
[12] Feuillet, J. P.: Nonambiguous beamforming for a high resolution twin-array. JASA, 97-5, pp. 3292-3294 (1995)
[13] Kaouri, K.: Left right ambiguity resolution of a towed array sonar, Somerville College. Univ. of Oxford, pp. 296-313, www.eprints.maths.ox.ac.uk
[14] Pasalic, N.: Towed array handling system on Collins-class submarine. Sea Technology, 42-11, pp. 51-56 (2001)
[15] Potter, J. R.: The "thin array": a light weight ultra-thin (8mm OD) towed array for use from small vessels of opportunity, http://www.arl.nus.edu.sg. Accessed 1 Jan 2011
[16] Allen, H., Cotterrill, P.: The manufacture and trials of an experimental vector sensor towed array to give good L/R ambiguity performance in a 65 mm diameter array. In: Proc. Of UDT'2006 Pacific, San Diego (2006)
[17] Bucker, H.: Least-squares target detection for a twin-line towed array. JASA, 95-3, pp. 1669-1670 (1994)
[18] SURTASS LFA: http://www.surtasslfa.com. Accessed 1 Jan 2011
[19] Gray, D. A., Riley, J. L.: Towed array shape estimation using Kalman filters:Theoretical models. IEEE J. Oceanic Eng., 18-4, pp. 543-556 (1993)
[20] Paidoussis, M. P.: Dynamics of flexible slender cylinders in axial flow. J. Fluid Mech., 26, pp. 717-736 (1966)
[21] Owsley, N. L.: Shape estimation for a flexible underwater cable. IEEE EASCON, Washington (1981)
[22] Varadarajan, V., Krolik, J.: Sensor array shape estimation using active sonar clutter. In: Proc 2nd IEEE Sensor Array and Multichannel Signal Processing Workshop, Rosslyn (2002)
[23] Hover, F. S., Grosenbaugh, M. A., Triantafyllou, M. S.: Calculation of dynamic motion and tensions in towed underwater cables. IEEE J. Oceanic Eng., 19-3, pp. 425-437 (1994)
[24] Smith, J. J., Leung, Y. H., Cantoni, A.: The Cramer-Rao lower bound for towed array shape estimation with a single source. IEEE Trans., SP-44-4, pp. 1033-1035 (1996)
[25] Karthikeyan, C.: Directional response of a towed array in shallow water sea. In Proc. IEEE, 133-F, pp. 138-145 (1986)
[26] Denyki, B., Griffin, B.: Smaller winch technology. Sea Technology, pp. 34-40 (1997)
[27] Preston, J. M.: Stability of towfish used as sonar platforms. In: Proc. of IEEE Ocean'92, pp. 888-893 (1992)

[28] Rickman, B.: Towed array hydrodynamics and dynamic beamforming. In: Proc. UDT'90, pp. 237-242 (1990)
[29] Milinazzo, F., Wilkie, M., Latchman, S. A.: An efficient algorithm for simulation the dynamics of towed cable system. J. of Oceanic Eng., 14-6, pp. 513-526 (1987)
[30] Ablow, C. M., Schechter, S.: Numerical simulation of undersea cable dynamics. J. of Oceanic Eng., 10-6, pp. 443-457 (1983)
[31] Lake, W.: Mechanics of Flow Induced Sound and Vibration. Academic Press, Boston (1986)
[32] Gordon, D. R.: Winch Technology, 39-7, pp. 17-24 (1998)
[33] Nikitakos, N. V., Leros, A. K., Katsikas, S. K.: Towed array shape estimation using multimodel partitioning filters. IEEE J. of Oceanic Eng., 23-4, pp. 380-384 (1998)
[34] Gloza, I.: Tracking the underwater noise source using vector sound: intensity probe. In: Proc. of ECUA'2002, pp. 247-250 (2002)
[35] Marra, W.: Applying technology solution to challenges of coastal security. Sea Technology, 46-3, pp. 16-23 (2005)
[36] Pinto, M. A., Bellettini, A., Hollett, R., Tesei, A.: Real and synthetic array signal processing of buried targets. IEEE J. of Oceanic Eng., 27-3, pp. 484-494 (2002)
[37] National Research Council: Illuminating the Hidden Planet: the Future of Sea Floor Observatory Science. National Academy Press, Washington (2000)
[38] Austin, T., Edson, J., McGillis, W., Chris, A., Purcell, M., Petitt, R., McElroy, M., Ware, J., Stokey, R.: The Martha's Vineyard coast observatory : a long term facility for monitoring air-sea processes. In: Proc. Oceans 2000, Providence, pp. 1937-1941 (2000)
[39] Detrick, R., Frye, D., Collins, J., Gobat, J., Grosenbaugh, M., Petitt, R., Plueddeman, A., Heydt, K., Wooding, F. B.: DEOS Moored Buoy Observatory Design Study. WHOI, Technical Report (2000)
[40] Butler, R., Lay, T., Creager, K., Earl, P., Fischer, K., Gaherty, J., Laske, G., Leith, B., Park, J., Ritzwoller, M., Tromp, J., Wen, L.:Hawaii-2 observatory pioneers opportunities for remote instrumentation in ocean studies, EOS, Trans. American Geophysical Union, 81-15, pp. 162-163 (2000)
[41] Ballard, R. D., Yoerger, D. R., Stewart, W. K., Bowen, A.: ARGO/JASON: a remotely operated survey and sampling system for full-ocean depth. In: Proc. Oceans 91, 1, pp. 71-75 (1991)
[42] Lemon, S. G.: Towed array history, 1917-2003. IEEE J. of Oceanic Eng., 29-2, pp. 365-373 (2004)
[43] Sonnichsen, F.: Communication options for underwater scientific networks. Sea Technology, 46-5, pp. 46-50 (2005)
[44] El-Sharkaw, M. A., Upadhye, A., Lu, S., Kirakham, H., Howe, B. M., McGinnis, T., Lancaster, P.: North east Pacific time integrated undersea networked experiments (NEPTUNE): cable switching and protection. IEEE J. of Oceanic Eng., 30-3, pp. 232-240 (2005)
[45] Whitman, E. C.: Submarine in network centric warfare. Sea Power Archives,

www.navyleague.org (2004)
[46] Sullivan, E. J.: Passive acoustic synthetic aperture processing. IEEE J. of Oceanic Eng. Society Newsletter, 38-1, pp. 21-24 (2003)
[47] Chatham, R. et al.: The synthetic aperture sonar revolution. In: Proc. of AUSI Workshop, Kona Hawaii (2000)
[48] Kessel, R. T., Hollett, R. D.: Underwater intruder detection sonar for harbour protection: state of the art review and implications. In: Proc. of 2nd IEEE Conference on Homeland Security and Safety, Istanbul (2006)
[49] Tanaka, T., Yamaguchi, I.: Low frequency synthetic aperture sonar. NEC Res.&Develop, 44-2, pp. 161-164 (2003)
[50] Whitman, E. C.: Submarine in Network Centric Warfare, http://www.navyleague.org (2004)
[51] Cebrowski, K., Garstka, J. J.: Network centric warfare: its origin and future. In: Naval Institute Proceedings (1998)
[52] Stefanick, T.: Strategic Antisubmarine Warfare and Naval Strategy. Lexington, New York (1987)
[53] SOSUS: http://www.fas.org/irp/program/collect/sosus.htm. Accessed 1 Jan 2011
[54] Dunn, P.: Navy unmanned undersea vehicle (UUV) master plan. In: Unmanned underwater Vehicle Showcase 2000 Conference Proceedings, pp. 105-126 (2004)
[55] NEPTUNE: www.neptunecanada.ca
[56] VENUS: www.venus.uvic.ca
[57] Fernandes, P. G., Brierley, A. S.: Using an autonomous underwater vehicle as a platform for mesoscale acoustic sampling in the marine environment. ICES CM, M:01-16 (1999)
[58] Curtin, T., Bellingham, J. G., Catipovic, J., Webb, D.: Autonomous ocean sampling networks. Oceanography,6-3, pp. 86-94 (1993)
[59] Leonard, N. E., Fiorelli, E., Bhatta, P., Leonard, N. E., Shulman, I.: Adaptive sampling using feedback control of an autonomous underwater glider fleet. In: 13th International Symposium on Unmanned Untethered Submersible Technology, Durham (2003)
[60] Eriksen C. C., Osse, T. J., Light, R. D., Wen, T., Lehman, T. W., Sabin, P. L., Ballard, J. W., Chiodi, A. M.: Seaglider: A long-range autonomous underwater vehicle for oceanographic research. IEEE J. Oceanic Eng., 26(4):424-436, October 2001.
[61] Graver, J. G., Bachmayer R., Leonard, N. E., Fratantoni, D. M.: Underwater glider model parameter identification, In 13th International Symposium on Unmanned Untethered Submersible Technology. Durham, NH, USA, August 2003.
[62] Graver, J. G. and Leonard, N. E.: Underwater glider dynamics and control" In 12th International Symposium on Unmanned Untethered Submersible Technology. Durham, NH, USA, August (2001)
[63] Leonard, N. E. and Graver, J. G.: Model-based feedback control of autonomous underwater gliders. IEEE Journal of Oceanic Eng., 26-4, pp.

633-645 (2001)
[64] Shermanet, J., Davis, R. E., Owens, W. B., Valdes, J.: The autonomous underwater glider. IEEE J. of Oceanic Eng., 26-4, pp. 437-446 (2001)
[65] Ulanov, A. V.: Optimization criteria and hierarchy of mathematical models of an underwater gliding vehicle. In: Third International Shipbuilding Conference ISC'2002, St. Petersburg (2002)
[66] Webb, D. C.: An underwater glider propelled by environmental energy. IEEE J. of Oceanic Eng., 26-4, pp. 447-452 (2001)
[67] Kong, J., Cui, J., Wu, D., Gerla, M.: Building underwater ad-hoc Networks and sensor networks for large scale real-time aquatic applications. In: Proceedings of MILCOM (2005)
[68] Hétet, A., Guyonic, S.: Experiments into detection and classification of buried mines: towards concept achievement. In: Proc. UDT Europe 2004, Nice (2004)
[69] Guyonic, S.: A technique for buried mines detection and classification—the application of a data acquisition tool and its subsequent process. Sea Technology, 44-6 (2003)
[70] Stanic, S., Kirkendall, C. K., Tveten, A. B., Barock, T.: Passive Swimmer Detection. NRL Review, 2004.
[71] Green, L. H., Raff, B. E.: Net-centric undersea warfare. Sea Technology, 40-11, pp. 19-26 (1999)
[72] Burns, R. F.: The naval research laboratory. Sea Technology, 34-11, pp. 66-77 (1993)
[73] Bealtie, G. A., Cotterill, P. A.: Assessing performance: Submarine flank arrays. Sea Technology, 36-11, pp. 45-49 (1995)
[74] Feintach, P. L.: Adaptive beamforming to suppress hull noise using PVDF array panels. In: Proc. of UDT'94, London, pp. 235 (1994)
[75] Stotts, S. A., Martin, J. L., Bedford, N. R.: Multiple source localization using GPS technology and received arrival time structure analysis in an air-deployed system. IEEE J. of Oceanic Eng., 22-3, pp. 576-582 (1997)
[76] Adams, A. E., Lawlor, M. A., Riyait, V. S., Hinton, O. R., Sharif, B. S.: Real time synthetic aperture sonar processing system. In: Proc. IEEE, 143-F, pp. 169-176 (1996)
[77] Stergiopoulos, S., Urban, H.: A new passive synthetic aperture technique for towed arrays. IEEE J. of Oceanic Eng., 17-1, pp. 16-25 (1992)
[78] Bruce, M.: A processing requirement and resolution capability comparison of side scan and synthetic aperture sonars. IEEE J. of Oceanic Eng., 17-1, pp. 106-117 (1992)
[79] Stergiopoulos, S., Urban, H.: An experimental study in forming a long synthetic aperture at sea. IEEE J. of Oceanic Eng., 17-1, pp. 62-67 (1992)
[80] Hares, M. P., Gough, P. T.: Broad band synthetic aperture sonar. IEEE J. of Oceanic Eng., 17-1, pp. 80-94 (1992)
[81] Stergiopoulos, S.: Implementation of adaptive and synthetic aperture processing schemes in integrated active and passive sonar system. In: Proc. IEEE, 86-2, pp. 358-398 (1998)

[82] Loggins, C. D., Christoff, J. T., Pipkin, E. L.: Results from rail synthetic aperture experiments. J. Acoust.Soc. Amer., 71, pp. 85-90 (1982)
[83] Garrood, D., Lehtomaki, N., Luk, T., Neudorfer, M., Palowitch, A.: Synthetic aperture sonar: an evolving technology. Sea Technology, 40-1, pp. 12-23 (1999)
[84] Hazell, P. A.: What's the future for ASW in NATO? Sea Technology, 39-11, pp. 10-17 (1998)
[85] Moffat, J.: Complexity Theory and Network Centric Warfare. CCRP, Washington (2003)
[86] Albert, D. S., Gartska, J. J., Stein, F. P.: Network Centric Warfare: Developing and Leveraging Information Superiority, 2nd Edn. CCRP, Washington (2002)
[87] Pinto, M. A., Bellettini, A., Hollett, R., Tesei, A.: Real and synthetic array signal processing of buried targets. IEEE J. of Oceanic Eng., 27-3, pp. 484-494 (2002)
[88] Kennedy, F. D. Jr.: Experimentation: the key to transformation. Undersea Warfare, 5-1, pp. 3-10 (2002)
[89] Walrod, J.: Sensor and actuator networks for acoustic signature monitoring and control. In: Proc. of UDT'1999, Nice (1999)
[90] Walrod, J.: Sensor networks for network centric warfare. In: Proc. of NCW Conference, Fall Church (2000)
[91] Applewhite, A., Kumagai, J.: Technology trends 2004. IEEE Spectrum, pp. 8-13 (2004)
[92] Jonson, Adm Jay L.: Address at US naval institute Annapolis seminar and 123rd annual meeting. Annapolis (1997)
[93] Bonito, G.: Sensor array projects and networks & other useful links. www.lnternet.edu (2003)
[94] Kaouri, K.: Left right ambiguity resolution of a towed array sonar, Somerville College, Univ. of Oxford, pp. 296-313, www.eprints.maths.ox.ac.uk
[95] ATM: http://www.atmforum.com. Accessed 1 Jan 2011
[96] How Loud is the Navy's SURTASS/LFA Sonar System ? http://www.earthislan.org. Accessed 1 Jan 2011
[97] Pruitt, T.: Maritime homeland security for ports and commercial operation. Sea Technology, 45-11, pp. 20-26 (2004)
[98] Sternlicht, D., Pesaturo, J. F.: Synthetic aperture sonar: frontiers in underwater imaging. Sea Technology, 45-11, pp. 27-34 (2004)
[99] Spiess, F., Kupperman, W. A.: The marine physical laboratory in Scripps. Oceanography, 16-3, pp. 45-54 (2003)
[100] Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T., Ferla, C.: Phase conjugation in the ocean: experimental demonstration of a time reversal mirror. JASA, 105, pp. 1597-1606 (1999)
[101] Sullivan, E. J.: Passive acoustic synthetic aperture processing. IEEE J. of Oceanic Eng. Society Newsletter, 38-1, pp. 21-24 (2003)
[102] Medwin, H.: Sound in the Sea, Cambridge Univ. Press, New York (2005)

[103] P.Cotterill, P. et al.: Processing considerations for vector sensor towed array: VESTA. In: Proc. of UDT' 2006 Pacific, San Diego (2006)
[104] Akyildiz, I., Pompili, D., Melodia, T.: Underwater acoustic sensor networks: research challenges. Ad Hoc networks 3, pp. 257-279, 257-279, http://www.seciencedirect.com (2005)
[105] Chatham, R. et al.: The synthetic aperture sonar revolution. In: Proc. Of AUSI Workshop, Kona, Hawaii (2000)
[106] Wocester, P. F., Spindel, R. C.: North Pacific Acoustic Laboratory, http://www.npal.ucsd.edu. Accessed 1 Jan 2011
[107] Anyang, S. Y.: Fundamentals of Complex-System Theories: in Economics, Evolutionary Biology and Statistical Physics. Cambridge Univ. Press, New York (1998)
[108] Cebrowski, K., Garstka, J. J.: Network Centric Warfare: its Origin and Future. Naval Institute Proceedings, JCS (1998)
[109] Fioravanti, A. et al.: A parametric aperture sonar. In: Proc. of 3rd ECUA, Heraklion, pp. 1085-1090 (1996)
[110] Camdy, J. V.: Model based signal processing in the ocean. IEEE Oceanic Eng. Society News Letter, 35-3, pp. 199-205 (2000)
[111] Sullivam, E. J., Middleton, D.: Estimation and detection issues in matched field processing. IEEE J. of Oceanic Eng., 18-3, pp. 156-167 (1993)
[112] Baggerroer, A. B., Kuperman, W.: An overview of matched field method in ocean acoustics. IEEE J. of Oceanic Eng., 18, pp. 379-387 (1993)
[113] Munk, W., Worcester, P., Wunsch, C.: Ocean Acoustic Tomography. Cambridge Univ. Press, New York (1995)
[114] Cox, H., Zeskind, R., Owen, M.: Robust adaptive beamforming. IEEE Trans., ASSP-35, pp. 1365-1376 (1987)
[115] Hampel, F., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A.: Robust Statistics: the Approach Based on Influence Function. Wiley, New York (1986)
[116] Green, T. J. Jr.: Robust Passive Sonar, DARPA Tech 2000, Dallas, http://www.darpa.mil (2000)
[117] Seong, W., Byun, S. H.: Robust matched field processing algorithm based on feature extraction. IEEE J. of Oceanic Eng., 27-3, pp. 642-652 (2002)
[118] Pandharipande, A.: Information, uncertainty and randomness. IEEE Potentials, Oct./Nov., pp.32-34 (2002)
[119] Porter, M. B.: Acoustic models and sonar systems. IEEE J. of Oceanic Eng., 18, pp. 425-437 (1993)
[120] Candy, J. V.: Model-Based Digital Processing. Wiley, New York (2005)
[121] Tolstoy, I., Clay, C. S.: Ocean Acoustics. AIP Press, New York (1987)
[122] Tolstoy, A.: Matched Field Processing for Ocean Acoustics. World Scientific Pub., New Jersey (1993)
[123] Sullivan, E. J., Edelson, G. S.: Model-Based Broadband Towed Array Processing. 147th meeting of the Acoustical Society of America, New York, (2004)
[124] Sullivan, E. J., Candy, J. V.: Space-time array processing: the model-based

approach. J. Acoust. Soc. Amer., 102-1, pp. 2809-2820 (1997)
[125] Candy, J. V., Sullivan, E. J.: Passive localization in ocean acoustics: a model based approach. J. Acoust. Soc. Amer., 98-3, pp. 1455-1471 (1995)
[126] SOSUS: http://www.globalsecurity.org. Accessed 1 Jan 2011
[127] Patria Aviation Oy, PFA:http://www.patria.fi. Accessed 1 Jan 2011
[128] Thales Underwater Systems: http://www.tahlesgroupe.com/naval. Accessed 1 Jan 2011
[129] Whitehouse, B. G., Hines, P., Ellis, D., Barron, C. N.: Rapid environmental assessment within NATO. Sea Technology, 45-11, pp. 10-14 (2004)
[130] ACTAS: http://www.atlas-elektronik.com. Accessed 1 Jan 2011
[131] Richman, B.: Towed array hydrodynamics and dynamic beamforming. In: Proc. of UDT'1990, London, pp. 243-248 (1990)
[132] Doisy,Y.: Port-starboard discrimination performance on activated towed arrays systems. In: Proc. of UDT'1995, Cannes, pp. 125-129 (1995)
[133] Letiche, M. et al.: TSM 2670 active low frequency sonar - the design of the transmitter system. In: Proc. of UDT'1995, Cannes, pp. 302-306 (1995)
[134] Van den Dool, T. C., Hopmans, L. J. M.: Flank array adaptive noise canceling. In: Proc. of UDT'1998, London, pp. 298-302 (1998)
[135] RAFAEL: http://www.rafael.co.il. Accessed 1 Jan 2011
[136] FAS 3-1 Flank Array Sonar: http://www.stn-atlas.de. Accessed 1 Jan 2011
[137] Godin, O. A., Palmer, D. R.: History of Russian Underwater Acoustics. World Scientific Pub., New Jersey (2008)
[138] Gerken, L.: ASW Versus Submarine Technology Battle. American Scientific Corp., California (1986)
[139] Bray, A. V.: Underwater hardware life testing. Sea Technology, 33-12, pp. 56-61 (1992)
[140] Marra, L. J.: Sharkbite of the submarine lightwave cable system: history, causes and resolution. IEEE J. of Oceanic Eng., 14-3, pp. 230-237 (1989)
[141] Blanhut, R. E., Miller, W., Wilcox, C. H.: Radar and Sonar. Springer, New York (1991)
[142] Kock, W. E.: Radar, Sonar and Holography. Academic Press, Boston (1973)
[143] Sherwin, C. W., Ruina, J. P., Rawcliffe, R. D.: Some early developments in synthetic aperture radar systems. IRE Trans. Military Electronic, 6, pp. 111-115 (1962)
[144] Tomiyasu, K.: Tutorial review of synthetic aperture radar (SAR) with applications to imaging the ocean surface. In: Proc. IEEE, 66, pp. 563-583 (1978)
[145] Brown, W. M., Porcello, L. J.: An introduction to synthetic aperture radar. IEEE Spectrum, pp. 52-62 (1969)
[146] Harger, R. O.: Synthetic Aperture Radar Systems. Academic Press, Boston (1970)
[147] Cutrona, L. J.: Synthetic aperture radar. In: Radar Handbook, M.I.Skolink (Eds.), McGraw-Hill, New York (1970)
[148] Loggins, C. et al.: Results from real synthetic aperture experiments. J. Acoust.Soc. Amer., 71-Suppl., pp. 85-90 (1982)

[149] Zhang, C. H, Tang, J. S. et al.: Final report of SAS prototype. Technical Report, Institute of Acoustics (2001)
[150] Li, Q. H. et al.: Advances of SAS study in China. In: Proc. of UAM'2005, Crete (2005)
[151] SAS: http://www.apl.washington.edu. Accessed 1 Jan 2011
[152] DDS: http://www.arstech.de; http://www.dsi.co.il, http://www.sonardyne.com; http://www/kongsberg-mesotech.com, http://www.QinetiQ.com. Accessed 1 Jan 2011
[153] Jiang, L. J. et al.: Advances of diver detection sonar study. Science in Press, 54-3, pp.269-272 (2009)
[154] Aeronet: http://www.gsfc.nasa.gov. Accessed 1 Jan 2011
[155] BATS: http://www.bbsr.edu. Accessed 1 Jan 2011
[156] Delaney, K. J.: Acoustic surveillance of sea minelaying. SPAWAR, San Diego
[157] Feely, M. et al.: ARCI heralds acoustic revolution. UDT Forum, 1, pp. 10-15 (2005)
[158] SURTASS: http://www.chinfo.navy.mil. Accessed 1 Jan 2011
[159] Roca, R.: AT&T Bell Labs presentation on information technology trends to GovTechNet'99. Washington (1999)
[160] WSS: http://www.wss2008.org. Accessed 1 Jan 2011
[161] Franceschetti, G., Lanari, R.: Synthetic Aperture Radar Processing. CRC Press, Boca Raton (1999)
[162] Robert, M. K., Beerens, S. P.: Adaptive beamforming algorithms for tow ship noise canceling. In: Proc. of UDT' 2002 (2002)

9

Software and Hardware Support and Performance Evaluation in Digital Sonar Design

As very complicated electronic equipment, a digital sonar system can be classified into two parts, the hardware and the software. The hardware consists of a large amount of electronic components, LSI, VLSI and DSP specific chips. It is the physical media for information acquisition, storage, transmission and processing. The software, running on the hardware, determines the information flow direction, combination format, logical relationship and data management.

Similar to a digital computer, the hardware and software are inseparable. In the entire design procedure, it is necessary to consider both hardware and software. In considering the hardware architecture, one should take care of the operating convenience of the software and leave the necessary redundancy for data processing. In the case that software modification is necessary, we have enough hardware support to meet the requirements. In considering the software design, one should pay attention to the system hardware structure, such that the data transmission and interchange between different hardware modules is easier.

It is difficult to make the hardware and software fit perfectly. The most important thing is to satisfactorily utilize the existing hardware and software resources, which is the basis of digital sonar design.

Principally speaking, the more we understand the hardware assembly language, the better the quality of software we can design. But sometimes it is hard to modify the assembly language for DSP chips. We prefer to use a high level language, which means we have to have more data memory or increase operating time. Therefore, a tradeoff is necessary in the design procedure.

In this chapter, several important support techniques for the hardware and software are introduced. Correctly utilization of these techniques will speed up the design procedure and reduce the cost.

We will also discuss a very important issue, which is often neglected by some designers. This is the problem of the laboratory testing of sonar systems and lake experiment and sea trials. The performance of sonar systems demonstrated in various situations varies, but can be compared in some ways.

9.1 Software Package Matlab in Digital Signal Processing

Matlab is a high performance interactive scientific and engineering computer language. It is an excellent software package for numerical calculation, matrix operation, signal processing and image processing. Because the application environment of Matlab is very simple, it is not necessary to have complicated compiling when using this software; therefore it is extensively used in many theoretical and applied engineering areas. Matlab has been used in digital sonar design for many years. It is the most important tool for system simulation in sonar design. Today, with the development of signal processing theory and applications, Matlab is still being improved and modified. The original material and basic description of Matlab software can be found in their website[1]. There are hundreds of monographs about Matlab and readers can benefit from the knowledge and programming skills in these books[2-3]. Some software based on Matlab have been developed in underwater acoustic field, which are extensively used in this area[4].

9.1.1 Several Frequently Used Commands in Signal Processing

In Matlab software, there are more than ten specific tool boxes for signal processing. They include software for image processing, spline interpolation, data assimilation, target recognition, etc. In order to compile these subroutines into a program, it is necessary to have specific control commands which can call the subroutine, change the flow direction in certain conditions, or jump the executing program to a new point. These control commands are similar to those in high level computer language.

(1) if: conditional execution;
(2) elseif: use with if;
(3) else: use with if;
(4) end: conclude if, for, while;
(5) for: repeat execution;
(6) while: cycle without number limitation;
(7) break: jump from for, while cycle;
(8) return: return from executing function;
(9) pause: temporary stop, wait for response.

The following commands are frequently used in digital signal processing:

(1) conv: convolution;
(2) cov: covariance;
(3) deconv: deconvolution;
(4) dft: discrete Fourier transform;
(5) fft: fast Fourier transform;
(6) fft2: two-dimensional Fourier transform;
(7) ifft: Inverse fast Fourier transform;
(8) butter: Butterworth IIR filter design;

(9) remez: Parks-McClellan FIR filter design;
(10) filter: filtering;
(11) freqz: frequency response of z transform;
(12) spectrum: spectrum analysis;
(13) fir1: FIR design.

9.1.2 Examples

Matlab software is extensively used in digital sonar design, including performance prediction, system simulation and underwater signal transmission loss, etc. The best way to learn Matlab is to compile programs in terms of the application environment. It is worth saying that "practice makes perfect". Here we are going to introduce some examples, which are very preliminary application programs in sonar design. The goal of introducing these examples is just to give readers an idea of how to use Matlab and show what Matlab can do.
(1) Samples of Random Number

```
% Samples of random number
aa = rand (1,1024);
bb = rand (1,1024);
k = 0 : 1000;
subplot (2,1,1);
plot (k,aa(k+1));
title ('Samples of random number: uniform distribution');
grid;
subplot (2,1,2);
plot (k,bb(k+1));
title ('Samples of random number: normal distribution');
grid;
```

Fig. 9.1 illustrates the results of this program.
(2) Fast Fourier Transform

```
% FFT, f_s = 5 kHz, SNR = -6 dB
% Frequency resolution = 5000/1024 = 4.8828 Hz
t = 0 : .0002 : .2048
pi = 3.1415926; snr = 4.887;
% signal frequency = 300 Hz
x = sin(2*pi*300*t) + snr*(rand(size(t)-0.5);
y = fft(x);
Pyy = y*conj(y);
f = 4.8828*(0:255);
plot (f,Pyy(1:256));
title ('Power spectrum of signal plus noise (SNR= -6dB)');
```

```
xlabel ('Frequency: Hz');
ylabel ('Spectrum level');
```

Fig. 9.2 shows the results of this program.

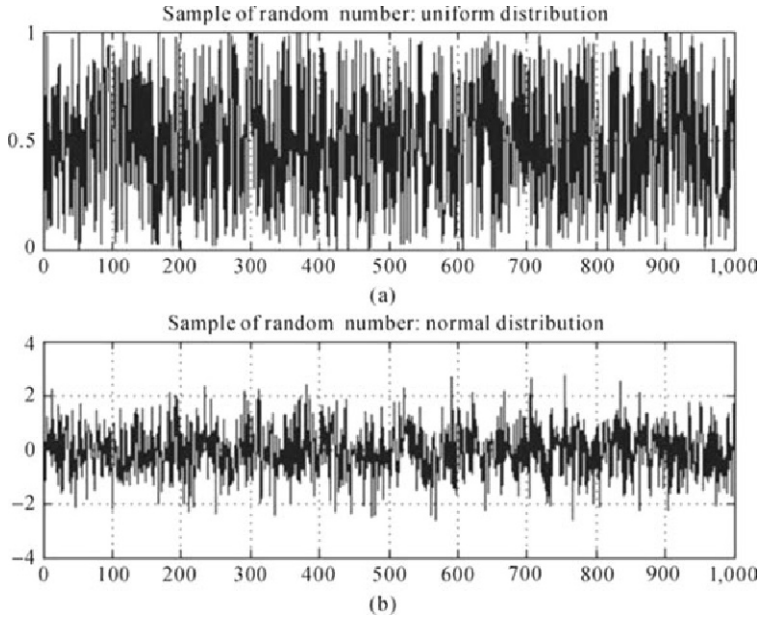


Fig. 9.1. Samples of random numbers generated by Matlab

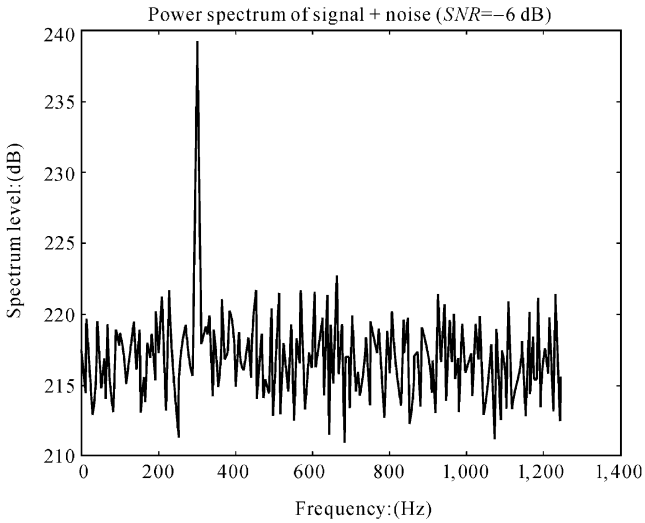


Fig. 9.2. Example of frequency analysis in Matlab

(3) FIR Filter Design

```
% FIR filter design
% 65 taps
% Sampling frequency = 40 kHz
% Band pass: 4-8 kHz
n = 65;
f = [0.20 0.40];
b = fir (nf);
pause;
[H,W] = freqz(b);
Plot (W*20/pi,abs(H)), grid;
xlabel ('Frequency (kHz)');
ylabel ('Magnitude (mV)');
title ('FIR filter design');
grid;
```

The frequency response designed by this program is shown in Fig.9.3.

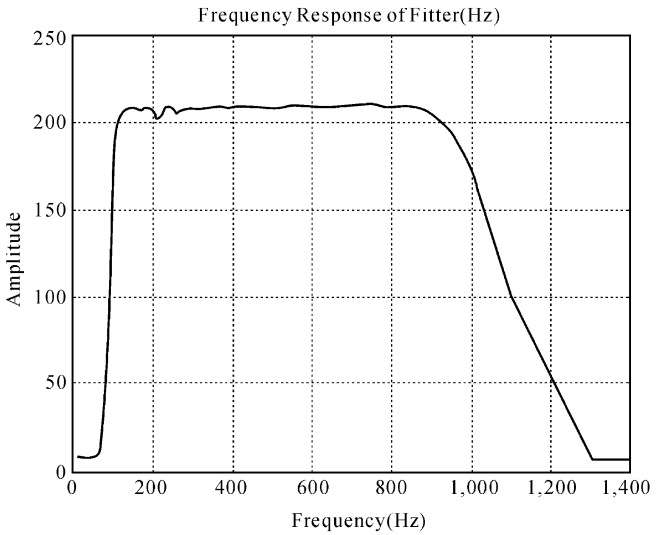


Fig. 9.3. Example of FIR filter design in Matlab

9.1.3 Other Features of Matlab

Tool Boxes

There are 24 specific signal processing tools and one simulation software. The

main tool boxes used in signal processing are:

(1) signal processing toolbox;
(2) system identification toolbox;
(3) optimization toolbox;
(4) neural network toolbox;
(5) control system toolbox;
(6) robust control toolbox;
(7) mu-analysis and synthesis toolbox;
(8) spline toolbox;
(9) image processing toolbox;
(10) fuzzy logic toolbox;
(11) wavelet toolbox.

The simulation software is called SIMULINK.

Self-Defined Function

Matlab allows the user to define his own function, if necessary. Once the function is defined, one can call it an ordinary function. For example, in calculating the beam pattern of a linear array, it is common to use function $\sin x/x$, but in Matlab there is no such function. We can define a function, called sinc_x, by the following statements.

```
% function s = sinc_x
% sinc_x computes the value of sin(x)/x
s = x;
set_1 = find (abs(x)<0.001);
set_2 = find (abs(x)>=0.001);
s(set_1) = ones (size(set_1));
s(set(2) = sin (x(set_2))/z(set_2);
```

To save this file as sinc$x.m$ when we use this function in Matlab, we can call it just call sqrt, cos etc.

```
% Illustration of sinc function
x = -15 : 0.05 : 15;
y = abs (sinc_x(x));
plot (x,y), title ('Sinc function');
xlabel ('X');, ylabel ('Y');
grid;
```

The plot of a sinc_x function is given in Fig. 9.4.

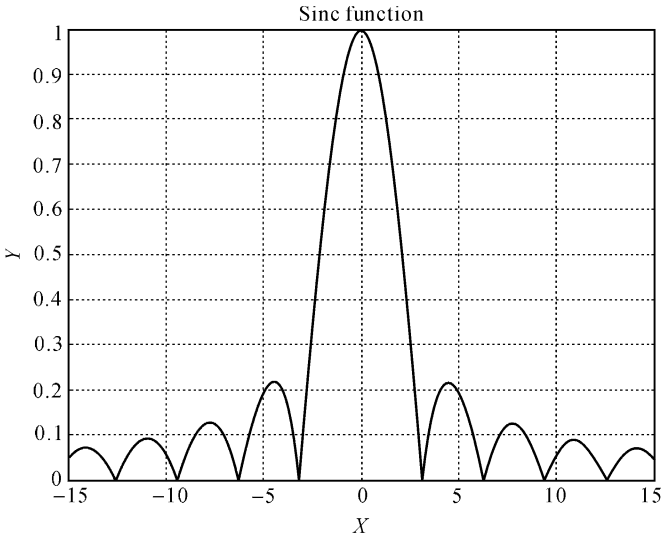


Fig. 9.4. Example of specific function sinc_x in Matlab

Sound Signal Processing Function

In Matlab, there is a specific command which transfers the vector data to the sound signal, i.e., the command of sound. It is also possible to transfer μ law sound code signal to a linear signal, and vice versa.

9.2 Programming Skills for C Language

The C language, including C++, Turbo C, Borland C, visual C in the C language family, is the most suitable programming language for the sonar designer[5,6], for the close relationship between the C language and hardware development. In the sonar design procedure, system simulation and performance testing, the C language plays a very important role. For the sonar designer, it is sufficient to know the high level language, but he should know a little about assembler code, which is necessary for hardware development. For a sonar designer, to be familiar with and be able to manage the C language is one of the most important tasks.

Some skills in writing and testing C language programs are shown below.

Make a Good Program Flow Chart

Don't try to write the program directly when we face a problem to be solved, unless the problem is very easy and simple. Otherwise, it is necessary to make a

good program flow chart. Based on the program flow chart, it is more efficient to write the program. In making a program flow chart, it is necessary to use standard notations or symbols (Fig. 9.5). The notations and symbols used in the flow chart include input, output, comparison and calculation, etc. In Fig. 9.5, the mission of this program is to calculate the area of a circle and print the result.

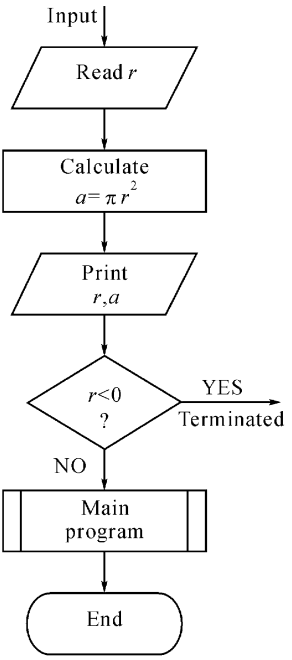


Fig. 9.5. Identification of flow chart for a program

Example

In this example, we want to simulate the ANC (adaptive noise canceling) algorithm, which is described in Chapter 3 (Eq. (3.124)). The key step in noise canceling is to perform the steepest descent algorithm

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + 2\mu\varepsilon(k)\boldsymbol{x}(k) \tag{9.11}$$

Fig. 9.6 shows the flow chart of the ANC program. In this program, it is necessary to input real data $\boldsymbol{x}(k)$ in each iteration step. The execution will finish when the convergence decision is confirmed (Fig. 9.6).

Decompose a Complicated Problem to Several Relative Simple Problems

If the problem to be solved is quite complicated, it is necessary to decompose it to several relative small and easy-to-solve problems. Then write a sub-program and link them together.

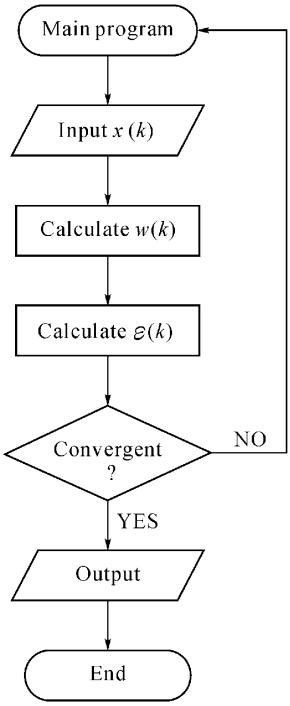


Fig. 9.6. Example of adaptive iteration program

The Program Testing Should Adopt a Reduced Dimension Method

In testing a program, a very efficient method is to reduce the dimension of the operation or variables. It means reducing the number of cycles or decreasing the dynamic range of data value. The goal of reduced dimension is to save time and check the correctness of the program, but without losing the program feasibility.

Testing of Compiled Program

Even for a program which has already passed program compiling, it is necessary to test this program by choosing special parameters and the program execution using these parameters will result in a known conclusion. In this case, the correctness of the program can be confirmed in the testing. At the same time, it is strongly recommended that one should insert enough interruption, which can provide an opportunity for man / machine interaction. Therefore, the problems produced in program execution can be corrected in time.

9.3 General Purpose DSP chips

The general purpose DSP chip is the hardware basis of a digital sonar system, a right choosing of DSP chips is substantially an important task of sonar designer[7-15]. Since the fast development of the semiconductor industry and the microcomputer, the performance of DSP chips has been changing quickly. Therefore, the hardware for digital sonar has also been changing quickly. DSP chips are extensively used in digital sonar, and the system architecture of digital sonar depends on the performance of DSP chips. There is no doubt that any modern digital sonar comprises an integration of various specific DSP chips and general purpose DSP chips. DSP chips are really the kernel of digital sonar. The design of a digital sonar system is based on DSP chips. A wide variety of special program and software have been developed for sonar signal processing[16-20], including model-based sonar signal processing, space-time array processing, detection/classification and auditory scene analysis[20-21].

In this section, the development trend of DSP chips is discussed. The applications of various DSP chips are introduced.

9.3.1 Concept of the General Purpose DSP Chip

The DSP chip is aspecific kind of microcomputer, which is suitable for digital signal operations. It is able to complete various kinds of digital signal processing in real time. DSP chips can be categorized into two classes. One is a general purpose DSP chip, and the other is a specific DSP chip. The general purpose DSP chip is used to complete general digital signal operations; it usually has a big space for program compiling and ease of address access. The specific DSP chip is designed for completing special digital algorithms, for example the FFT chip, FIR chip and so on. In this section when we talk about DSP chips, it means the general purpose DSP chip, unless specifically mentioned.

The main characteristics of DSP chips are as follows:

(1) The system architecture of a DSP chip is designed to consist of Harvard architecture, i.e., the program space and data space are separated. There is a high speed RAM in the chip and the RAM can be accessed by independent data bus.
(2) The program command sets the support flow line operation and in one instruction period it can complete one addition and multiplication.
(3) The fast interrupt processing and hardware I/O processing can be completed in real time; it is usually has the ability of a multiple address generator in a single period operation.

The DSP chips can be also categorized by the data format. One class is a fixed point DSP and one is a floating point DSP. The fixed point DSP chips are all the same format for different brands of DSP; but the different type of floating DSP

adopts quite different formats of floating digits in different products. Some floating point DSP chips adopt the IEEE standard floating point format, for example the Motorola series floating point DSP chips. But some floating point DSP chips adopt a self-defined floating point format, for example the TI TMS 320C3X series DSP chips.

9.3.2 Main Specifications of DSP Chips

For sonar designers, the selection of DSP chips is multiple choices. In the following, we will present some main specifications for DSP chips, which describe the most important features of DSP chips.

(1) Instruction period: the time for performing one instruction, usually in unit ns.
(2) The time for one multiplication and addition.
(3) MIPS (mega instruction per second), the time for performing one mega instruction per second.
(4) MOPS (mega operation per second), the time for performing one mega operation per second.
(5) MFLOPS (mega floating operation per second), the time for performing one mega floating point operation per second.
(6) Accuracy of operation: for the fixed point DSP, there are 16-, 24-, 32- and 64-bit word lengths; for the floating point DSP, it is usually a 32- bit word length.
(7) Power consumption.
(8) Reliability: for military products, the failure rate is less than 10^{-7} h. The temperature range is –50 °C – 125 °C.
(9) Development software: a well-established software package is usually compiled in C language and can be embedded in Matlab software. The development tools for DSP chips often include simulation software of signal processing, which usually can be run on a PC.

There are two kinds of DSP chips extensively used in digital sonar systems. One is the C60 series of DSP chips from the TI Company. The main specification is listed in Table 9.1.

Table 9.1 Specification of TMS C60 series chip of TI Company

Item	C62X 16 bit, fixed point	C64X 16 bit, fixed point	C67X 32 bit, floating point
Main freq. (MHz)	150 – 300	300 – 600	100 – 225
MIPS / MFLOPS	1,200 – 2,400	2,400 – 4,800	600 – 1,350
Data RAM (Mb)	1	2	0.512
Prog. RAM (Mb)	2	8	0.512
Voltage (V)	3.3	1.4	2.4

A useful DSP chip is the SHARC ADSP 2100 series. “SHARC” means Super Harvard Architecture. In addition to the separate data space and program space,

there is a special I/O processor on the chip. The specification of this series chip is given in Table 9.2.

Table 9.2 Specifications of ADSP 2100 series of AD Company

Item	21065 32 bit, floating point	21060 32 bit, floating point	21160 32 bit, floating point
Main freq. (MHz)	66	40	95
MFLOPS	130 – 200	80 – 120	380 – 600
On-chip SRAM (Mb)	0.544	4	2
1024 point FFT (μs)	279	460	97
Voltage (V)	3.3	5.0	1.8/3.3

9.3.3 Development Tools for DSP Chips

It is necessary to have a set of complete development tools for programmable DSP chips. By using these development tools, the user can compile a configurable program, which is supposed to operate on a DSP chip. The development tools can be also used to test the simulation program of a DSP chip. The development tools of a DSP chip can be categorized into two classes. One is a code generating tool. The main function of this tool is to transfer the C program language or program compiler to a DSP program and link them to generate and execute a DSP program. The other one is a code testing tool. The main function of this kind of tool is to test the DSP program and help the user to debug and modify a program.

9.3.4 System Integration

System integration based on DSP chips is an important step in digital sonar design. The main contents of system integration consists of mutual correlated steps, including the right choice of DSP chips, enough program space, convenient means of program testing and online/offline testing of the sonar function. As we have shown before, there are no big differences for different types of DSP chip of the same generation. The most important thing is that, once a DSP chip is chosen, it is necessary to establish a complete test environment and software development tools. It is worth pointing out that to continuously use a chip from one company is a good strategy. It is easy to renew our knowledge and software library, when we trace the same type of DSP chip development.

It is also important to learn how many third parties can provide technical support for the chosen DSP, because the more companies which are involved in DSP chip development, the less time you pay in DSP software development.

9.4 Hardware Adjustment of Digital sonar

After hardware and software installation in the digital sonar, it is necessary to have a series testing procedure for this sonar. This testing can be divided into two classes: electronic function testing and sonar specification testing. The task of sonar designers doesn't finish, even if the sonar testing is successfully completed. They have to participate in the full life management of the sonar system. Only in this case can the sonar designers understand if their design meets the user's requirement.

The electronic function testing includes the correctness of the wire / cable connection between printed boards, cabinets and subsystems. It is also necessary to test if the DSP chip is running normally, if the software is performing correctly, etc.

Sonar specification testing is to test the main technical / tactical specification of the sonar system in a laboratory. The test procedure can be online or offline. Any parameters which appear in the sonar equation have to be tested in the digital sonar. Generally speaking, this is to find any possible problems and to try to solve them. Laboratory testing is the basis of lake testing and sea trial.

9.4.1 Evaluation Procedure of Digital Sonar

The evaluation procedure of a digital sonar system is given in Fig. 9.7. Once a prototype of the digital sonar is manufactured, it is necessary to test it in a laboratory or workshop. The main means for this testing is a sonar simulator. If the digital sonar passed the laboratory testing, then lake testing or sea trial can be carried out. Only when all testing has succeeded can the design of the digital sonar system be confirmed.

In each phase of sonar testing, any problems we find should be corrected until the problem is solved. In this procedure, it is not surprising if part of the hardware and software must be modified or renewed.

The testing procedure in a laboratory can be divided into several phases. In the first phase, one can adopt a little flexibility to see if the digital sonar works normally. In subsequent phases, it is necessary to test the entire system following national, military and naval standards. Reliability and maintainability is one of the most important testing procedures. Generally speaking, the testing contents include anti-vibration, anti-shock, high / low temperature storage, electronic / magnetic compatibility, anti-salt fog and so on. Each failure should be deleted and any problem should be solved in the testing procedure. Fig. 9.8 illustrates the testing procedure in a laboratory.

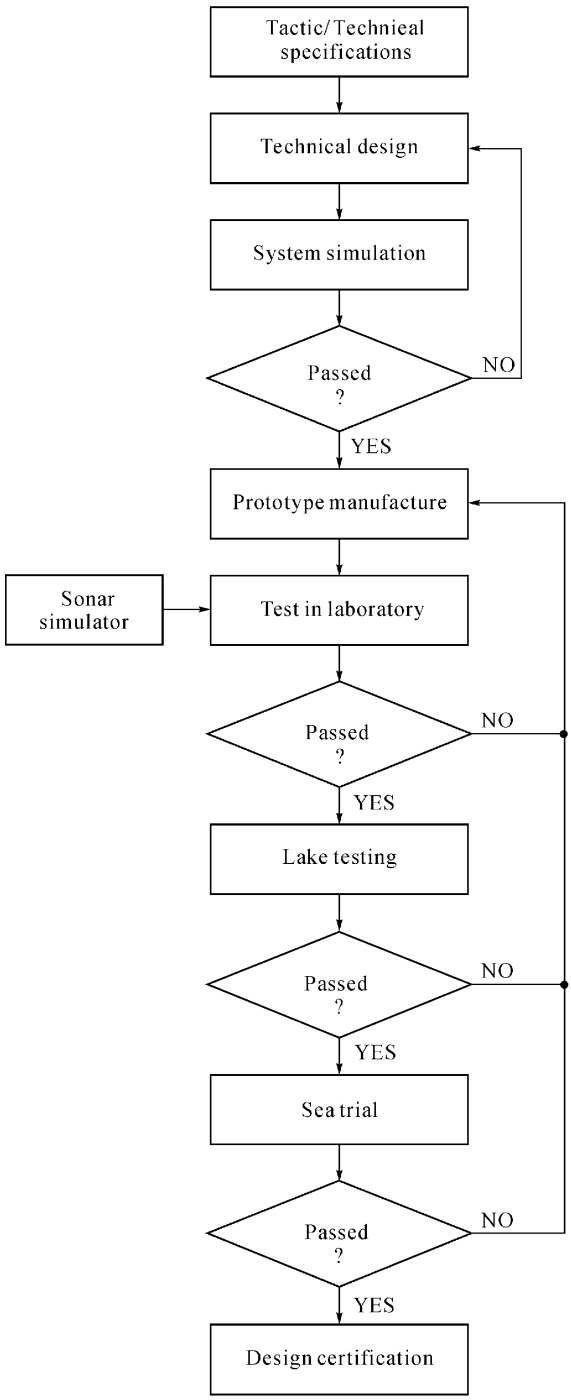


Fig. 9.7. The typical overall procedure of a digital sonar design

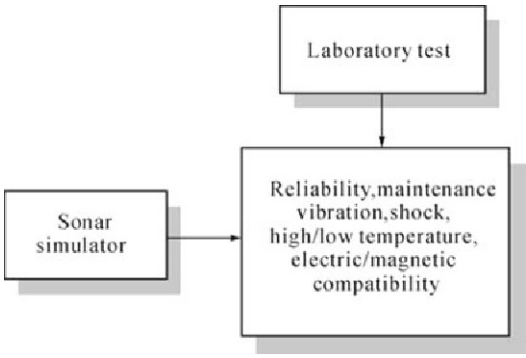


Fig. 9.8. Test items of digital sonar in a laboratory

9.4.2 Step by Step Testing in a Laboratory of a Digital Sonar System

The user of digital sonar concerns the functions of sonar systems, i.e., if the sonar system meets the technical and tactical requirements. But many sonar specifications are impossible to evaluate in a laboratory. As we have shown before, in the sonar design procedure, the tactical specification of a sonar system has already been associated with the technical specification. And some important requirements of a sonar system have been broken down into corresponding sub-systems. It is necessary to show that the technical specifications of a sub-system can usually be tested in a laboratory. Some parameters of the sonar equation, which is the basis of sonar design in some sense, can be evaluated, for example the time gain GT and spatial gain GS.

Of course, there are quite a lot of parameters of the sonar equation that cannot be tested in a laboratory, for example the source level, noise level and transmission loss, etc.

The test method in a laboratory for digital sonar is shown in Fig. 9.9 As we can see, the wet end and dry end of the sonar system should be tested separately. Usually, the components of the wet end, which can be tested in a laboratory, are only a small part of the entire wet end. But most of the dry part should be able to be tested in a laboratory.

Once we complete the wet end and dry end testing, the limited function testing of the entire sonar system is possible. The limited function testing of digital sonar is incomplete function testing, because we can't establish a necessary working environment for the wet end. The main tool of dry end testing is a sonar simulator; it is shown in Fig. 9.9(b). The sonar simulator can provide both analogue and digital signals. An important principle in dry end testing is module-dependent processing.

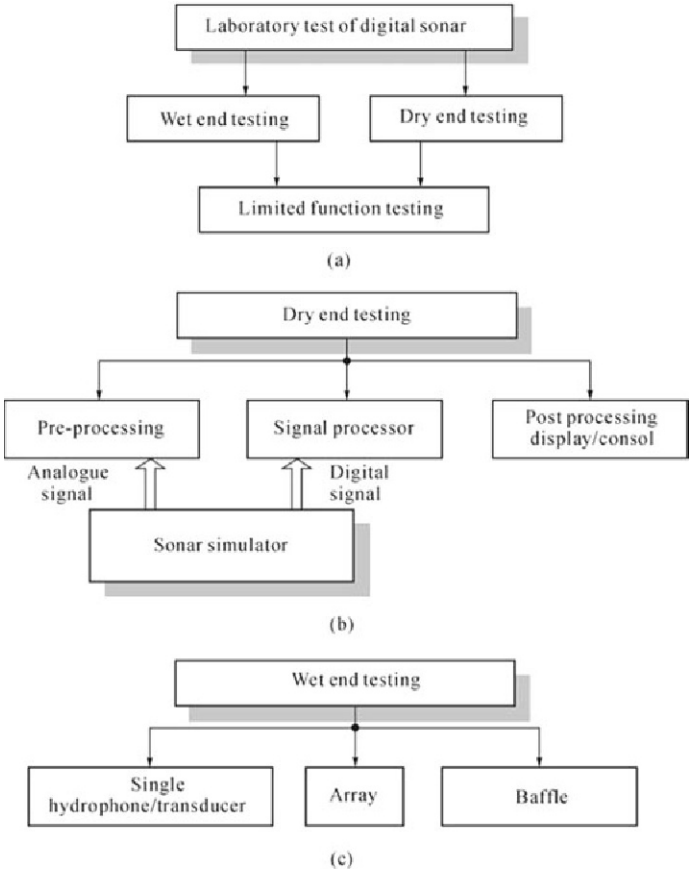


Fig. 9.9. Test of digital sonar in laboratory. (a) Test of entire sonar system; (b) Dry end testing; (c) Wet end testing

Pre-processing

The pre-processor of a sonar system includes a pre-amplifier, filtering and A / D converter. First, we should test if the self noise of the pre-amplifier is low enough for extracting a weak signal. In Chapter 4, we have pointed out that the broadband self noise of a pre-amplifier should be much less than the minimum ocean environment noise; the value is about 10^5 μPa, 100 – 10,000 Hz. Suppose the sensitivity of the receiving hydrophone is about −195 – 185 dB (0 dB = 1 V /μPa), then the output voltage for a hydrophone is 10 – 30 μV. Therefore, the self noise for a pre-amplifier should be less than, for example, 1 μV (broadband, open circuit).

It is also necessary to test the phase shift between any two channels. The test can be carried out in the geometry center frequency of a sonar system. Because the phase error of a channel will finally affect the time delay compensation in beamforming, therefore the specification of the phase error between two channels

becomes a part of the time delay compensation errors.

The role of a pre-filter is to avoid an aliasing effect resulting from sampling. Therefore, the minimum requirement of the upper band frequency of a filter is half of the Nyquist sampling rate. If the sampling frequency is f_s, then the upper band frequency W of a pre-filter should be less then $f_s/2$. But usually the f_s value is greater than $5W$.

For the A/D converter, it is necessary to focus on the dynamic range and the operating point selection. One has to check the uniform performance of signals from multiple channels. We should keep the quantization amplitude as similar as possible for different channels. But the sampling procedure for different channels does not have to happen simultaneously. This is because the time delay differences can be compensated in the beamforming operation. Fig. 9.10 shows two kinds of sampling models, one is a synchronized sampling and the other is an asynchronized sampling. The main type of asynchronized sampling is sequential sampling.

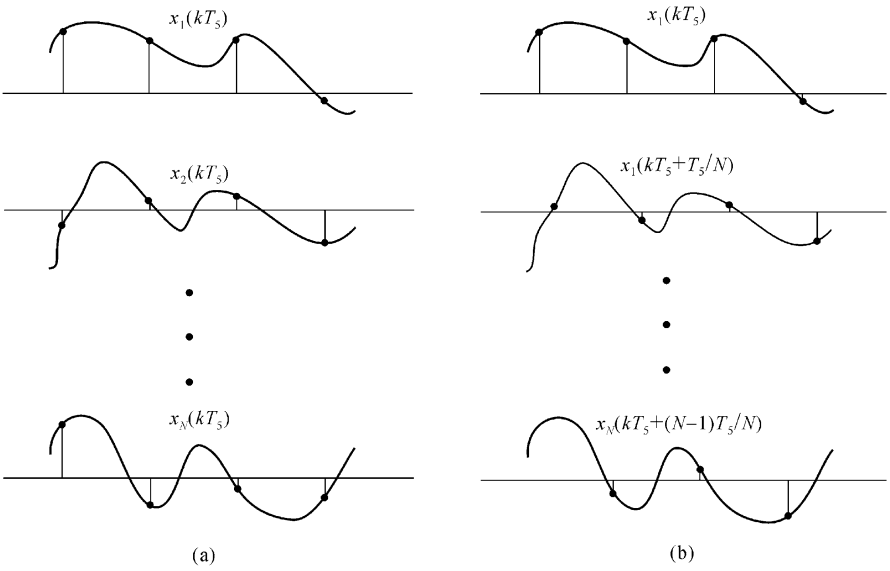


Fig. 9.10. Two different signal sampling methods. (a) Synchronize sample; (b) Asynchronize spample

Synchronized sampling is for sampling the multi-channel signals at the same time. Suppose there are N-channel signals, the sampling duration is T_s. Before sampling, the signals are:

$$x_1(t),\ x_2(t),\ ...,\ x_N(t) \tag{9.2}$$

After synchronized sampling, the data we get are

$$x_1(kT_s),\ x_2(kT_s),\ ...,\ x_N(kT_s) \tag{9.3}$$

This series becomes the input data of the signal processing module.

Sequential sampling is a technical process, which sequentially samples the N-channel signals, the time delay T_s / N is added to each channel

$$x_1(kT_s),\ x_2(kT_s + T / N_s),\ ...,\ x_N(kT_s + (N-1)T_s / N) \tag{9.4}$$

Any additional time delay can be compensated in the beamforming algorithm. Because the distribution of the time delay error is essentially random, therefore the additional time delay doesn't result in any new errors. In a laboratory test, the important thing is to know the relative phase errors between any two channels.

Signal Processing Module

The basic specifications of a digital sonar system should be tested in a signal processing module. These specifications include the directivity function, system gain, and so on. The performances of bearing accuracy and resolution ability of the multi-target are defined by the directivity. Therefore, it is necessary to measure the detailed behavior of the directivity function, which usually depends on the frequency band. The measurement of the directivity function should be carried out not only in a static, but also in a dynamic condition. That means the virtual target is moving along a preset trace.

The measured directivity function should be compared with theoretical calculation results; any abnormal phenomena should be corrected.

The measurement of the system gain of sonar in a laboratory is an important item in digital sonar testing. The recommended method is to measure the beam response in the direction of the incidental angle of the target. Once we know the value of the direct current jump, the system gain of a sonar system can be resolved from Eq. (4.149). Fig. 9.11 illustrates the relationship between direct current jump and system gain for an array with N=128 elements.

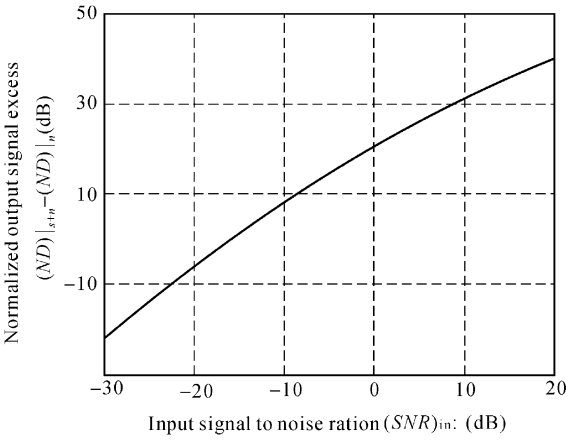


Fig. 9.11. Relationship between input signal to noise ration and output signal excess

Generally speaking, the system gain with a small SNR is of most concern to the sonar designer. A particular value we are interested in is the minimum detectable input SNR, since the FOM can be deduced in terms of the minimum detectable input SNR.

According to the sonar, $SL - NL - TL + GS + GT = DT$, can be rewritten in the following form

$$GS + GT - DT = TL + NL - SL \tag{9.5}$$

The negative value of the right hand of Eq. (9.15) is the minimum detectable input SNR, i.e.

$$(SNR)_{\min} = SL - NL - TL \tag{9.6}$$

We found that the minimum detectable input SNR is just equal to the difference in system gain $GS + GT$ and detection index DT.

Testing of Post Signal Processing Module and Display / Consol

The task of the testing of the post-processing module is to check if the designed function and performance of a digital sonar can be reached and realized. And also we have to test the interface between the signal processing module and display / consol.

The algorithms performed in the post-processing module are various and complicated. It is necessary to test them individually and to see if the expected operation is going well. The most important items are the amplitude display of beam pattern, time / bearing display of multi-beam system, automatic tracking function and target recognition. Finally, one should check if the symbols, menu and image display on the sonar screen, are well configured and have a friendly interface with the sonar operator.

If the sonar system has a failure self-testing program, it is necessary to verify this in laboratory testing.

Actually, the entire hardware testing of a sonar system is carried out by sonar simulator. The more comprehensive the sonar simulator is, the better the system testing should be. Otherwise, it is almost impossible to test a sonar system in a laboratory without a sonar simulator.

9.4.3 Wet End Testing of the Sonar System in a Laboratory

Generally speaking, for a digital sonar system, only a small part of the wet end can be tested or measured in a laboratory. In particular, if the sonar system has a large aperture array, it cannot be tested in a water tank or pool. Therefore, it is impossible to make associate testing with the dry end. But some performances which belong to the wet end can be tested in a laboratory, including:

(1) The transmitting power of the high frequency transducer and the sensitivity of the high frequency hydrophone can be measured in a laboratory. Some low

frequency hydrophones can be qualitatively tested in the atmosphere, for example the towed array.
(2) The absorption, reflection and transmission characteristics of the sound baffle can be tested in a laboratory.
(3) The attached pre-amplifier of the low noise hydrophone can be tested.
(4) Some specific components of digital sonar can be tested in a laboratory, for example, the cable, the winch, acceleration sensor, etc.

9.4.4 Limited Function Associate Testing of the Dry End and Wet End

The limited function associate testing of the dry end and wet end can be understood as shore-based testing of the sonar system. Once we complete the separate testing of the wet end and dry end, one can carry out associate testing for both the dry end and wet end. But we have to remember that this kind of testing is limited function testing because the wet end cannot work properly in a laboratory. The complete testing of the entire sonar system must be carried out in a lake or at sea.

The main task of the associate testing of the dry end and wet end is to see if the connection of the entire digital system is complete and if some basic specifications are well established.

9.5 Tactical and Technical Performance Evaluation of Digital Sonar: Comparison of Results of Laboratory Testing, Lake Testing and Experiment at Sea

The final place for sonar performance evaluation is field testing, meaning sea trials. There are many research laboratory and facilities involved in basic research of underwater acoustics and sonar design[23-30], the experiences of research work in underwater acoustics and ocean engineering is "do experiment at sea". Only the results of sea trials can pass judgment on the sonar system. Is this statement exaggerated? No. Underwater acoustics is an experimental science. Almost all software and programs, which related to underwater acoustics are evaluated in sea trials, again and again[31-34]. The performance of a sonar system strongly depends on the ocean environment. It is hard to theoretically and completely describe real ocean conditions. The multiple and changing characteristics of the ocean environment demonstrate its changing space / time performance and frequency spread phenomena. All of this is difficult to describe exactly. It is necessary to stress that any system simulation on a computer cannot replace experiments at sea. Therefore, the results of sea trials become the final judgment and authority[35-40]. Why are experiments so important? Just as Kennedy said[36]: "We experiment to

learn, and we learn to transform the force from what it has been, to what it needs to be." Production of sonar systems will have a minimal probability of failure if they have survived rigorous experimentation throughout their development.

In this section, the basic outline of field testing techniques on a lake or at sea are discussed. We will try to describe the potential differences between the results of testing in a laboratory, in a lake and at sea.

9.5.1 From Sonar Design to Experiments at Sea

When we were talking about the sonar equation, we mentioned that some parameters appeared in the sonar equation that must be measured in the real ocean environment. For example, the source level SL, background noise level, reverberation level RL and transmission loss TL.

In Fig. 9.12, the procedure leading from sonar design to experiment at sea is illustrated. Suppose the technical and tactical specification in the very beginning phase is 1.0, which is the value of acceptance by the sonar designer. In the design phase, one has to increase this specification by as much as 30%. Indicated by symbols, we can write $Q_0 = 1.0$, $Q_1 = 1.3$. Based on this value, we can carry out our design and finally develop a hardware / software system. The test result based on laboratory specifications is about the same as $Q_{lab} = Q_1 = 1.3$, but there is some potential error existing in the designed specification and test result. The fluctuation value could be more or less than 5%. In lake testing specifications, the value may be decrease to $Q_{lake} = 1.2$, And the dispersive level could be 10%. In experiments at sea, the test result based on specifications may decrease to $Q_{sea} = Q_0 = 1.0$, and the dispersive level could be 15%.

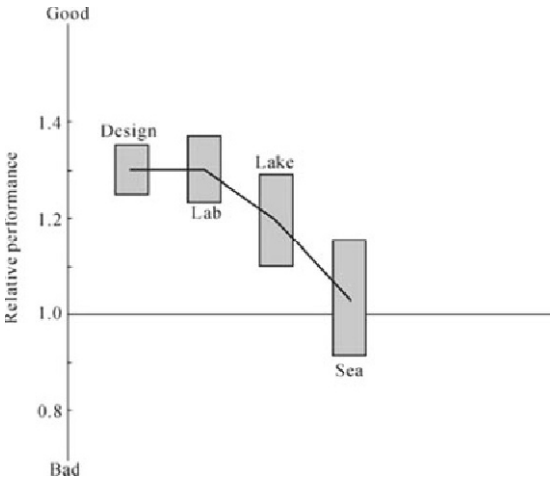


Fig. 9.12. The empirical results of sonar performance in different develop phase

The relationship between Q_0, Q_1, Q_{lab}, Q_{lake}, and Q_{sea} is listed in Table 9.3.

Table 9.3 Specification comparison from sonar design to at sea experiment

Item	Notation	Value	Dispersive level
Accepted specification before design	Q_0	1.0	
Design specification	Q_1	1.3	5%
Testing in laboratory	Q_{lab}	1.3	5%
Testing in lake	Q_{lake}	1.2	10%
Testing at sea	Q_{sea}	1.0	15%

The differences between the predicted value and the measured value of sonar specification are not fixed. It varies with the practical objects. There is no doubt that the theoretical results of research work strongly depend on the knowledge we have. The empirical results are illustrated in Fig. 9.13. They qualitatively present the dispersive level of the main specifications of a digital sonar system. The results are shown in Table 9.4.

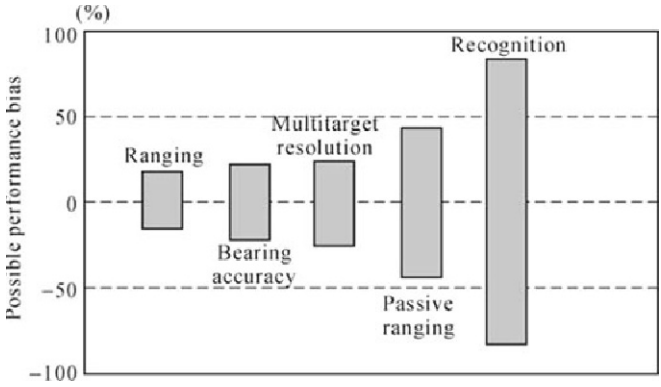


Fig. 9.13. Empirical results of sonar performance in various condition

Table 9.4 Dispersive level between the expected value and measured value of sonar specifications

Item	Dispersive level (%)
Ranging distance	±15
Accuracy of bearing angle and range estimation	±25
Angle resolution of multiple target	±35
Passive ranging	±50
Target identification / classification	±85

The empirical results presented in Table 9.4 reflect the real status of research work in this field.

9.5.2 Measurement of Source Level and Sound Pressure Index

Fig. 9.14 illustrates the basic idea of source level measurement of radiated noise of an underwater target. There are usually two methods that can be adopted. One is

single hydrophone measurement and the other is hydrophone array measurement. When we use a single hydrophone to measure the source level of radiated noise of an underwater target, it is necessary to have the hydrophone as close to the target as possible, because a big enough SNR is necessary. If the measured target is not too close to the hydrophone, or if the radiated noise is very quiet, then we should use the hydrophone array to measure the target noise, since the hydrophone array has spatial gain to enhance the input SNR. In both situations, the hydrophone or geometrical center of the hydrophone array should be on the same horizontal level as the measured target and the hydrophone must be inspected before use.

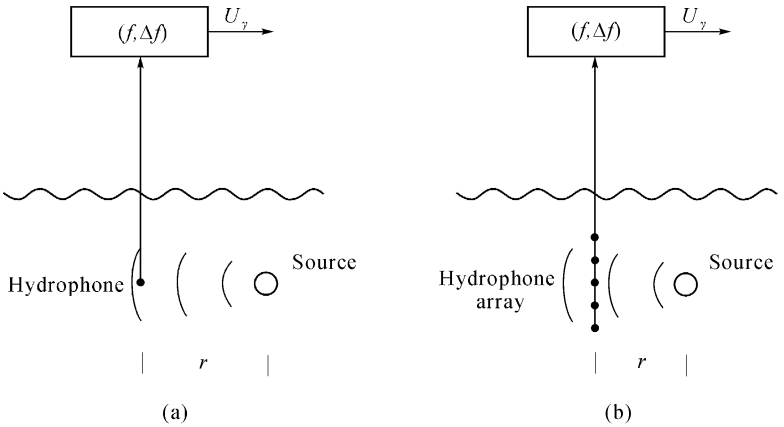


Fig. 9.14. Measurement of radiated noise of underwater target. (a) Single hydrophone; (b) Hydrophone array

Example

Suppose the sensitivity of a standard hydrophone is $M = -195$ dB, reference 0 dB value is M_0=1 V / μPa, if the target passes through the hydrophone at a distance of r = 100 m. At frequency f = 1,000 Hz and $\Delta f = 10$ Hz, the measured voltage output of sound pressure is U_r = 100 μV. We are going to find the source level SL in 1 m. Based on this assumption, we have

$$Q_0 = 20 \lg \frac{M}{M_0} = -195 \ \text{(dB)} \tag{9.7}$$

At distance 1 m, the decibel value of the sound pressure should be

$$SL = 20 \lg \frac{p_1}{1 \ \mu\text{Pa}} \tag{9.8}$$

If we can find the sound pressure P_r in r m, then it is easy to find SL in terms of square attenuation law of near range propagation. We have

$$SL = 20\lg\frac{p_1}{1\ \mu\text{Pa}} = 20\lg\frac{U_r}{M\times 1\ \mu\text{Pa}} + 20\lg\frac{r}{1\ \text{m}} - 10\lg\Delta f$$

$$= 20\lg\frac{U_r}{M_0} + 20\lg\frac{M_0}{M} + 20\lg r - 10\lg\Delta f$$

or

$$SL = 20\lg U_r - Q_0 + 20\lg r - 10\lg\Delta f \tag{9.9}$$

where the unit of U_r is V, the unit of r is m, and the unit of Δf is Hz.

Substituting the assumed value $U_r = 10^{-4}$, $r = 100$, $\Delta f = 10$ into Eq. (9.9), we have $SL = -80+195+40-10=145$ (dB).

It is worth showing that the distance between the hydrophone and target should be as close as possible, since the SNR will substantially affect the accuracy of the measurement. We will briefly discuss the effect of SNR in source level measurement.

Suppose the received signal is $x(t) = s(t) + n(t)$, where $s(t)$ represents the radiated noise of the target and $n(t)$ is ambient noise. Let $\text{Var}[s(t)] = \sigma_s^2$, $\text{Var}[n(t)] = \sigma_n^2$. If $(SNR)_{\text{in}} = \sigma_s^2 / \sigma_n^2 = 10$ (dB), when the output of noise is only 1.0, then the output of signal plus noise should be 11.0. The standard deviation value is $\sqrt{10}$ and $\sqrt{11}$. The noise effect can be neglected. However, if $(SNR)_{\text{in}} \le 10$ dB, the measurement error cannot be neglected.

From this discussion, we know that if the radiated noise producing the SNR at 1 km is 0 dB, then the distance for measuring should be no greater than 100 m, otherwise the accuracy of the measurement will be poor.

Now we turn to discuss the measurement of transmitting power. We can measure the transmitting power of a transducer in a lake or at sea (Fig. 9.15). It is also necessary to use a standard hydrophone to measure the transmitting power.

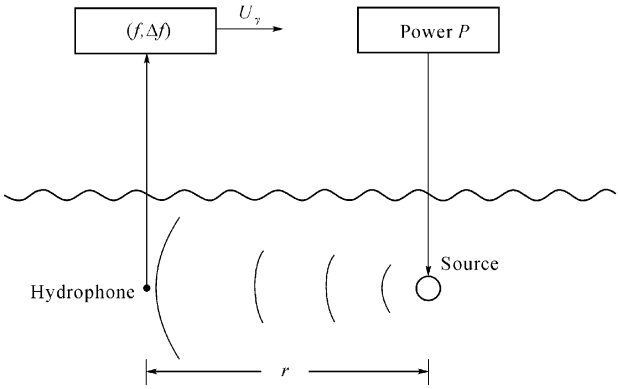


Fig. 9.15. Measurement of acoustic source level

Example

Suppose the sensitivity of a standard hydrophone with amplifier is $Q_0 = -160$ dB (ref. 0 dB = 1 V /μPa). By using this hydrophone, we measure the sound intensity of a transmitting transducer, which is omni-directional, $U_r = 800$ mV at distance $r = 20$ m. How large is the sound intensity of this transducer?

To solve this problem, we have first to transfer U_r to sound pressure p_r, and then to calculate the index sound pressure p_1 at 1 m. Once we get the index sound pressure, it is easy to find the value of the sound intensity.

Denote the reference sound pressure as $M_0 = 1$ V/μPa and we know

$$Q_0 = 20\lg\frac{M}{M_0} = -160 \text{ (dB)} \tag{9.10}$$

Since the reference index sound pressure 1 μPa is as 0 dB, therefore

$$\begin{aligned} SL &= 20\lg\frac{p_1}{1\ \mu\text{Pa}} = 20\lg\frac{U_r}{M\times 1\ \mu\text{Pa}} + 20\lg r \\ &= 20\lg\frac{U_r}{M_0} + 20\lg\frac{M_0}{M} + 20\lg r \\ &= -Q_0 + 20\lg r + 20\lg U_r \end{aligned} \tag{9.11}$$

where, the unit of r is m, the unit of U_r is V. Substituting $r = 20$, $U_r = 0.8$ into Eq. (9.11), we get $SL = 160 + 26 - 2.4 = 183.6$ (dB). In Chapter 4, we have shown that the relationship between sound intensity P and index sound pressure SL for an omni-directional transmitting transducer is given by

$$10\log P = SL - 170.5 \text{ dB} \tag{9.12}$$

The unit for P is W, and 1 W = 170.5 dB. Substituting SL=183.6 dB in to Eq. (9.13) and we have P = 20.4 W.

Similarly, if the source level is know as SL, and the sensitivity of the hydrophone is M_0, then the voltage value of the index sound pressure at 1 m should be

$$p_1 = 10^{(SL+M_0)/20} \text{ V} \tag{9.13}$$

In measuring the transmitting sound power of the transducer, it is necessary to have a certain distance between the receiver and the transducer, i.e., it must satisfy the far field condition. The far field condition depends on the transmitting frequency and the aperture of the transducer. For high frequency ($\geq$10 kHz), a

distance greater than l^2 / λ can be considered as far field. This distance is called the Rayleigh distance and l is a quantity related to the aperture of the transducer.

In general case, it is necessary to place the receiver beyond the Fresnel area and the magnitude is in the order of $r \approx (\lambda z)^{1/2}$ $(z >> \lambda)$, where λ represents the wavelength of the transmitting signal and z is the distance between receiver and transmitting transducer. Usually, $z \geq 100\lambda$.

9.5.3 Measurement of Target Strength and Underwater Signal Channel

In the active sonar equation, there is a very important parameter, namely the target strength. For a small object or model, the value of the target strength can be measured at the lake, but for a submarine the only way to find its reflective characteristics, or target strength, is at sea.

From the theoretical point of view, a spheroid is the ideal model in simulation testing. Some parameters can be measured on a small scale by changing the frequency and model size. However, in underwater acoustic studies many parameter measurements cannot be scaled down to a small size, since many environmental conditions cannot be scaled down. Therefore, the uniquely effective method of finding the target strength of a submarine is measurement of a real submarine at sea.

Fig. 9.16 is a record of the receiving signal of an active sonar. It consists of a transmitting signal, the echo from submarine and various multiple paths effects.

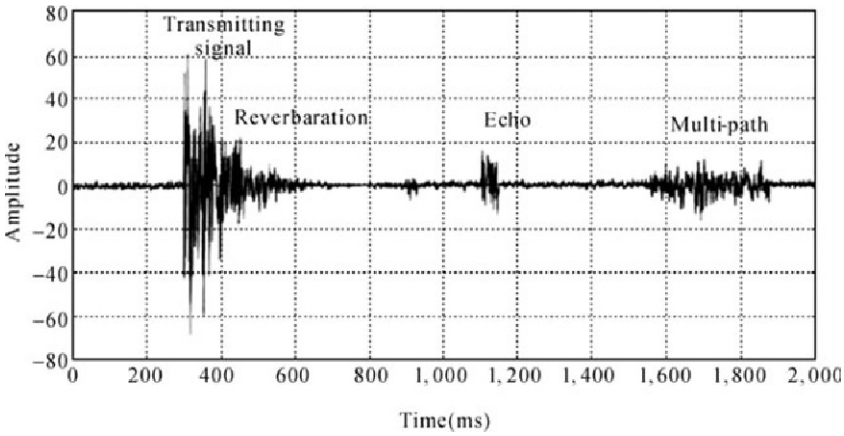


Fig. 9.16. Active sonar signal and echo signal

The transmitted signal is usually a narrow rectangular pulse. The echo signal is obviously increased in length. There also exists frequency spread; the magnitude

depends on the architecture and propagation range. The multiple paths effect consists of a reflected signal from sea bottom to sea surface. The echoes from different places on the sea bottom, sea surface and some water bodies, result in an echo-like signal; because the existence of multiple paths of that signal the description of an underwater acoustic channel become more complicated. A simply way to describe an underwater acoustic channel is to consider the channel as a signal filter. Whose transfer function of this filter can be divided into two parts. One part is stable, like a general linear filter, whose transfer function is the Fourier transform of a certain impulse response function; The other part is random variable; The magnitude of this part reflects the channel's varying characteristics. For the underwater acoustic channel in any real ocean, it is necessary to measure the transfer function and find a suitable model to best match the performance of this channel. Then sonar performance prediction is possible.

As a matter of fact, if we perform detailed analysis for the echo signal of a submarine, we will find that it is possible to distinguish the echoes from different parts of the submarine. That means we can consider the submarine as a distributed body object; the transmitting signal highlights this body and gets different echoes from different parts of the body. Some parts of a submarine have strong reflective characteristics, and some parts of a submarine have weak reflective characteristics. Therefore, the reflected signal is distributed in different ways, highlighted or dimmed. It is reported that[41] the echo signal reflected from stern win and bridge of submarine usually is plane wave; the echo signal from bow of submarine is spherical wave and the echo signal from hull of submarine is cylindrical.

The extraction of the highlight effect not only depends on the echo signal analysis, but also depends on the measurement of the underwater acoustic channel. It has become a hot topic in active sonar signal analysis.

There are several important applications of the measurement of the transfer function of an underwater acoustic channel. We are going to discuss two problems. The first problem is channel equalization and the second is the time reversal problem (Fig. 9.17).

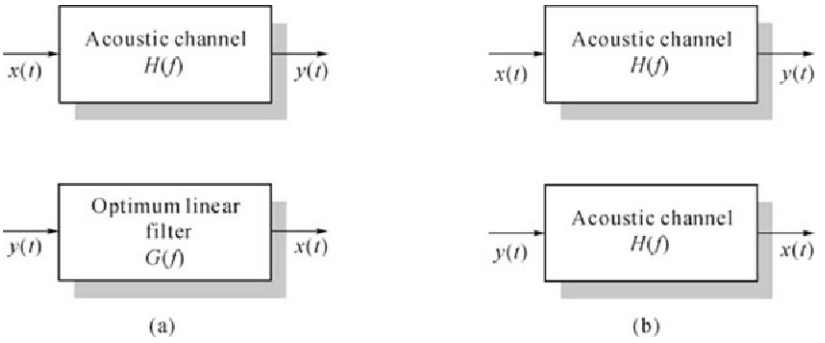


Fig. 9.17. Signal recovery problem in underwater acoustics. (a) Channel equalization; (b) Time inverse

(1) Channel equalization.

Suppose the transfer function of an underwater acoustic channel is $H(f)$, signal $x(t)$ propagating through this channel becomes $y(t)$. The question is what kind of channel can transfer $y(t)$ to $x(t)$, given signal $x(t)$ is known. If the receiving signal $y(t)$ has a big enough SNR, it is possible to solve the unknown channel transfer function $H(f)$ and the answer to the question is

$$G(f) = 1 / H(f) \tag{9.14}$$

The solution is simple, but the realization is complicated since the transfer function $H(f)$ may have zero points. Therefore it is hard to implement the transfer function in a digital filter.

(2) Time reversal.

When we receive the signal $y(t)$, what kind of signal should be re-transmitted on the same channel to obtain the original signal $x(t)$? The solution to this question is easy, but not obvious. The solution will lead to a very important concept: the time reversal signal.

In fact, the spectrum of $y(t)$ is the product of the spectrum of $x(t)$ and $H(f)$, i.e.,

$$Y(f) = X(f)H(f) \tag{9.15}$$

The signal to be solved should have the spectrum of

$$Z(f) = \frac{Y^*(f)}{|H(f)|^2} \tag{9.16}$$

since, in this case, the spectrum of the output signal of the channel is

$$\frac{Y^*(f)}{|H(f)|^2} \times H(f) = \frac{Y^*(f)}{H^*(f)} = X^*(f) \tag{9.17}$$

The signal with the spectrum in Eq. (9.16) is actually related to the signal $y(-t)$. because the spectrum of $y(-t)$ is $Y^*(f)$ and $y(-t)$ is just the reversal in time of $y(t)$.

Generally speaking, the measurement of an underwater acoustic channel is a very complicated problem. And until now, there has not been an effective method for doing this. In fact, it is very difficult to discuss this problem abstractly, since any problem concerning the underwater acoustic channel is always associated with the ocean environment and is also related to the position of the transducer and hydrophone.

A useful method for measuring the underwater acoustic channel is to use the sound of the underwater explosion. The explosive sound usually has a quite wide frequency range. At the receiving end we can put a horizontal and vertical array to pick up the explosive sound, and the comb filter bank can find the spectrum of the received sound pressure. Therefore, the frequency dependent relationship of the underwater acoustic channel can be estimated. This is important for sonar range prediction and estimation.

9.5.4 Calibration of Sonar Specification in Terms of the Result of Experiments at Sea

The tactical and technical specifications of a digital sonar system strongly depend on the ocean environment. Some specifications can be considered environmentally independent quantities, for example the minimum detectable SNR. But most of the specifications are environmentally dependent, for example the range.

A popular method is to indicate the environmental conditions when assigning the sonar specification. This is a reasonable method. But we cannot expect that every specification of the sonar system can be tested in the environmental conditions, as described. Therefore, it is an important problem within the sonar specification calibration problem. If the assigned environmental condition is L_0, in this condition we have sonar specification Q_0. But if the measured result is Q_1 in environment L_1, how can we compare Q_0 and Q_1? Should we accept Q_1? The specification calibration problem is a really important issue in sonar applications. We have to pay more attention to this issue, to find a suitable resolution.

As an example, the ranging distance calibration is discussed below. The law of transmission loss $TL(r, f/L)$ of the underwater acoustic signal plays a key role in range calibration, where r represents distance, f is frequency and L denotes some SSP (sound speed profile). For active sonar, since the signal is sinusoidal or in a narrow band, the transmission loss function $TL(r, f/L)$ is easy to find. But for passive sonar, the signal is broadband; the transmission loss function depends on the broadband signal and it is necessary to find the value of $TL(r, f/L)$ in each frequency bin and distance. The calculation will be a little more complicated than that in the case of active sonar.

Suppose the power spectrum density of signal $x(t)$ is $K_x(1, f)$, where the digit 1 indicates a quantity of 1 m. The energy transmitted by the radiated noise is

$$\sigma_s^2(1) = \int K_x(1, f)\mathrm{d}f \tag{9.18}$$

At the distance of r, the signal power spectrum becomes

$$K_x(r, f) = K_x(1, f)TL(r, f \mid L) \tag{9.19}$$

and the signal power becomes

$$\sigma_s^2(r) = \int K_x(r, f)\mathrm{d}f \tag{9.20}$$

The relationship between $\sigma_s^2(r)$ and r is the transmission loss of the broadband signal. Sometimes we prefer to use the transmission loss of the geometrical center frequency in the band to replace the broadband transmission loss. The calculation will be simplified, but the resulting error should be noted.

It is worth showing that it is not necessary to measure the value of $TL(r, f \mid L)$ in the arbitrary frequency f. As a rule, it is sufficient to measure the

value in the 1 / 3 octave frequency band. From 100 Hz, the frequency bins are listed below: 100, 125, 160, 200, 250, 315, 400, 500, 630, 800, 1,000 (Hz). The bandwidth is about 23% of center frequency.

Now we turn to an analysis of the procedure for range calibration. For simplification, suppose for a fixed frequency we already know the transmission loss laws $TL(r, L_0)$ and $TL(r, L_1)$. The frequency sign is omitted. Fig. 9.18 shows the procedure for calibration.

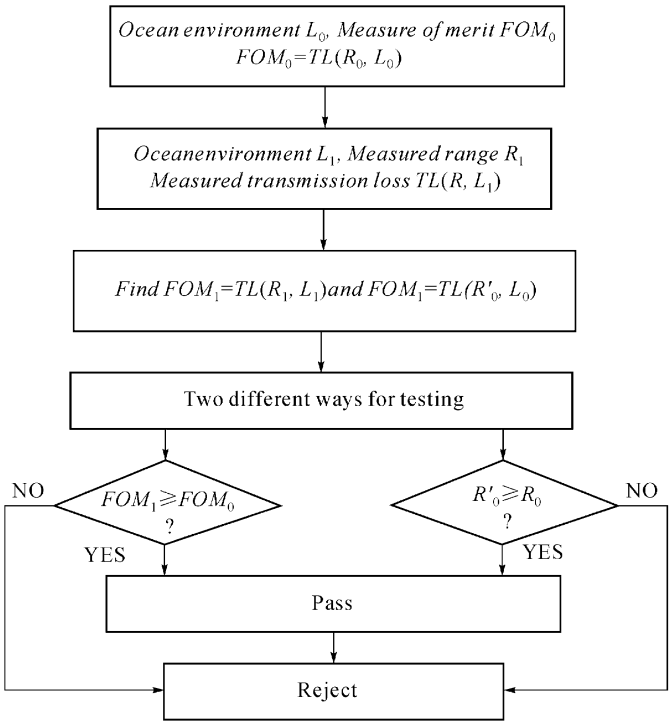


Fig. 9.18. The problem of sonar range re-calculation

Suppose in the sonar design phase, the specification of range distance R_0 is given by SSP condition L_0. Based on this condition, the figure of the merit value is FOM_0 and

$$FOM_0 = TL(R_0, L_0) \tag{9.21}$$

In practical experiments at sea, which are carried out in SSP condition L_1 and where the measured sonar range is R_1 of course we have

$$FOM_1 = TL(R_1, L_1) \tag{9.22}$$

Are there any conclusions resulting from this at sea experiment? Obviously,

FOM_0 and FOM_1 may be different. Based on FOM_1, in SSP L_0 we can find another sonar range R_0'. That is

$$FOM_1 = TL(R_0', L_0) \tag{9.23}$$

It is necessary to compare $R_0' \geq R_0$? $FOM_1 \geq FOM_0$? The YES/NO answer to these two questions is the same. If the answer is positive, i.e., "YES", then we confirm that the sonar range distance passes the test; otherwise, the sonar specification of the range distance failed in at-sea testing. Fig. 9.19 illustrates this procedure. If L_0 represents an isothermal SSP and L_1 represents a weak negative gradient SSP, then the results will be similar to the plot shown in Fig. 9.19.

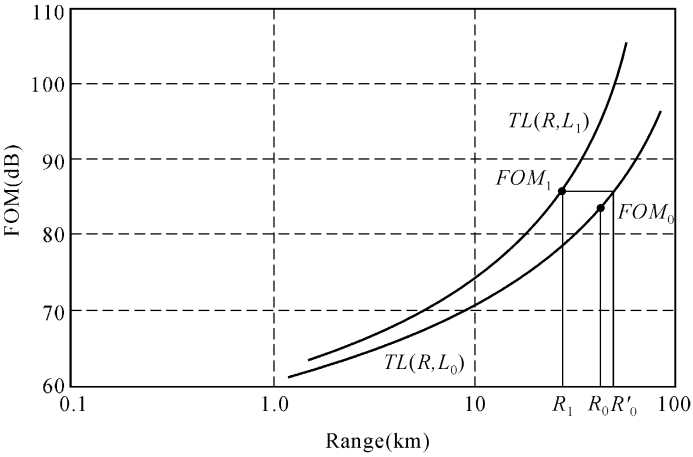


Fig. 9.19. Example of sonar ranging calculation

Table 9.5 shows some at sea experimental results of the US navy, reported by Jenkins[42].

Table 9.5 At sea experimental results of US navy sonar (Sea State 5, 1993)

Frequency (kHz)	1	3	5	7.5	10	15
Source level (dB)	232	232	232	232	232	232
Target strength (dB)	15.0	15.0	15.0	15.0	15.0	15.0
Noise level (dB)	66.9	58.8	55.0	52.0	49.9	46.9
Array gain (dB)	25.0	25.0	25.0	25.0	25.0	25.0
Detection threshold (dB)	19.0	19.0	19.0	19.0	19.0	19.0
FOM	93.1	97.1	99.0	100.5	101.6	103.1
Ranging distance (km, Aegean Sea)	116	87	58	36	28	15
Ranging distance (km, Korean Sea)	9.8	10.0	10.6	9.8	9.2	7.2

In this table, we can see that the ranging distance of the same sonar has big differences in different places. For example, at 1 kHz in the Aegean Sea, the

ranging distance is 116 km, but in the Korean Sea it is 9.8 km.

9.5.5 Statistical Average Problem in Experiment at Sea

Now we turn to discuss a special problem in sea trial, the statistical average problem. This problem is so common that any sonar designer and user may have to face this problem when digital sonar is passed on from the sonar designer to the military user. Suppose the bearing accuracy of a sonar system is 0.5°, then we have to test this specification in the same conditions many times (for example, more than 30 times), and then make a statistical average. If the result of the statistical average is less or equal to 0.5°, the testing is passed. Otherwise, this specification has failed to pass. Here, the probability of a wrong decision or rejection must be a small probability event. But how many times is testing necessary and how small a probability of rejection can be accepted? Since a sea trial is a high cost business and it is hard to run testing in the same conditions many times, a reasonable testing time and rejection rule is necessary and it must be confirmed by both the sonar designer and the military user.

A very important and useful inequality called the Chebychev inequality, which is the basis of statistical average error, is introduced in the following.

Chebychev Inequality

Suppose Y is a random variable with mathematical expectation $E(Y)$ and variance Var(Y), then for any given positive value $\varepsilon > 0$, we have

$$P\{|\, Y - E(Y)\,| \geq \varepsilon\} \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2} \tag{9.23}$$

This inequality is the basis of many theorems in probability theory, for example the large number law.

Now, if we have N independent observations $x_1, \ldots, x_N$ $x_1, \ldots, x_N$, we can think this is the result of N independent target bearings, N independent target recognitions, or N independent torpedo alarms. If $E(x_i) = \mu$, $\mathrm{Var}(x_i) \leq \sigma^2$, based on the Chebychev inequality,

$$P\left\{\left|\frac{1}{N}\sum_{i=1}^{N} x_i - \mu\right| > \sigma\right\} \leq \frac{1}{N} \tag{9.24}$$

It means when the number of observations is large enough, the average value of multiple testing must steadily improve the expectation value. Otherwise, the assumption of $\mathrm{Var}\,(Y) \leq \sigma^2$ is not correct. In mathematical statistics, the number

of N is usually supposed to be 30. This is considered as the standard of a "large quantity sample". The following "small probability event" rule is accepted in practice.

Proposition of Small Probability Event

In one experiment, the occurrence probability of a "small probability event" is zero. A well-known standard is that 0.05 can be considered as a small probability. It is worth showing that, "one experiment" should be understood as the result of one package experiment for a certain specification, and not indicate just one time experiment. For example, if we want to carry out a target recognition experiment, testing 10 times as one package, then this package, which consists of 10 tests, should be considered as one experiment.

Theoretically, it is easy to prove that if the occurrence probability of a certain event A is p $(0 < p < 1)$, if the number of independent observations N is very large, then the probability of occurrence Np times of event A in N times experiments is close to 1. The proof of this fact is simple.

Denote the N independent experiment as $x_1, \ldots, x_N$. Suppose

$$x_i = \begin{cases} 1 & \text{event } A \text{ occurrence} \\ 0 & \text{event } A \text{ not occurrence} \end{cases} \tag{9.25}$$

$P\{x_i = 1\} = p$, $P\{x_i = 0\} = 1 - p$ $(1 \le p \le 1)$. Define a random variable $x_0 = x_1 + \ldots + x_N$.

Then x_0 represents the number of event A occurrences in N independent experiments. The random variable x_0 satisfies the Bernoulli distribution, i.e.,

$$P\{x_0 = k\} = C_N^k p^k (1-p)^{N-k} \tag{9.26}$$

In terms of the central limit theorem described in Chapter 3, we have

$$\lim_{N\to\infty} P\left\{\sum_{k=1}^{N} x_k - Np \le \sqrt{N}\sigma x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)\mathrm{d}y \tag{9.27}$$

where $\sigma = p(1-p)$.

Based on Eq. (9.37), it is easy to prove that

$$P\left\{\sum_{k=1}^{N} x_k > (1-\varepsilon)Np\right\} \to 1 \quad \text{when} \quad N \to \infty \tag{9.28}$$

and ε is an arbitrary given positive value.

Example

Suppose the correct alarm probability of a torpedo alarm system is 0.75, then

in an experiment of 100 times, the probability of 75 correct alarms should be close to 1. The result presented previously has an important meaning in practical sea trial. Fig. 9.20 illustrates the case of sample numbers $N = 12$ and $N = 15$. This is an alarm calculation example. Suppose the correct alarm probability of one experiment is $p = 0.75$. If the number of experiments is 12, then the probability of a correct alarm 6 times is 0.98. The probability of having at least 8 times a correct alarm is 0.84. From this we can see that the small quantity sample has potential risk. If the sample number is increased, then the situation will be improved. Fig. 9.21 is the case of $N = 30$. This time the probability of a correct alarm at least 20 times (note: 8:12 = 20:30) increases to 0.90.

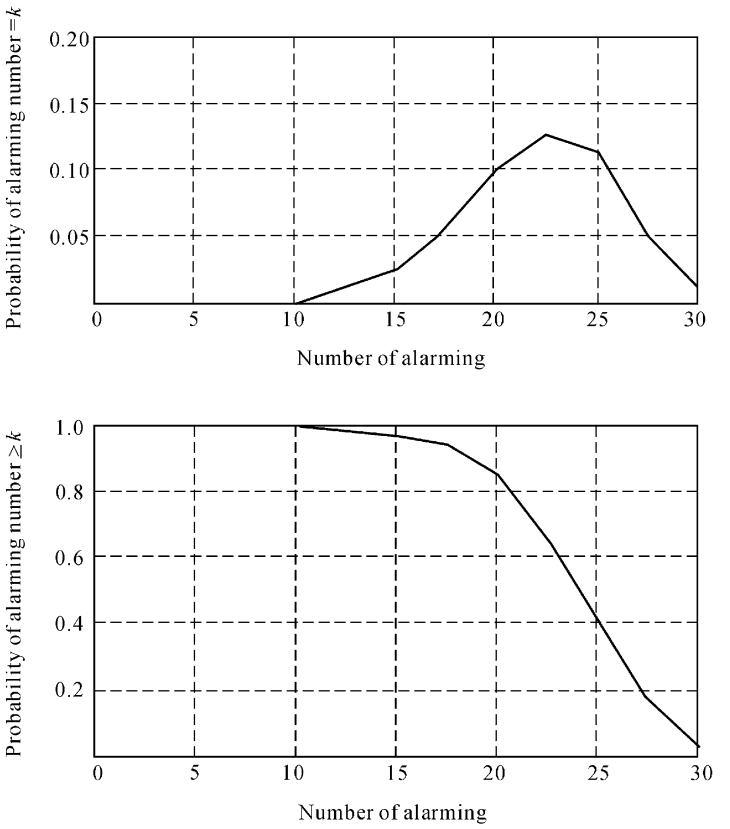


Fig. 9.20. Probability of small size samples ($n = 12$, $p = 0.75$)

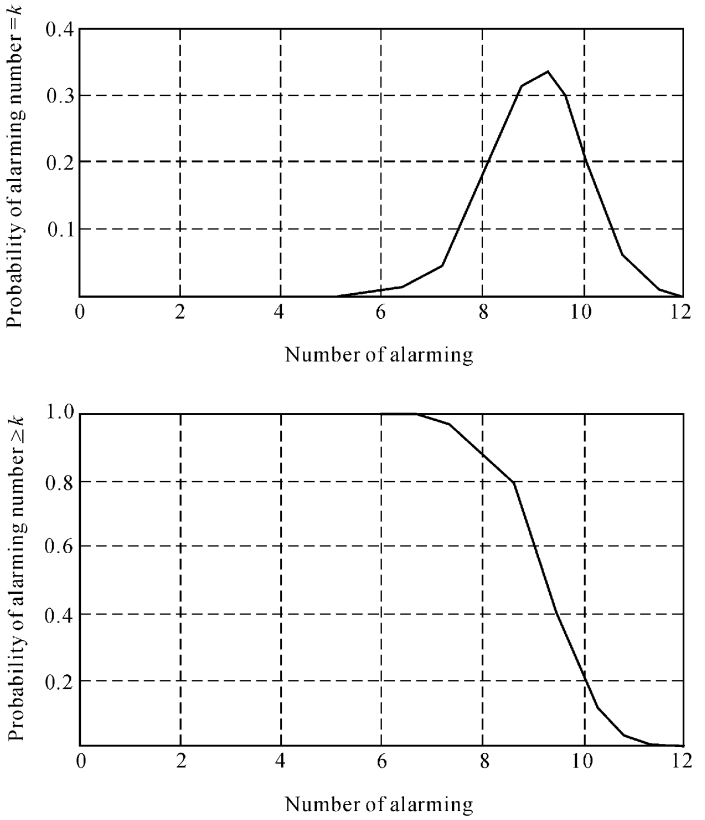


Fig. 9.21. Probability of large size samples ($n = 30$, $p = 0.75$)

9.5.6 *Central Moment and Original Moment*

In sea trial testing of digital sonar, there is an important matter we have to be concerned with, i.e., the measurement error problem. Suppose the bearing accuracy of a sonar system is 1°. How can we verify this specification? The frequent measure is randomly to choose a set of bearing directions, for example, $\theta_1, ..., \theta_N$. These values must be "true value", i.e., the values can be confirmed by other means, rather than by the sonar system itself. The practical test results of a sonar system are x_i $(i = 1, ..., N)$. The bearing accuracy of a sonar system might be considered as

$$\varepsilon = \left[\frac{1}{N} \sum_{i=1}^{N} (x_i - \theta_i)^2 \right]^{1/2} \tag{9.29}$$

This definition is easy to understand intuitively. But compared with basic statistical theory, there are still some problems. The main reasons are:

(1) Theoretically, the true value to be tested should be fixed. That means we have to carry out multiple testing for the same true value and then perform a statistical average. However, in practice, it is sometimes difficult to realize.

(2) The true value itself may contain a systematical error, which doesn't result from the sonar system, but from other factors, for example, the radar calibration results.

Based on these reasons, we have to find a new statistical average algorithm, which may not only avoid the problems described above, but can also be easily realized in sea trials.

Suppose X is a random variable with k-order origin moment

$$v_k = E(X^k) \tag{9.30}$$

The k-order center moment is defined as

$$\mu_k = E[(X - E(X))^k] \tag{9.31}$$

If the realization of random variable X is $x_1, ..., x_N$, then the statistical value v_k and μ_k is

$$v_k = \frac{1}{N}\sum_{i=1}^{N} x_i^k \tag{9.32}$$

$$\mu_k = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{j=1}^{N} x_j \right)^k \tag{9.33}$$

In the general case, there is no closed form relationship between the origin moment and the center moment. However, for the first order and second order moments there are closed form expressions between them. That is

$$\mu_1 = 0 \tag{9.34}$$

$$\mu_2 = v_2 - v_1^2 \tag{9.35}$$

Eq. (9.45) can be also expressed by the data $x_1, ..., x_N$

$$E[(X - E(X))^2] = E(X^2) - (E(X))^2 \tag{9.36}$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{j=1}^{N} x_j \right)^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{j=1}^{N} x_j \right)^2 \tag{9.37}$$

We can straightforwardly prove Eq. (9.47). In fact

$$\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\frac{1}{N}\sum_{j=1}^{N}x_j\right)^2=\frac{1}{N}\sum_{i=1}^{N}x_i^2-\frac{2}{N^2}\sum_{i=1}^{N}x_i\sum_{j=1}^{N}x_j+\frac{1}{N^2}\left(\sum_{i=1}^{N}x_j\right)^2$$

$$=\frac{1}{N}\sum_{i=1}^{N}x_i^2-\left(\frac{1}{N}\sum_{j=1}^{N}x_j\right)^2$$

In testing sonar specifications, the center moment is more beneficial than the origin moment, since the application of the center moment can avoid systematical error, yielded by non-sonar factors.

If the true value of the parameter to be measured is φ, the *i*-th measurement data is $\theta_i=\phi+\mu+\Delta\theta_i$ $(i=1,\ ...,\ N)$, where μ represents the systematical testing error, $\Delta\theta_i$ is the measurement error of the sonar system. The first order origin moment of random variable θ_i is $v_1=\frac{1}{N}\sum_{i=1}^{N}\theta_i=\phi+\mu+\frac{1}{N}\sum_{i=1}^{N}\Delta\theta_i$. The estimation error is

$$v_1-\phi=\mu+\frac{1}{N}\sum_{i=1}^{N}\Delta\theta_i \tag{9.38}$$

Obviously, in Eq. (9.48), the systematical error is contained in the estimation error. If we use the center moment, then the systematical error can be deleted from the estimation error. This is because the first order of center moment is

$$\mu_1=\frac{1}{N}\sum_{i=1}^{N}\Delta\theta_i\quad \mathrm{R} \tag{9.39}$$

The second order of center moment is

$$\sqrt{\mu_2}=\left[\frac{1}{N}\sum_{i=1}^{N}\left(\theta_i-\frac{1}{N}\sum_{j=1}^{N}\theta_j\right)^2\right]^{1/2}=\left[\frac{1}{N}\sum_{i=1}^{N}\left(\Delta\theta_i-\frac{1}{N}\sum_{j=1}^{N}\Delta\theta_j\right)^2\right]^{1/2} \tag{9.40}$$

The value expressed on the left of Eq. (9.40) is called rms (root of mean square) value of estimation.

Now we turn to answer the question proposed at the beginning of this section. Suppose we have *N* measured results x_i for *N* parameters θ_i, do we make any conclusion about estimation error? Generally speaking, the answer is "No". But if we assume that x_i is the unbiased estimation of θ_i, i.e., $E(x_i)=\theta_i$ $(i=1,\ 2,\ ...,\ N)$. And the variance $\mathrm{Var}(x_i)=\sigma^2$ is independent of *i*. Then we

can find the mean square error of estimation by the following equation

$$E\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\theta_i)^2\right]=\frac{1}{N}\sum_{i=1}^{N}E(x_i-\theta_i)^2=\sigma^2 \tag{9.41}$$

Example

Suppose in a sea trial, the target incidental angles (true value) are $\theta = 0$ °, 10.3 °, 15.9 °, 25 °, 48 °, 130 °, 215.7 °, 310 °. The measured results are $x = -1.2$ °, 9.8 °, 17 °, 24.3 °, 47.8 °, 131.5 °, 309 °. If the estimation is unbiased, then the rms value of estimation in terms of Eq. (9.51) is $\Delta = 1.007$ °.

9.5.7 Sea State Description and Some Results of Sonar Ranging Prediction

The sea state description is listed in Table 9.6, which is necessary knowledge for any sonar designer[43].

Table 9.6 Sea state description

Beaufort wind scale	Wind speed (k)	Wind speed (km / hr)	Weather terms	Observed effects	Sea description	Wave heights (m)	Douglas sea state
0	<1	<1	Calm	Sea like a mirror	Calm	0	0
1	1 – 3	1 – 5	Light	Ripples	Smooth	<1	1
2	4 – 6	6 – 11	Light breeze	Small wavelet	Slight	0.3 – 1	2
3	7 – 10	12 – 19	Gentle	Wave begin to break	Moderate	1 – 2	3
4	11 – 16	20 – 28	Mode-rate	Nume-rous whitecaps	Rough	2 – 3	4
5	17 – 21	29 – 38	Fresh	Some spray	Rough	2 – 3	4
6	22 – 27	39 – 49	Strong Wind	Large waves	Rough	2 – 3	4
7	28 – 33	50 – 61	Stiff wind	White foam	Very rough	3 – 4	5
8	34 – 40	62 – 74	Stormy wind	Small Branches Broken	Very rough	3 – 4	5
9	41 – 47	75 – 88	Strong gale	Slight Structural Damage	High	4 – 6	6
10	48 – 55	89 – 102	Whole gale		Very high	6 – 12	7
11	56 – 63	103 – 117	Storm		Moun-Tainous	12	8
12	>64	>118	Hurricane		Confused		9

Some useful data appeared in the sonar equation and the performances of modern sonar are listed in Tables 9.9 – 9.10. From these data readers can get a better understanding of the real ocean and modern sonar performance[44].

Table 9.7 Array gain of AN / BQQ-5 submarine sonar of US navy

Frequency (Hz)	Spherical array (dB)	Cylindrical array (dB)	Towed array (dB)
40		6	12
200	0	10	20
500	9	15	20
1,000	15	18	20

Table 9.8 Ranging distance of AN / BQQ-5 sonar (towed array, deep sea)

TL (dB)	SL (dB)	NL (dB)	AG (dB)	DT (dB)	Addt.Loss (dB)	Distance (km)
30 Hz						
106	140	55	12	–14	5	>1000
86	120	55	12	–14	5	180 – 210
66	100	55	12	–14	5	1 – 2
300 Hz						
109	130	50	20	–14	5	>500
89	110	50	20	–14	5	55 – 65
69	90	50	20	–14	5	1 – 2

Note: Addt.loss means additional loss.

Table 9.9 Ranging distance of towed array sonar (shallow sea)

TL (dB)	SL (dB)	NL (dB)	AG (dB)	DT (dB)	Addt.Loss (dB)	Distance (km)
30 Hz						
91	140	70	12	–14	5	<100
71	120	70	12	–14	5	3 – 10
51	100	70	12	–14	5	0.3 – 0.5
300 Hz						
99	130	60	20	–14	5	<100
79	110	60	20	–14	5	3 – 30
59	90	60	20	–14	5	1 – 5

Table 9.10 Ranging distance of fixed shore-based sonar (the detection distance for submarines)

TL (dB)	SL (dB)	NL (dB)	AG (dB)	DT (dB)	Addt.Loss (dB)	Distance (km)
30 Hz						
82	140	70	3	–14	5	5 – 40
62	120	70	3	–14	5	2 – 8
42	100	70	3	–14	5	0.1 – 0.3
300 Hz						
92	130	60	12	–14	5	<100
72	110	60	12	–14	5	<10
52	90	60	12	–14	5	0.3 – 0.5

References

[1] MatLab: http://www.mathworks.com. Accessed 1 Jan 2011
[2] Etter, D. H.: Engineering Problem Solving with ANSIC. Prentice Hall, New Jersey (1995)
[3] Etter, D. H.: Engineering Problem Solving with Matlab, 2nd Edn. Prentice Hall, New Jersey (1997)
[4] Matlab Front-End for Acoustic Toolbox, URL: http://www.curtin.edu.au/curtin/centre/cmst/products/actoolbox/. Accessed 1 Jan 2011
[5] Ritchie, D. M., Thompson, K.: Unix time-sharing system : the C programming language. Bell System Tech. J., 57-6, pp. 1191-2019 (1978)
[6] Kernighan, B. W., Ritchie, D.: The C Programming Language. Prentice Hall, New Jersey (1978)
[7] Lapsley, P., Bior, J.: Choosing DSP processor. DSP Applications, 3-1, pp. 24-32 (1994)
[8] Davis, A.: DSPs and the age of specification. DSP and Multimedia Technology, 3-7, pp. 20-32 (1994)
[9] Kung, S. Y.: VLSI and Modern Signal Processing. Prentice Hall, New Jersey (1985)
[10] Bier, J.: Selecting the right tools for DSP processor software development. DSP and Multimedia Technology, 3-7, pp. 55-62 (1994)
[11] Daly, P.: How to make RAM and KRAKEN Agree in Range-Dependent Nevironment, http://www.mit.edu. Accessed 1 Jan 2011
[12] Acoustic Toolbox, URL: http://stommel.tamu.edu/~baum/linuxlist/linuxlist/node7. html#AcousticsToolbox. Accessed 1 Jan 2011
[13] Porter, M. B., Reiss, E. L.: A numerical method for ocean-acoustic normal modes. J. Acoust. Soc. Amer., 76, pp. 244-252 (1984)
[14] Porter, M. B., Reiss, E. L.: A numerical method for bottom interacting ocean acoustic normal modes. J. Acoust. Soc. Amer., 77, pp. 1760-1767 (1985)
[15] Porter, M. B.: The KRAKEN normal mode program. Rep. SM-245, SACLANTSEN, La Spezia (1991)
[16] Sullivan, E. J., Edelson, G. S.: Model-based broadband towed array processing. In: 147th meeting of the Acoustical Society of America, New York (2004)
[17] Sullivan, E. J., Candy, J. V.: Space-time array processing: the model-based approach. J. Acoust. Soc. Amer., 102-1, pp. 2809-2820 (1997)
[18] Candy, J. V., Sullivan, E. J.: Passive localization in ocean acoustics: a model based approach. J. Acoust. Soc. Amer., 98-3, pp. 1455-1471 (1995)
[19] Brown, G. J., Wrigley, S. N.: Feasibility study into the application of computational auditory scene analysis techniques to sonar signals. Technical Report, University of Sheffield, Department of Computer Science (2000)
[20] Baldacci, A., Haralabus, G.: Signal processing for an active sonar system

suitable for advanced sensor technology applications and environmental adoption schemes. In: Proc. of EUSIPCO'2006, Florence Italy (2006)
[21] Boashash, B., O'Shea, P.: A methodology for detection and classification of some underwater acoustic signals using time-frequency analysis techniques. IEEE Trans. on Acoustics, Speech, and Signal Processing, 38-11, pp. 1829-1841 (1990)
[22] Bregman, A. S.: Auditory Scene Analysis: the Perceptual Organization of Sound. The MIT Press, London (1990)
[23] Moon, T. K.: Similarity methods in signal processing. IEEE Trans., SP-44-4, pp. 827-883 (1996)
[24] Spiess, F., Kuperman, W.: The marine physical laboratory at Scripps. Oceanography, 16-3, pp. 45-49 (2003)
[25] Scripps Institute of Oceanography: http:// www.scripps.edu. Accessed 1 Jan 2011
[26] Woods Hole Oceanographic Institution: http://www.whoi.edu. Accessed 1 Jan 2011
[27] Wocester, P. F., Spindel, R. C.: North Pacific Acoustic Laboratory, http://www. npal.ucsd. Edu. Accessed 1 Jan 2011
[28] Office of Naval Research: http://www.onr.navy.mil. Accessed 1 Jan 2011
[29] Burns, R. F.: The naval research laboratory. Sea Technology, 34-11, pp. 66-77 (1993)
[30] Curtin, T. R.: ONR program in underwater acoustic communications. Sea Technology, 40-5, pp. 17-27 (1999)
[31] Porter, M. B., Reiss, E. L.: A numerical method for ocean-acoustic normal modes. J. Acoust. Soc. Amer., 76, pp. 244-252 (1984)
[32] Porter, M. B., Reiss,E. L.: A numerical method for bottom interacting ocean acoustic normal modes. J. Acoust. Soc. Amer., 77, pp. 1760-1767 (1985)
[33] Porter, M. B.: The KRAKEN normal mode program. Rep. SM-245, SACLANTSEN, La Spezia (1991)
[34] Scott, R.: Offboard countermeasure technology. Naval Forces, XV-4, pp. 16-22 (1994)
[35] National Research Council of USA (Eds.): The Role of Experimentation in Building Future Naval Forces. NA Press, Washington (2002)
[36] Kennedy, F. D. Jr.: Experimentation: the key to transformation. Undersea Warfare, 5-1, pp. 3-10 (2002)
[37] Carey, W., Dillman, L. M..: Shallow water transmission measurements taken on the New Jersey continental shelf. J. Acoust. Soc. Amer., 89, pp. 1981-1991 (1986)
[38] Tobin, P. E.: Year of the ocean: a chance to consider our own role. Sea Technology, 39-1, pp. 10-14 (1998)
[39] Hamblen, W.: Next generation of stealth submarine. Sea Technology, 39-11, pp. 59-62 (1998)
[40] Douglass, J. W.: Undersea dominance: don't leave home without it. Sea

Technology, 37-11, pp. 10-16 (1996)
[41] http://www.tods.com/defense. Accessed 1 Jan 2011
[42] Jenkins, J. W.: Sonar principles and antisubmarine warfare. Sea Technology, 34-2, pp. 61-71 (1993)
[43] Bradley, M.: Environment Acoustics Pocket Handbook, 2nd Edn. PSI McLean, VA (1996)
[44] Miasnikov, E.: Can Russian strategic submarine survive at sea? the fundamental limits of passive acoustics, Science & Global Security, 4, pp. 213-251 (1994)

Index

A
A-law, 29, 345, 346
A/D conversion, 294, 298, 300
Absolute value detector, 267
Absolute value detection, 332, 333
Acceleration sensitivity of hydrophone, 539
Acoustic mode, 209
Acoustic module of towed array sonar, 536
Active sonar signal capture, 297, 446
Adama matrix transform, 461
Adaptive beamforming (ABF), 409, 412
Adaptive canceling of self noise, 544
Adaptive equalization, 455
Adaptive filtering, 126
Adaptive line enhancer, 131, 317
Adaptive noise canceling (ANC), 287, 501
Adiabatic equation, 203
Advanced Deployable System, 577
Along-track resolution ability, 563
Ambient noise, 169
Ambient noise level, 177
Amplitude modulation signal, 321, 454,
Amplitude quantization, 26
Amplitude Shift Keying (ASK), 459
Anti-aliasing filtering, 20
Arc array, 242
Arithmetical central frequency, 293
Array baffling, 226
Array signal generator, 504
Artificial neural network, 436, 438
Asynchronized sampling, 609
Attractor, 71
Auto-correlation function, 43
Auto-power spectrum function, 120
Automatic gain control (AGC), 298
Automatic target tracking, 307, 399
Axiom of probability, 84

B
Background equalization technique, 339
Back propagation model, 438
Basic reliability, 166
Batch processing, 286, 287
Bayes estimate, 138, 140
Beam data extension, 400
Beamforming, 7, 8, 226
Beamforming of multiple elements with different SNR, 549
Beamforming in frequency domain, 287
Beamforming in time domain, 287
Beam pattern, 226
Beam width of main lobe, 229, 241
Bearing accuracy of sonar, 380
Bearing ambiguity, 533
Bernoulli binomial distribution, 87
Bernoulli large number law, 93
Binary test, 106
Black box principle, 438
Bottom reverberation, 193
Bow sonar, 217, 242
Broadside direction, 233

Buffon's needle experiment, 481
Built in failure self test, 359

C
Calibration of sonar specification, 621
Capture threshold, 452
Cauchy series, 68
Central limit theorem in probability theory, 176
Central moment, 627
Channel capacity, 461
Characteristics of model based sonar, 521
Chaos theory, 67
Chebychev inequality, 624
Chebychev large number law, 93
Chirp-Scaling algorithm, 569
Circle array, 226
Circle plate array, 258
C language, 599
Classic definition of probability, 83
Classify before detection, 369
Cluster analysis, 437
Coding theorem, 103, 104
Combined architecture of ANC and ALE, 328
Companding, 28, 29
Conditional entropy, 99
Conditional probability, 85
Confidence level, 424
Conformal array, 543
Consistent estimate, 139
Contract mapping, 70
Conventional beamforming, 230
Convolution, 37
Coordinate rotation digital computer (CORDIC), 414
Cosine transform (DCT), 60
Cramer-Rao low bound, 142
Criterion of maximum SNR, 109
Criterion of minimum total error probability, 108
Critical angle towing, 536
Cross-correlation function, 47
Cross power spectrum, 41, 117, 118, 382
Cross track resolution ability, 563
Cylindrical and spherical model of ambient noise, 181
Cylindrical array, 6, 226, 257

D
Data fusion, 361
Data interpolation, 347
Data sieve, 424, 425
Data fusion, 361
Decorrelation time delay, 326
De-coupled material, 548
Delay line time compressor (DELTIC), 4
Demodulation on noise (DEMON), 184, 317
Dependent observations, 363
De-sampling, 296, 514
Detecting distance, 164. 398
Detection index, 163
Detection probability, 108
Deterministic signal, 20, 23
Digital filter, 32
Dipole-element hydrophone, 534
Discrete cosine transform (DCT), 60
Discrete Fourier transform (DFT), 43, 44
Diver detection sonar (DDS), 1, 580
Digital Interference Canceling Adaptive Network Nulling Equipment (DICANNE), 4, 314
Digital multi –beam system (DIMUS), 4
Digital signal processing (DSP), 1, 4, 19
Direct current jump, 264
Directivity function, 226
Directivity index, 229
Discrete Fourier transforms (DFT), 44
Distribution function of random variable, 87, 92
Diver detection sonar (DDS), 580
Dolph-Chebychev weighting, 248, 250
Double check decision strategy, 422
Dry end (of sonar system), 5, 6, 7
DSP (Digital signal processing) chip, 1, 8

E
Echo ranging, 154
Effective estimate, 140
End fire direction, 233
Energy detector, 112
Entropy, 10, 97
Entropy of information source, 98
Equal spaced line array, 232
Equipment module, 526
Ergodic random signal, 42, 46
Ergodic theorem of correlation function, 95
Ergodic theorem of mean value, 95
Estimation of auto-correlation function, 146
Estimation of signal parameters via rotational invariant Techniques (ESPRIT), 413
Expert system, 437, 440, 441

F
Failure rate function, 352, 353
False alarm probability, 108
Fast Fourier Transform (FFT), 4, 47
Feature level of data fusion, 361
Figure Of Merit (FOM), 170
Finite Impulse Response (FIR), 8, 452, 486
Five-point parabolic interpolation, 390
Flow noise, 191, 535
Flank array sonar, 3, 154, 543
Four point passive ranging, 430
Fractal transforms, 67
Frequency spread technique, 461
Frequency modulation, 461
Frequency shift keying (FSK), 460
Full automatic tracking, 400

G
Gaussian bell shape impulse, 454
Gaussian distribution, 87, 88
Gaussian random signal, 27
Geometry definition of probability, 83
General purpose sonar simulator, 504, 505
Generating a Gaussian random number, 487
Generation function, 75
Geometry center frequency, 608
Global positioning system (GPS), 456
Grazing angle analysis, 422
Grazing angle, 210, 421, 422, 426
Grey scale conversion (GSC), 297, 344

H
Halftone technique, 65
Hamming window, 307
Hanning window, 307
Hardware reliability, 351
Harvard architecture, 602
Hausdorff distance, 70
Hawaii-2 observatory, 577
Heisenberg uncertainty principle, 74
Height of side lobe, 229, 237, 241
Homogeneous and isotropic noise field, 178
Homogeneous sound field, 545
Hopfield/Kohonen model, 438
Hull mounted sonar, 189, 211
Human/computer interface, 341
Human/machine interface, 5

I
Independent observations, 363
Index sound level, 164
Infinite Impulse Response (IIR), 14
Infinite non-cyclic random number, 507
Information superiority, 573
Invariant, 33
Interception sonar, 6
Interpolation technique, 508
Intrude detection sonar (IDS), 582
Inverse beamforming (IBF), 541
Isothermal layer profile, 205
Iterated Function System (IFS), 71
Iterative algorithm of adaptive noise canceling (ANC), 501

J
Joint detection and estimation, 369

K
Kalman filtering, 132
Khinchine large number law, 93
Kinetic equation of the sound wave, 545

L
Laboratory testing of sonar systems, 593
Lagrange multiplier method, 363
Latitudinal wave, 202
Learning curve of adaptive filtering, 130, 502
Left beam, 381
Left/right ambiguity, 527, 533, 534
Left sub-array, 381
Lenna's photo, 63
Likelihood ratio, 107, 109
Limited function associate testing, 612
Line array, 226
Linear digital filter, 32
Linear frequency modulation signal, 454
Linear towed array sonar, 523
Line spectrum components, 547
Local average filtering, 51
Local area network (LAN), 51
Low frequency analysis record (LOFAR), 184, 317

M
Main lobe, 229
Maintainability, 154, 165, 298
Matched filtering, 116
Matlab software, 594
Maximum à posterior probability (MAP) test criteria, 106
Maximum entropy, 104
Maximum likelihood estimation, 142
Maximum likelihood ratio test, 111
Maximum posterior estimate, 142
Matched field processing (MFP), 519, 522,
Mean Time Between Critical Failure (MTBCF), 166, 298
Mean Time Between Failures (MTBF), 166, 356
Mean Time To Failure (MTTF), 166, 352, 353
Mean Time To Repair (MTTR), 166, 352, 353
Mean value estimate, 144
Mean value filtering, 55, 56, 338
Median value estimation, 141
Median value filter, 54
Mega Floating Operation Per Second (MFLOPS), 603
Mega Instruction Per Second (MIPS), 603
Mega Operation Per Second (MOPS), 603
Mellen formula, 213
Metcalfe's law, 574, 578
Method for generating a random number, 482
Metric space, 68
Micro-disturb method, 121
Mine avoidance sonar, 154
Mine hunting sonar, 154
Minimum ambient noise, 177
Minimum mean square error (MSE), 109
Model based processing (MBP), 519
Model based sonar, 517, 518, 521
Model-free estimates, 145
Model mismatch, 517
Mode theory, 522
Modified Bessel function of zero order, 200
Module 1/2 addition, 507
Monte Carlo method, 480, 482
Moore's law, 10, 11
Moving average, 337
M series, 447
μ-law, 29, 345, 346
Multi-layer sensing model, 438
Multi path, 464
Multiple independent noise simulation, 495
Multi-target resolution, 307, 398, 404
Multiple signal classification (MUSIC), 413

N
Nearest neighborhood criteria, 441
Negative gradient, 205
Negative thermal cline, 204, 221
Network Centric Warfare (NCW), 1, 573
Neyman-Pearson criterion, 109
Non-linear shift register, 512
Non-recursive filter, 37
Nyquist sampling frequency, 23

O
Omni-directional transducer, 202
Optimum detection criterion, 108
Optimum linear data fusion, 363
Optimum spatial filter, 307, 311
Optimum transfer function, 122
Optimum weighting coefficients, 363
Order truncate average (OTA), 347
Origin moment, 164, 628
Orthogonal principle in parameter estimation, 134
Own Doppler null, 7

P
Parabolic approximation of wave propagation, 522
Parabolic interpolation, 389
Parallel operation, 286
Parameter estimation of random signal, 138
Parseval equality, 42
Passive ranging sonar (PRS), 155, 429
Performance evaluation, 593
Personal Computer Interface (PCI), 14
Phase Shift Keying (PSK), 460
Platform noise, 183
Point to point processing, 62, 339
Point-wise imaging algorithm, 569
Polyvinylidene fluoride (PVDF) material, 548
Positive gradient, 204
Post-processing, 307
Power density spectrum, 25, 145
Pre-processing, 298, 608
Pre-setting tracking, 415
Probability density function, 20
Product theorem of directivity, 244
Programmable digital sonar, 292
Proposition of Small probability event, 625

Q
Quantization error of amplitude, 499
Quantization error of time delay, 499
Quantization noise, 31, 302
Quiet submarine, 518

R
Rader's algorithm, 508
Radiated noise, 162, 183
Radiated noise of torpedo, 188
Random variable, 86
Range-Doppler (R-D) algorithm, 569
Ranging distance of sonar, 222
Rayleigh distribution, 92
Ray theory, 207, 522
Receiver operating curve (ROC), 111
Recognition coefficient, 164
Recognition index, 169
Rectangle window function, 306
Rectangular plate array, 258, 259
Recursive filter, 37
Recursive moving average, 52, 53
Redundancy design, 358
Rejection response of the array filter, 312
Relaxation frequency, 213
Reliability, 165
Resolution ability of the sonar, 398
Resolution invariant principle of SAS, 564
Reverberation dynamic range, 202
Reverberation level, 194
Rice distribution, 200
Right beam, 381
Right sub-array, 381
Rigid design, 359
Ring array, 258
Robust signal processing, 520

S
Sampling duration, 21
Sampling frequency, 22
Sampling theorem in frequency domain, 24
Saturation noise, 302
Schwartz inequality, 91, 123
Second accumulation, 391
Seeds of the random number, 512
Self mapping model, 438
Semi-automatic tracking, 401
Sequential operation, 286
Sequential test, 115
Shape distortion of towed array, 530
Shore based sonar station, 553
Short Time Fourier transform (STFT), 73
Side lobe, 229
Side scan sonar, 154
Σ-Δ modulation, 30
Signal to noise ratio (SNR), 259
Signal to noise ratio of quantization (SQNR), 27
Simulation of post processing, 502
Similar weighting, 251
Simulation of reverberation, 496
Simulation of the sonar environment field, 493
Single rectangle wave, 453
Small probability event, 625
Snell's law, 209
Software reliability, 358
Sonobuoy, 153
Sonar decoy, 153
Sonar equation, 151
Sound pressure, 159
Sonar range prediction, 265
Sonar signal simulator, 503
Sound Surveillance Sonar (SOSUS), 553
Sound speed profile (SSP), 204
Source level, 160
Space gain of the sonar system, 168
Spatial correlation radius, 231
Spatial filter, 226
Spectrum integration, 522
Spectrum level, 160
Spectrum spread, 200
Spherical array, 226
Square detector, 262
Standard hydrophone, 615
Stationary random process, 94
Stationary random signal, 25, 116
Statistical definition of probability, 84
Statistical hypothesis testing, 106
Stealth submarine, 523
Superfine time delay signal, 514
Super Harvard architecture, 603
Surface reverberation, 193
Surveillance Towed Array Sensor System (SURTASS), 524
Swimmer detection sonar (SDS), 582
Synchronized sampling, 609
Synthetic Aperture Sonar (SAS), 559
System simulation technique, 479

T
Tapped delay line, 326
Target classification and identification, 298, 441
Target strength, 422, 618
Target strength change analysis, 422
Target trace extraction technique, 58
Task reliability, 166
Target moving analysis (TMA), 413
Target strength, 169
Testing in lake, 614
Testing at sea, 614
Theory of complexity system, 575
Three level of data fusion, 361
Three-point parabolic interpolations, 389
Three point ranging, 427
Threshold filtering, 56
Time delay estimation, 427
Time/frequency domain signal capture technique, 448
Time gain of the sonar system, 169
Time invariant (filter), 33
Time reversal mirror, 4

Time Varying Gain (TVG), 197
Tone frequency, 184
Torpedo alarming, 287, 420
Tow cable, 526
Towed array sonar, 184, 524
Trace correlation technique, 340
Trace identifying algorithm, 425
Tracking distance, 167
Transmitting source level, 166
Transversal wave, 203
Triangle window, 307
True bearing angle calibration, 394
Truncate filtering, 56
Transfer function, 34
Transmission loss, 169
Twin linear array system, 534
Two dimensional Fourier transform, 569

U
Unbiased estimation, 629
Underwater acoustic channel, 455
Underwater acoustic communication, 455
Underwater monitoring network, 579
Uniform distribution, 87
Unit impulse response, 33
Upsampling, 296

V
Variable depth sonar (VDS), 3, 154
Variance of random variable, 89
Versa Module Europe (VME), 14
Vertical directivity of ambient noise, 182
Vibration isolated module, 7, 542
Virtual array, 564
Volume reverberation, 193

W
ω-k Algorithm in wave number domain, 569
Wald sequential test, 115
Wavelet transform, 67, 72
Wave theory, 207
Wet end, 2, 7
White noise, 162
Wide Aperture Array (WAA), 435
Widrow's algorithm, 130
Wiener filtering, 116
Wiener-Hopf Equation, 121
Wiener-Khinchine theorem, 43
Wild value filtering, 332
Winch, 525

Z
Zero order Bessel function, 388