

Mathematics for Industry 3

Mikio Tohyama

# Waveform Analysis of Sound

 Springer

# **Mathematics for Industry**

Volume 3

*Editor-in-Chief*

Masato Wakayama (Kyushu University, Japan)

**Scientific Board Members**

Robert S. Anderssen (Commonwealth Scientific and Industrial Research Organisation, Australia)

Heinz H. Bauschke (The University of British Columbia, Canada)

Philip Broadbridge (La Trobe University, Australia)

Jin Cheng (Fudan University, China)

Monique Chyba (University of Hawaii at Mānoa, USA)

Georges-Henri Cottet (Joseph Fourier University, France)

José Alberto Cuminato (University of São Paulo, Brazil)

Shin-ichiro Ei (Hokkaido University, Japan)

Yasuhide Fukumoto (Kyushu University, Japan)

Jonathan R.M. Hosking (IBM T.J. Watson Research Center, USA)

Alejandro Jofré (University of Chile, Chile)

Kerry Landman (The University of Melbourne, Australia)

Robert McKibbin (Massey University, New Zealand)

Geoff Mercer (Australian National University, Australia) (Deceased, 2014)

Andrea Parmeggiani (University of Montpellier 2, France)

Jill Pipher (Brown University, USA)

Konrad Polthier (Free University of Berlin, Germany)

Osamu Saeki (Kyushu University, Japan)

Wil Schilders (Eindhoven University of Technology, The Netherlands)

Zuowei Shen (National University of Singapore, Singapore)

Kim-Chuan Toh (National University of Singapore, Singapore)

Evgeny Verbitskiy (Leiden University, The Netherlands)

Nakahiro Yoshida (University of Tokyo, Japan)

**Aims & Scope**

The meaning of “Mathematics for Industry” (sometimes abbreviated as MI or MfI) is different from that of “Mathematics in Industry” (or of “Industrial Mathematics”). The latter is restrictive: it tends to be identified with the actual mathematics that specifically arises in the daily management and operation of manufacturing. The former, however, denotes a new research field in mathematics that may serve as a foundation for creating future technologies. This concept was born from the integration and reorganization of pure and applied mathematics in the present day into a fluid and versatile form capable of stimulating awareness of the importance of mathematics in industry, as well as responding to the needs of industrial technologies. The history of this integration and reorganization indicates that this basic idea will someday find increasing utility. Mathematics can be a key technology in modern society.

The series aims to promote this trend by (1) providing comprehensive content on applications of mathematics, especially to industry technologies via various types of scientific research, (2) introducing basic, useful, necessary and crucial knowledge for several applications through concrete subjects, and (3) introducing new research results and developments for applications of mathematics in the real world. These points may provide the basis for opening a new mathematics-oriented technological world and even new research fields of mathematics.

More information about this series at <http://www.springer.com/series/13254>

Mikio Tohyama

# Waveform Analysis of Sound



Springer

Mikio Tohyama  
Tokyo, Japan

ISSN 2198-350X                   ISSN 2198-3518 (electronic)  
ISBN 978-4-431-54423-4       ISBN 978-4-431-54424-1 (eBook)  
DOI 10.1007/978-4-431-54424-1  
Springer Tokyo Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014955147

© Springer Japan 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, 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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

## About the Author

Mikio Tohyama holds the Dr. Eng. from Waseda University in Tokyo, Japan. Dr. Tohyama was involved in research projects in acoustics and signal processing at the NTT (Nippon Telegraph and Telephone) research laboratories for 18 years beginning in 1975. He was a professor at Kogakuin University (1993–2003) and



Waseda University (2003–2011) in sound and perception. Since 2012 he has operated his research consulting firm, Wave Science Study (WSS). His present interest is sound signature expression in the temporal domain oriented to sound perception. Dr. Tohyama enjoys playing the piano every day.

E-mail address: m\_tohyama@fuji.waseda.jp

# Preface

What is this sound? What does that sound indicate? These are two questions frequently heard in daily conversation. Sound results from the vibrations of elastic media, and in daily life provide informative signals of events happening in the surrounding environment. In interpreting auditory sensations, the human ear seems particularly good at extracting the signal signatures from sound waves. Although exploring auditory processing schemes may be beyond our capabilities, source signature analysis is a very attractive area in which signal processing schemes can be developed using mathematical expressions.

This book is inspired by such processing schemes, and oriented to signature analysis of waveforms. Most of the examples in this book are taken from data of sound and vibrations; however, the methods and theories are mostly formulated using mathematical expressions rather than by acoustical interpretation. This book might therefore be attractive and informative for scientists, engineers, researchers, and graduate students, who are interested in the mathematical representation of signals and the applications of Fourier analysis.

The book can be described as being nearly self-contained, but does assume readers to be familiar with introductory topics in discrete signal processing as in the discrete Fourier transform. Hence this book might be also usable as a textbook for use in graduate courses in applied mathematics on topics as in complex functions.

Almost all scientific phenomena are sensed as waves propagating in some space. Over the years, waveform analysis has therefore been one of the resilient academic areas of study, and still is seen as fertile ground for development. In particular, waveform analysis based on the theory of linear systems would be a good example where a physical interpretation can be given to the mathematical theory of complex functions in terms of magnitude, angle, poles, and zeros of complex functions. Readers who are interested in the physical aspects of sound and vibration data or elementary formulation of wave equations and their solutions are recommended the book, M. Tohyama, *Sound and Signals*, Springer 2011, which can be used as a complementary companion to this book or a good reference.

This book is organized as follows.

The sinusoidal representation of discrete signals dealt with in this book is fundamental, and still newly developing from both a theoretical and a practical perspective in signal analysis. Sinusoidal functions are the conceptual roots from which most, if not all, signals are generated. Therefore sinusoids are inescapable when analyzing linear systems. In contrast, discrete Fourier analysis is formulated to develop representations of discrete sequences of finite record length, assuming sequences can be periodic extended outside the finite-length interval. Consequently, extrapolation (or prediction) even for a sinusoidal sequence is problematic when a sequence observed in a finite interval is represented by its Fourier transform. From the perspective of theory and practice, this issue in Fourier analysis typically motivates the introduction of new schemes for signal processing.

Following an elementary introduction into the manipulation of discrete sequences and polynomials, this book is mostly devoted to the analysis and representation of compound sinusoidal sequences and their related polynomials. The compound sinusoidal functions are representatives of auditory sound waveforms, which range from periodic to almost periodic waves. The almost periodic functions might form the conceptual basis for discrete analysis of sequences; however, the functions would be suitable in representing non-periodic waves in a deterministic sense. This book develops spectral peak selection from the interpolated spectral function so that the spectral components might be estimated, and thus a finite-length sequence could be extrapolated outside its interval of observation.

Signal analysis as in power spectral analysis is mostly performed in the frequency domain. Nevertheless, signal signatures in the temporal domain as well as those in the frequency domain are dealt with in this book. Narrow-band temporal envelopes and auto-correlation functions characterize the signal signatures in both these domains. The envelopes are sensitive to phase spectral properties and spectral correlations in the frequency plane rather than the power spectral functions, whereas the auto-correlation functions are essentially independent of phase properties; however, both are functions of time and frequency.

This book takes a frame-wise approach to the time domain instead of the usual mathematical formulation of the temporal envelopes and auto-correlation functions of the time and frequency. The reason is that an open question remains at present as to which windowing (or framing) functions are adequate in making the functions informative. Triangular windowing sequences, developed for calculations of group-delay functions without explicitly including the derivatives, are applied to the frame-wise approach. The frame-wise auto-correlation analysis of a reverberant speech sample would be a good example that illustrates why a speech sample could be still recognizable even under adverse reverberant conditions. The example illustrates the benefits of the frame-wise approach to the time domain in preference to the long-term analysis in which temporal information is lost.

The envelopes are decomposed into clustered line-spectral components around a central frequency. The clustered components are too closely distributed (with a very narrow spacing) over the frequency interval to be separately picked up by spectral peak selection. This indicates that a sinusoidal compound sequence

characterized by the envelope might not be well represented by such a selection. This book develops clustered line-spectral modeling, called CLSM for short, so that the sequence, including its envelope, can be extrapolated outside the interval of observation. CLSM is formulated using the least-squares error solution of a set of simultaneous equations over the frequency domain.

The complementarity between complex time and frequency planes indicates interesting properties stemming from the zeros for a time sequence in the complex time planes, as well as the zeros of the transfer function (or  $z$ -transform of a time sequence) in the complex frequency domain. The zeros of the complex time plane gives a clear interpretation of the relationship between the instantaneous magnitude (or the envelope) and the instantaneous frequency (or the frequency of the carrier), as zeros in the complex frequency domain identify the minimum-phase property for the spectral function. The relationship between the magnitude and phase properties that is expected for minimum-phase systems might be formally realized between the temporal envelope and carrier frequency for minimum instantaneous-phase systems.

The methods of analysis are based on the complex complementarity between the time and frequency planes where the signals are represented using sinusoidal functions or discrete sequences. Following the complex complementarity, the analytic signals can be interpreted in terms of the complex spectral properties, and, clustered time-sequence modeling (CTSM) is developed based on the complementary with CLSM. In contrast to CLSM which characterizes the slowly changing temporal envelopes in the time domain, CTSM identifies pulse-like but frequency-band limited waveforms (sequences) by solving the least-squares error solution in the time domain. The Fourier transform of a sequence of finite record length is also a sinusoidal compound function on the frequency plane. Therefore frequency-band limiting of the sequence is formally equivalent to narrowing the observation interval of the compound sinusoidal function on the time domain.

As understood through the spectral peak selection, the peaks identify resonant or eigenfrequencies, the spectral functions corresponding to those eigenfrequencies, and the effect of poles. However, troughs and zeros are also significant signatures that characterize the phase properties of spectral functions, in particular, short waveforms as in pulse-like sound waves. Troughs or zeros of the spectral functions are strongly sensitive on the waveforms of brief sequences, even if the envelopes of the magnitude spectral functions look more or less identical. Indeed, the residual of the CLSM (modeling error) for piano-string vibration yields a brief impulsive sequence corresponding to the source waveform excited by the action of the hammer on the piano string. It is an example of how the zero or trough of the spectral function uniquely characterizes the source waveform by separating the characteristics of the transfer function.

The zeros are formally identified from the factorization of the  $z$ -transform given as a polynomial of finite order, subject to the constraint that the sequence has finite length in the temporal domain. Every zero is, in principle, identified by a pair of adjacent time pulses composed of a single unit pulse followed by a single pulse with a one-unit sample of delay and complex magnitude. The time sequence can be formally reconstructed by convolution of all the pairs of adjacent time pulses.

Inspired by polynomial factorization, adjacent pairing time-pulse modeling (APTM) is developed in this book. The pairs of time pulses can be estimated by deriving iteratively the least-squares-error solutions in a localized range around each spectral trough. The solving procedure is similarly repeated in spectral peak selections until the dominant spectral troughs are removed. Examples of CLSM and APTM applied in the analysis of piano-string vibrations confirm that CLSM and APTM are possible approaches in extracting source effects in the response through a linear system by separating its properties. Extraction or separation schemes of source and path information from the response would be desirable for systems monitoring and diagnostics. As described above, this book provides a theoretical basis and introduction to waveform analysis that would cover much that is necessary in wave science and engineering including acoustics from a signal theoretic point of view.

The author thanks Satoru Gotoh, Yoshihumi Hara and Tomomi Hasegawa for their intensive research cooperation and assistance in preparing drawings. In particular, the author gratefully thanks Satoru Gotoh, Yoshihumi Hara, and Yuta Enomoto for their great effort in preparing figures and setting the Tex manuscript with finalized drawings. The author also thanks Edanz Group Japan Co. Ltd. for checking the author's written English.

The author acknowledges that this book is inspired and motivated by the very kind and responsive editorship of Yuko Sumino of Springer (Japan). The author greatly appreciates the suggestions and informative discussions of Masato Wakayama (Professor, University of Kyushu, Japan, and Editor-in-chief, Mathematics for industry) during the preparation of this book. Finally, the author extends his appreciation to all authors of research articles referred to in this book.

Tokyo, Japan  
August 2014

Mikio Tohyama

# Contents

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
	References .....	10
<b>2</b>	<b>Discrete Sequences and Their Fourier Transform .....</b>	<b>13</b>
2.1	Function and Sequence .....	13
2.2	Generating Functions and $z$ -Transforms .....	13
2.2.1	Convolution and Generating Function .....	13
2.2.2	Correlation and Generating Function .....	15
2.2.3	$z$ -Transformation of Discrete Sequence .....	16
2.2.4	Convergence of $z$ -Transformation .....	16
2.2.5	Inverse $z$ -Transformation .....	17
2.3	Zeros, Poles, and Unstable Poles .....	18
2.3.1	Zeros of Polynomial and Poles of Series .....	18
2.3.2	Unstable Pole .....	20
2.4	Fourier Transform .....	22
2.4.1	Fourier Transform of Discrete Sequence .....	22
2.4.2	Fourier Transform of Real Causal Sequence .....	23
2.4.3	Fourier Transform for an Analytic Sequence .....	25
2.4.4	Exponential Windowing and $z$ -Transform .....	27
	References .....	27
<b>3</b>	<b>Temporal and Spectral Characteristics of Discrete Sequence .....</b>	<b>29</b>
3.1	Magnitude and Phase of Spectrum .....	29
3.1.1	Magnitude and Phase Spectrum of Real Sequence .....	29
3.1.2	Power Spectral Density and Auto-correlation Sequence .....	30
3.1.3	Flat Magnitude Spectral Function and Its Phase Correlation .....	33
3.2	Frame-Wise Fourier Transform and Filter Bank .....	36
3.2.1	Frame-Wise Fourier Transform .....	36
3.2.2	Sub-band Filters .....	39
3.2.3	Perfect Reconstruction Filters .....	39
3.3	Modulation Envelope and Group Delay .....	42
3.3.1	Sinusoidal Modulation and Envelope .....	42

3.3.2	Group Delay for Real Discrete Sequence.....	46
3.3.3	Triangle Windows and Spectral Derivatives.....	48
3.4	Triangular Windowing and Auto-correlation Sequence.....	49
3.4.1	Short-Term Auto-correlation Sequence.....	49
3.4.2	One-Side Auto-correlation Sequence and Envelope of Power Spectrum .....	52
	References .....	55
<b>4</b>	<b>Temporal and Spectral Enhancement by Sound Path .....</b>	<b>57</b>
4.1	Steady-State Response of Sound Path in Rooms and Early Reflection	57
4.1.1	Impulse Response and Frequency Characteristics .....	57
4.1.2	Direct and Early Reflected Sounds.....	59
4.1.3	Auto-correlation of Direct Sound Followed by a Single Reflection .....	60
4.1.4	Temporal and Spectral Representation of Transmission Sound .....	61
4.2	Source Effects on Sound Path .....	63
4.2.1	Frame-Wise (Short-Term) Auto-correlation of Transmission Sound .....	63
4.2.2	Source Effect and Distance to Receiver .....	66
4.2.3	Auto-correlation of Impulse Response and Distance from Source .....	68
4.2.4	Envelope Energy and Distance from Sound Source.....	70
4.3	Decay of Reverberation .....	74
4.3.1	Reverberation Decay Curve of Sound Waves in Rooms .....	74
4.3.2	Decay Curves and Distance from Sound Source .....	75
	References .....	76
<b>5</b>	<b>Modulation and Periodic Properties of Temporal Envelope .....</b>	<b>77</b>
5.1	Modulation Spectrum of Envelope .....	77
5.1.1	Magnitude Spectrum and Modulation Index for Speech Envelope .....	77
5.1.2	Phase Spectrum, Cross-Correlation, and Modulation Index for Envelopes .....	79
5.1.3	Complex Modulation Transfer Function.....	80
5.1.4	Temporal and Spatial Range of Direct Sound Envelope .....	81
5.2	Narrow-Band Envelopes and Speech Intelligibility .....	82
5.2.1	Envelope Recovery and Speech Intelligibility.....	82
5.2.2	Phase Effect on Speech Intelligibility for Time-Reversed Speech .....	84
5.3	Fundamental Frequency and Period of Envelopes.....	86
5.3.1	Auto-correlation of Sequence and Period Estimation .....	86
5.3.2	Missing Fundamental .....	89
5.3.3	Period of Envelope .....	90
	References .....	94

<b>6 Transfer Function of Linear Systems .....</b>	95
6.1 Zeros of the Transfer Function .....	95
6.1.1 $z$ -Transform of Impulse Response .....	95
6.1.2 Magnitude Frequency Response for Zero .....	96
6.1.3 Power Spectral Density Function for Zeros .....	96
6.1.4 Auto-correlation Sequence for Output Response Through a Single Echo System .....	100
6.1.5 Logarithmic Expression for Power Spectral Density Function.....	100
6.2 Phase and Accumulated Phase .....	104
6.2.1 Phase Characteristics for Zeros .....	104
6.2.2 Analytic Expression for Phase Frequency Response.....	106
6.2.3 Geometric Interpretation of Phase Response .....	111
6.2.4 Poles, Zeros, and Accumulated Phase .....	112
6.3 Minimum Phase and Cepstral Sequences .....	116
6.3.1 Magnitude and Phase Characteristics for Minimum-Phase Transfer Function .....	116
6.3.2 Decomposition of Minimum-Phase Cepstrum into Magnitude and Phase Terms .....	117
6.4 Decomposition of Transfer Function into Minimum-Phase and All-Pass Transfer Functions .....	119
6.4.1 Decomposition of Transfer Function by Pole-Zero Diagram ..	119
6.4.2 Cepstral Decomposition of Transfer Function into Minimum-Phase and All-Pass Components .....	122
6.4.3 Samples of Cepstral Sequences and Decomposition .....	123
6.5 Linear Phase and Ideal Low-Pass Filter.....	126
6.5.1 Linear Phase System.....	126
6.5.2 Ideal Low-Pass Filter and Fourier Transform .....	126
6.5.3 Resonator and Cut-Off Frequency of Low-Pass Filter .....	127
6.5.4 Zeros for a Truncated Ideal Low-Pass Filter .....	129
6.5.5 Linear Phase and Ideal Low-Pass Filter .....	132
References .....	135
<b>7 Sampling Theorem and Discrete Fourier Transform .....</b>	137
7.1 Sampling of Spectral Function .....	137
7.1.1 Simultaneous Equations for Representation of Periodic Sequence .....	137
7.1.2 Periodic Property in Time and Frequency Planes .....	138
7.2 Discrete Fourier Transform and Periodic Property .....	139
7.2.1 Discrete Fourier Transform Pair .....	139
7.2.2 Interpolation of Time Sequence .....	140
7.2.3 Interpolation of Discrete Spectral Sequence.....	142
7.3 Sampling Theorem .....	142
7.3.1 Partial Sum of Fourier Series .....	142
7.3.2 Periodic Property of Spectrum for Sampled Sequence .....	145

7.3.3	Sampling Partial Sum .....	147
7.3.4	Sampling Theorem .....	147
7.4	Discrete Fourier Transform and Sampling Theorem .....	149
7.4.1	Sampling of Sinusoidal Function .....	149
7.5	Interpolation and Decimation of Sequences .....	151
7.5.1	Interpolation of Sequences .....	151
7.5.2	Interpolation and Decimation Samples .....	152
7.5.3	Sampling Frequency Conversion .....	156
	References .....	160
<b>8</b>	<b>Sinusoidal Representation of Sequence</b> .....	161
8.1	Spectral Peak Selection .....	161
8.1.1	Representation of Sinusoidal Sequence by Discrete Fourier Transform .....	161
8.1.2	Spectral Estimation of Sinusoidal Sequences by Spectral Interpolation .....	163
8.1.3	Spectral Estimation of Compound Sequence by Spectral Peak Selection .....	165
8.1.4	Harmonic Sequence on Logarithmic Frequency Scale .....	169
8.1.5	Scaling Structure of Harmonic Spectral Function on Logarithmic Frequency Scale .....	170
8.1.6	Auto-correlation Analysis of Musical Tones in the Frequency Domain .....	171
8.2	Clustered Line Spectral Modeling .....	174
8.2.1	Windowed Sinusoidal Function and Its Spectral Sequence ...	174
8.2.2	Formulation of CLSM .....	176
8.2.3	LSE Solution of the Simultaneous Equations .....	178
8.2.4	CLSM Examples .....	180
8.3	Prediction of Compound Sinusoidal Sequences .....	186
8.3.1	Compound Sinusoidal Sequence and Almost Periodic Function .....	186
8.3.2	Sinusoidal Compound Sequences and Spectral Functions .....	187
8.3.3	Spectral Peak Selection and Extensions Outside the Observation Interval .....	187
	References .....	189
<b>9</b>	<b>Modeling for Zeros in Complex Time and Frequency Plane</b> .....	191
9.1	Sinusoidal Modeling and Zeros for Transfer Functions .....	191
9.1.1	Frequency Characteristics for Single-Degree-of-Freedom System .....	191
9.1.2	Residues and Zeros of Transfer Function .....	194
9.1.3	Sinusoidal Modeling and Zeros for External Source .....	200
9.2	Clustered Time-Sequence Modeling (CTSM) .....	204
9.2.1	Correspondence Between Time and Frequency Regions .....	204
9.2.2	Formulation of CTSM .....	210

Contents	xv
9.2.3 Source Signature Analysis by CTSM .....	211
9.3 Adjacent Pairing Time-Pulse Modeling of Zeros .....	217
9.3.1 Formulation of Adjacent Pairing Time-Pulse Modeling (APTM) .....	217
9.3.2 Spectral Trough Tracking .....	218
9.3.3 Representation of Source Signature by Spectral Trough Selection .....	219
References .....	220
<b>Erratum .....</b>	E1
<b>Index .....</b>	223

# Chapter 1

## Introduction

The issues dealt with in this book are summarized in this introduction to provide a helpful and informative overview of all chapters so that the purpose of each can be separately sought. The content or motivation of each chapter is briefly described as well as how each issue is developed in every chapter from the mathematical expressions of basic concepts in Chap. 2 through to the advanced topics in Chap. 9.

Signals of sound are conveyed in both the spectral and temporal characteristics of sound waves. To assess these characteristics, waveform analyses of a discrete sequence modeled by a linear system of equations are performed in both time and frequency domains. Based on the theory of linear systems, the Fourier transformation of these sequences represents the sequence over both temporal and frequency domains, and is the fundamental starting point to waveform analysis in this book.

Following this introduction, Chap. 2 gives a brief summary and introduction of mathematical expressions and operations of the representation of discrete sequences. These include convolution, correlation,  $z$ -transforms, and Fourier transforms, which are defined using instantaneous values sampled from continuous functions over a set of sampling periods. The  $z$ -transforms use continuous functions defined on the complex frequency domain, whereas Fourier transforms use both continuous and periodic functions defined on the unit circle centered on the origin ( $z = 0$ ) of the complex frequency domain. This summary will be helpful to readers who are not familiar with discrete signal analysis; however, it is recommended that all readers become acquainted with the basics of signal processing theory from introductory textbooks [1] complementing this book. However, it may surprise the cognizant reader to see the  $z$ -transformation introduced through the generalization

of the generating functions of discrete signals into the complex domain. The concept of generating functions is well known in combinatorial mathematics [2].

The dominant characteristics that distinguish the temporal features of a discrete sequence arguably are the period and envelope of the waveform. Chapter 3 introduces the well-known relationship between the power spectral density function and the auto-correlation sequence following the definition of the magnitude and phase spectrum of a sequence. Auto-correlation analyses are the first steps to estimations of the period of a sequence.

Magnitude (or power) and phase spectral properties of discrete signals are the basis of signal analysis, but the importance of phase spectral characteristics has been overlooked in general [3, 4]. The auto-correlation sequence, defined by the power spectral density function, yields only the sequence under 0-phase condition. A sequence cannot be reconstructed using the auto-correlations, even if the period can be estimated by auto-correlation analysis. Indeed, phase spectral properties are necessary to represent the sequence uniquely. Chapter 3 explains the concept of phase correlation and group delay functions in the frequency plane, so that phase effects affecting the temporal envelopes might be understood more intuitively, and also that group delay can be visually interpreted in terms of the delay of the envelope.

An introduction to signal analysis requires appreciating the temporal and spectral analysis of sequences. The temporal envelopes show the dynamics of the sequences, which typically go unseen in the auto-correlation functions. A frame-wise approach is taken to temporal analysis to see the temporal behavior in the context of power spectral analysis. The envelopes represent the temporal changes in macroscopic structures, whereas the frame-wise spectral properties can be interpreted as related to the time-dependent fine structure of a sequence. The local behavior of phase spectral records or group delay is crucial in the construction of the envelopes.

A triangular window is a good candidate as a weighing function for sampling a short frame of the sequence from the whole sequence. The inspiration for introducing triangular windows into the frame-wise approach of, for example, frame-wise auto-correlation analysis, which enables the temporal behavior of a sequence to be seen using the power-spectral properties, is drawn from the mathematical expressions developed for the group delay function where differential operations are not explicitly used. The analogous context here, the relationship between the spectral derivatives (or averaging) in the frequency domain and the triangular windowing with a positive (or negative) slope in the time domain, is intriguing. Moreover, the short-term auto-correlations are obtained keeping its non-negative Fourier transform by applying the triangular window.

Chapter 4 formulates the transient and steady-state response of a sound field. A formulation with respect to the power spectral and auto-correlation analysis in the time and frequency domains of sound radiating from a source into a room helps in interpreting the transient and steady-state responses in terms of the impulse response and its frequency characteristics between the source and receiver in the room. The responses might be intuitively understood from a perceptual point of view. However, a deeper question can be posed here. The magnitude and phase spectral

characteristics of the steady state in a room look like random processes in the frequency plane. This is because a lot of reflected waves superpose with the direct sound (travelling from the source to the receiving positions without reflections) with random magnitude and phase components [5, 6]. Randomness over either the frequency or temporal domain seems unlikely from sound signals as in intelligible speech circulating from source to listeners in a room. Human ears must be well structured so that speech can be recognizable, even if the frequency characteristics of the transmission path seem random [7]. The smart scheme is one that develops a waveform analysis that extracts the source information (or the source effects) by separating the path information from the response [8].

Frame-wise auto-correlation analysis reveals some of the significant source signatures from reverberant speech in which the original speech is embedded. The early portion of the auto-correlation sequences for impulse responses between the source and receiving positions yields macroscopic power spectral features, such as the power spectral envelopes, that are not highly sensitive to the reverberation condition and thus are preserved even in a reverberation field. This partly explains why intelligible speech can be delivered under adverse reverberation.

The auto-correlation analysis of sound followed by a single reflection produces a condition where the direct sound is enhanced by the reflection without spectral degradation [9]. The source effects hidden in the responses are strongly dependent on distance from the source to listening positions as long as that distance is relatively short. The interest here is seeing the coherent range in which sound appears as a spherical travelling wave radiating from a point source even in a reverberant space [3, 10, 11]. The source signature depends on the distance from the source. Chapter 4 presents examples in which the envelope spectrum from the coherent region is remarkably sensitive to the distance. The best distance (position) to listen to intelligible speech can be estimated using power spectral analysis of the narrow-band temporal envelopes. In addition, seeing the differences in the spectral changes of the envelopes between the free and reverberant fields is impressive. The effects of sound reflections, even in the coherent region, are interpreted by the comparison between the two-types of fields, such as in the anechoic and echoic rooms.

Speech intelligibility or period is a major aspect regarding the signals of audible sound. “What is the dominant factor in regard to intelligibility and pitch perception?” would be an interesting question in waveform analysis of sound. Following the formulation of the magnitude and phase spectral properties of temporal envelopes in terms of the complex transfer function, Chap. 5 describes interesting experimental results related to intelligibility, narrow-band temporal envelopes, and the frame-wise magnitude and phase spectral properties of speech samples [4]. The narrow-band temporal envelopes essentially convey intelligibility [12]. The intriguing aspect is seeing a frame-length-dependent phase dominance rather than the frame-wise magnitude spectrum of speech sample while maintaining narrow-band envelopes. The phase spectral properties must be maintained to produce samples of intelligible speech if the frame-length is very long or extremely short. A time-reversed speech sample, briefly mentioned in Chap. 5, is a good example that shows that no intelligible speech is produced under conditions where the magnitude

spectral property for a whole sentence is identical with the original one; however, its phase is the complex conjugate of the original [13]. The temporal envelopes of the original speech sample were no longer seen for the time-reversed speech.

Pitch perception, which depends on the fundamental frequency, is one of the more important attributes of hearing. Historically, place theory asserts that pitch perception depends on the most dominant power spectral component and is generated by the resonance features of the basilar membrane; however, the theory does not explain pitch perception when no fundamental component is contained in the sound of interest. The time theory, an alternative to place theory, assumes the pitch perception arises from the period of the sound. Chapter 5 introduces time theory and provides samples, including those with missing fundamentals, of period analysis using the auto-correlation of sound. In particular, auto-correlation sequences are introduced after removing power spectral characteristics by setting the spectral magnitude to unity for the sinusoidal components contained in the sound [14]. Consequently, it is possible to estimate the period in the time domain where time theory works without the power spectral effect in the frequency domain.

An example of period estimation for a musical chord, however, poses a question from a perceptual point of view. The estimated period is too long; specifically, the fundamental frequency is two-octaves lower than that for the root of the chord. This example suggests more or less that pitch perception might follow both types of theories.

Chapter 5 also discusses period analysis of narrow-band temporal envelopes. Such envelopes essentially govern the periodic nature of sound as well as the intelligibility of speech. Nevertheless, Chap. 5 provides a counter example of period estimation in which the envelope model does not work well [14]. Examples of period estimation are also given in Chap. 8.

Linear systems theory is the fundamental basis of waveform analysis. Linear systems are identified by their  $z$ -transforms (transfer functions), which can be represented by their poles and zeros in the complex frequency plane. Chapter 6 deals with the frequency response of a linear system in terms of the poles and zeros of the transfer function. Recall that the transfer function is represented by a polynomial factorized as products of first-order polynomials. Except at the origin of the complex frequency plane, it is an analytic function subject to the constraint that the length of its impulse response is finite. The transfer function can be expressed then as a product of transfer functions for unit-sample delay systems (sum of the no-delay and a single sample delay) samples. The transfer function for a unit-sample delay system can be identified by a single zero. These zeros are the main concern in Chap. 6 and indeed throughout this book.

Following the logarithmic expression for the transfer functions, the phase- and group-delay characteristics are formulated for the zeros. Introducing the logarithmic expressions provides the basis from which the cepstral sequences are derived [1]. The expressions show that the phase characteristics cannot be a continuous function as the zeros are close to the unit circle in the complex frequency plane, and thus the group delay makes an abrupt sign change if the location of the zero moves across the unit circle [15]. Consequently, the phase responses can be formally represented as

the Fourier series expansion of the logarithmic expression for the Fourier transform. The phase responses can be interpreted geometrically as the angle on the unit circle made by the complex transfer function, and also understood in light of the residue theorem governing the integration of complex functions over complex domains.

The relationship between the real and imaginary parts of a spectral sequence can be theoretically formulated for a real causal sequence. Chapter 6 formally shows the relationship between the magnitude and phase frequency responses for a minimum-phase system by introducing the cepstral sequences from the logarithmic expression of the  $z$ -transform. Decomposition of the transfer function into the minimum-phase and all-pass components is the fundamental theorem of a linear system. Chapter 6 covers the analytical decomposition of the zeros of the transfer function as well as giving a visual assignment of the location of zeros.

The transfer function of an all-pass system comprises symmetric pairs each composed of a zero and a pole with respect to the unit circle; the exceptions are the poles at the origin. The zero (pole) is located outside (inside) the unit circle. These symmetric pairs characterize the flat-magnitude frequency response and explain why the system is called all-pass. Interestingly, replacing the inside poles with zeros characterizes a linear phase system. Specifically, a linear phase system, whose phase response is linearly proportional to the frequency instead of the flat magnitude for the all-pass system, is identified by a transfer function associated with pairs of symmetric zeros, apart from the poles at the origin, with respect to the unit circle. In addition, the impulse response of a linear-phase system has also a symmetric sequence on the time axis; however, a symmetric sequence does not necessarily produce a linear phase. Following the introduction of the all-pass and linear phase systems, Chap. 6 deals with the ideal low-pass filter, which is important in constructing the sampling theorem described in the next chapter. The ideal low-pass filter can be interpreted as a band-limited all-pass system for which the frequency response is essentially discontinuous at the cut-off frequency. However, it would be intriguing to understand the response around the cut-off frequency by analogy with the frequency response of a single resonator for which the resonant frequency is the cut-off frequency. This analogy arises because the record length of the impulse response is infinitely long for the ideal low-pass filter for which the transfer function is characterized by the poles.

In contrast, the transfer function can be represented by the zeros if the impulse response is truncated to finite length. Chapter 6 describes how the frequency response can be written as the Dirichlet kernel [16]. To be specific, the truncation of the impulse response of the ideal low-pass filter can be interpreted as that for the Fourier series expansion. Interestingly, Chap. 6 illustrates the effect of hidden resonance by a virtual pole of the frequency response [17]. This effect even occurs for the truncated impulse response by applying the triangular window with positive slope.

Chapter 6 closes with issues related to linear phase and minimum-phase low-pass filters. From a theoretical standpoint, the ideal low-pass filter cannot have a linear phase, although the impulse response is a symmetric sequence. Chapter 6 introduces the auto-correlation sequence for the truncated impulse response of the ideal low-

pass filter by adding the time delay corresponding to the number of samples in the non-causal part. In general, the linear-phase systems are important tools for signal processing, because the narrow-band envelopes of an input sequence might be simply delayed without degradation. However, it might be problematic from the point of view of audio-engineering-type applications. This is because a linear-phase system generally has a slow transient response because of the symmetric impulse response sequence.

The sampling theorem is the backbone supporting discrete signal processing. Chapter 7 deals with the discrete Fourier transform in parallel with the sampling theorem. The spectral function that is derived by the Fourier transformation of a discrete sequence composed of  $N$ -entries can be sampled by the discrete Fourier transform to give a spectral sequence or line-spectral components. Specifically, an  $N$ -point sequence can be represented using the continuous spectral function in accordance with the Fourier transform or the discrete spectral sequence in accordance with the discrete Fourier transform (DFT). Chapter 7 formulates the DFT as the solution of a set of simultaneous linear equations for which the orthogonality of the sinusoidal sequences holds. Recalling that a periodic function is composed of line-spectral components, the periodic properties for the sequences over both the time and frequency domains can be understood. The Fourier transformation of a discrete sequence produces the periodic and continuous spectral functions. In contrast, the DFT function is given by the periodic and discrete spectral sequence, where the time sequence is assumed to be expanded to the periodic sequence. These periodic structures for the sequences in both the time and frequency domains provide the basis to the application of the sampling theorem.

The main theoretical issues dealt with in Chap. 7 are interpolation and decimation of the sequences that can be interpreted as sampling frequency conversion without violating the sampling theorem. Of interest here is developing the interpolation of a discrete time (or spectral) sequence, which makes essentially a continuous function from a discrete sequence. It can be formulated using the convolution of the sequence and the Dirichlet kernel. This fact explains the sampling theorem. A discrete sequence sampled from a periodic function composed of  $N$ -sinusoidal components (or line-spectral components) yields a periodic line-spectral (or discrete spectral) sequence using the discrete Fourier transform. The periodic sequence can be interpreted as the periodic expansion of the  $N$ -point spectral sequence of the original periodic continuous function. If the period of the expanded spectral sequence by the sampling is long enough to preserve the original  $N$ -point spectral sequence in the fundamental period without deformation, the original continuous function can then be reconstructed even after sampling (from the sampled time sequence) by taking the partial sum of the spectral sequence in terms of the Fourier series expansion (or ideal low-pass filtering in terms of signal processing). The partial sum is expressed by the convolution of the sampled time-sequence and Dirichlet kernel. Chapter 7 formulates the sampling theorem by taking a periodic function as an example where the Fourier series expansion holds, and thus the ideal low-pass filtering can be represented by the Dirichlet kernel. The simple example

of sampling a sinusoidal function provides an intuitive means to see the so-called aliasing that occurs following the violation of the sampling theorem.

The interpolation and decimation can be expressed in a single unifying equation involving the convolution of the analytic expression of the sequence with the Dirichlet kernel. Chapter 7 provides some examples of sampling frequency conversion given by this analytic convolution formula.

Waveform analysis, in principle, is based on Fourier analysis in which sinusoidal functions play an important role. A sinusoidal function is only a periodic function exhibiting a single frequency and hence has no harmonics. When a waveform is decomposed into sinusoidal functions, these are viewed in an analogy as prime numbers by which any number can be factorized as a product. Therefore, each frequency of a sinusoidal signal comprising the waveform must be determined by signal analysis.

However, when a sinusoidal signal is observed or recorded in a finite interval, the determination may not be possible. Spectral estimates show many frequency components different from the original signal frequency of a given sinusoid. Although true from the mathematical perspective of Fourier analysis, it goes against the intentions of spectral analysis used in practice. An example is shown in Chap. 8 using the Fourier series expansion of a sinusoidal function that recalls the mathematical manner in which signals are expressed. Only the original sinusoidal frequency, along with magnitude and phase spectrum, requires estimating from a practical point of view. In this book, the true frequency is defined as the frequency that might be estimated if a sinusoidal signal were observed in an infinite interval. In Chap. 8, spectral peak selection is introduced so that an audio signal might be better represented in accordance with its true frequencies from which its complex spectrum can be estimated by interpolating the frequencies at the spectral peaks [18, 19]. The sequence to which the spectral peak selection is applied is not necessarily composed of harmonic components. A sequence composed of harmonics on the logarithmic scale, a kind of Weierstrass function [7], is an intriguing prospect to see how the spectral peak selection works.

Following the representation of a compound sinusoidal sequence by spectral peak selection, spectral peak-frequency selection and its auto-correlation analysis is developed so that period analysis might be carried out on a frame-by-frame basis. Here the power spectral effects on the period estimation can be avoided (or discarded) in every single frame because of the peak-frequency selection; however, the temporal (or frame-wise) statistics of the power spectral nature can be observed without losses [20]. The spectral auto-correlation analysis on the frequency domain shows the histogram of the spacing between peak frequencies. The example of period analysis of a musical chord presented in Chap. 8 determines that the fundamental frequency can be estimated by the statistical mode of frequency spacing shown in the frame-averaged spectral auto-correlation sequence. This result implies a possible approach to integration of the place and time theories for the fundamental frequency estimation.

An audio signal might be well represented by its true frequencies, their complex spectra being estimated by interpolating the frequencies at the spectral peaks.

In contrast, spectral peak selection is not always well suited to provide a representation of beats, which are typical examples of auditory events. This is because beats are generally composed of sinusoidal waves, the frequencies of which are too closely clustered to be separately estimated. Chapter 8 develops a method called clustered line spectral modeling (CLSM) [19]. CLSM decomposes the spectral peak, which is formed as an overlap of Fourier-transformed functions over the frequency domain, into the clustered sinusoidal functions in a finite window. Here, this overlap represents the effect of clustering of the sinusoidal frequencies on the spectral domain. The decomposition of the overlapped spectral components is performed by solving the set of linear equations in the frequency plane following the least-squares-error criterion.

CLSM, in principle, gives the spectral estimate for the true sinusoidal components in the observation interval (the so-called window length). It impresses seeing the sinusoidal components forming beats, even if the observation interval is much shorter than the beat period. An exponentially decaying signal is not represented by a single sinusoidal component, even if it is a resonant response whose resonance frequency can be uniquely determined. Such a spread in resonant response around the resonant frequency can also be subjected to the CLSM approach. Chapter 8 features examples of piano-tone analysis (string vibration) [19, 21, 25]. The effect of decay (damping for the resonant response) is represented by the clustered sinusoidal components around the spectral peaks corresponding to the resonant frequencies.

Prediction or extrapolation outside the observation interval is an attractive aspect in waveform analysis. As stated above, it is quite unlikely that a sequence is predicted using conventional Fourier analysis. In contrast, a compound sinusoidal sequence, which is not always periodic from a theoretical point of view, can be predictable, provided that the spectral peak selection or CLSM works well. Of interest is seeing the spectral functions, given by the Fourier transforms of sequences with finite record lengths, represented by the compound sinusoidal functions in the frequency domain. The prediction for spectral functions can be read by expanding the frequency range of interest as long as that for the sampling theorem is not violated. However, extrapolation of the temporal signature including the narrow-band envelopes is not an easy task, in general. The number of clustered sinusoidal components required for CLSM depends on the frequency range of the envelope. A large (small) number of the sinusoids is necessary, if the envelope contains a wide (narrow) frequency range. In particular, almost stationary envelopes might be predictable even using a relatively small number of clustered components, whereas transient envelopes need for predictions a large number of sinusoids clustered around the central frequency component.

The identification of a source signature represented by zeros is an important issue in waveform analysis. Chapter 9 discusses the occurrences of zeros of the transfer function in the complex frequency plane [3, 15, 22, 23]. The zeros lie between two adjacent poles as a result of the superposition of the two dominant responses from the poles and the remainder responses from the other poles that can be approximated as a slow-varying function for the pole pair. Therefore, the occurrences depend on the residues of the poles and the remainder response (function). There are four

instances, i.e., no zeros, a single zero, double zeros, and a symmetric pair of zeros, in the interval of the pole-line connecting the pole pair on the complex frequency plane.

However, the zeros arising from the source condition can be produced independent of the residue and remainder condition [24]. Chapter 9 formulates the source effect whereby zeros are created from modeling error (or remainder function) after CLSM. An example of a piano-string vibration analysis shows the zeros resulting from the source effect. Here, the remainder function creating such zeros corresponds to a brief time sequence observed in the initial portion of the vibration waveform.

Spectral functions regarded as compound sinusoidal functions leads to an interesting relationship between the time and frequency regions [25]. Chapter 9 provides the theoretical correspondence between the complex time and frequency planes with respect to the representation of analytic signals. The correspondence is expressed as the complementarity between complex conjugates [3]. Analytic signals are represented in terms of the envelopes and carriers, or the instantaneous magnitude and phase in the time domain. This pair of envelope and carrier exhibits magnitude-phase complementarity for the spectrum in the complex frequency plane.

Zeros of the complex spectrum are crucial in establishing signal signatures as in the minimum-phase characteristics. Chapter 9 introduces the zeros of the waveforms into the complex time domain based on the complementarity between the time and frequency planes. Following the introduction of the zeros in the complex-time domain, the instantaneous phase or the instantaneous frequency can be interpreted as the complementarity with the phase spectrum or the group delay. The sign of the instantaneous frequency can be explained in terms of the signal zeros in the complex time domain. It is actually an example of complementarity between the instantaneous frequency in the time domain and the group delay in the frequency domain. Here, the instantaneous frequency is defined in the time domain, whereas the group delay is defined on the frequency plane. A modulated sequence in the time domain and a modulated spectrum is an instance where complementarity can be observed. Recovery of the time envelope from its carrier is another instance of complementarity with respect to the minimum-phase property of sequences from a theoretical point of view.

In regard to complementarity between the complex time and frequency planes, Chapter 9 formulates clustered time sequence modeling (CTSM) in the time domain instead of CLSM in the frequency domain [21]. Following the introduction of compound complex sinusoidal functions, the effects of spectral windowing on the time sequence can be understood in accordance with the time-frequency complementarity. Complementarity is used to derive CTSM as a dual algorithm of a discrete time sequence to CLSM of a spectral sequence.

CLSM, based on the compound sinusoidal sequence model, assumes that the sinusoidal components do not change frame by frame (locations of observation intervals). Thus, the example of vibration analysis of a piano string taken in Chap. 8 actually determines that the modeling error is concentrated in the initial portion of the string vibration rather than the whole vibration sequence. This result may indicate that a transient part as in the initial portion of string vibration can be well

represented by an approach in the time domain rather than the frequency domain approach as in CLSM. An actual reflected wave travelling on the piano string, including the initial portion, is displayed using CTSM analysis in Chap. 9 [21]. The direct and single cycles of the waveform represent the informative source signature including the spectral trough (corresponding to each zero) on the frequency domain.

A brief time sequence is factorized into products in which each term defines a single zero from the adjacent pulse pair with a single unit of time delay. Chapter 9 formulates adjacent pairing time pulse modeling (APTM) to represent zeros on the complex frequency plane [26]. Zeros are localized by iteratively solving sets of simultaneous equations using least-squares-error on the frequency plane. Consequently spectral trough selection is developed to represent the transfer function or the source signatures as for the spectral peak selection. The example of a piano-string vibration shows that CLSM, CTSM, and APTM are possible approaches to separate the zeros of the source waveform from those of the transfer function.

## References

1. M. Tohyama, T. Koike, *Fundamentals of Acoustic Signal Processing* (Academic Press, 1998)
2. R. Nelson, *Probability, Stochastic Processes, and Queueing Theory* (Springer, 1995)
3. M. Tohyama, *Sound and Signals* (Springer, 2011)
4. M. Kazama, S. Gotoh, M. Tohyama, T. Houtgast, On the significance of phase in the short term Fourier spectrum for speech intelligibility. *J. Acoust. Soc. Am.* **127**(3), 1432–1439 (2010)
5. M.R. Schroeder, Statistical parameters of the frequency response curves in large rooms. *J. Audio Eng. Soc.* **35**(5), 299–306 (1987)
6. R.H. Lyon, Statistical analysis of power injection and responses in structures and rooms. *J. Acoust. Soc. Am.* **45**(3), 545–565 (1969)
7. M.R. Schroeder, *Fractals, Chaos, Power Laws* (W.H. Freeman and Company, 1991)
8. R.H. Lyon, *Machinery Noise and Diagnostics* (Butterworth, 2000)
9. Y. Ando, *Auditory and Visual Sensation* (Springer, 2009)
10. P.M. Morse, R.H. Bolt, Sound waves in rooms. *Rev. Mod. Phys.* **16**, 69–150 (1944)
11. Y. Takahashi, M. Tohyama, Y. Yamasaki, Phase response of transfer functions and coherent field in a reverberation room. *Electron. Commun. Jpn.* **90**(4 Part 3), 1–8 (2007)
12. T. Houtgast, H.J.M. Steeneken, R. Plomp, A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* **77**(3), 1069–1077 (1985)
13. K. Saberi, D.R. Perrot, Cognitive restoration of reversed speech. *Nature* **398**, 760 (1999)
14. M.R. Schroeder, *Computer Speech* (Springer, 1999)
15. M. Tohyama, R.H. Lyon, Zeros of a transfer function in a multi-degree-of-freedom system. *J. Acoust. Soc. Am.* **86**(5), 1854–1863 (1989)
16. A. Zygmund, *Trigonometric Series* (The University Press, Cambridge, 1988)
17. Y. Takahashi, M. Tohyama, Y. Yamasaki, Cumulative spectral analysis for transient decaying signals in a transmission system including a feedback loop. *J. Audio Eng. Soc.* **54**(7/8), 620–629 (2006)
18. M. Kazama, M. Tohyama, Estimation of speech components by ACF analysis in a noisy environment. *J. Sound Vib.* **241**(1), 41–52 (2001)
19. M. Kazama, K. Yoshida, M. Tohyama, Signal representation including waveform envelope by clustered line-spectrum modeling. *J. Audio Eng. Soc.* **51**(3), 123–137 (2003)

20. Y. Hirata, Estimation of the frequency response of a structure using its non-stationary vibration. *J. Sound Vib.* **313**, 363–366 (2008)
21. T. Hasegawa, M. Tohyama, Analysis of spectral and temporal waveforms of piano-string vibration. *J. Audio Eng. Soc.* **60**(4), 237–245 (2012)
22. R.H. Lyon, Progressive phase trends in multi-degree-of-freedom systems. *J. Acoust. Soc. Am.* **73**(4), 1223–1228 (1983)
23. R.H. Lyon, Range and frequency dependence of transfer function phase. *J. Acoust. Soc. Am.* **76**(5), 1435–1437 (1984)
24. T. Hasegawa, M. Tohyama, Separation of zeros for source signature identification under reverberant path condition. *J. Acoust. Soc. Am.* **130**(4), EL271–EL275 (2011)
25. R. Kumaresari, A. Rao, On minimum/maximum/all-pass decomposition in time and frequency domains. *IEEE Trans. SP* **48**, 2973–2976 (2000)
26. T. Hasegawa, M. Tohyama, Source signature identification by using pole/zero modeling of transfer function. *Inter Noise* **2011**, 431616 (2011)

# Chapter 2

## Discrete Sequences and Their Fourier Transform

A brief summary is given of mathematical terms, expressions, and complex functions defined using discrete sequences, including convolution, correlation,  $z$ -transform, Fourier transform, analytical representation of sequences, magnitude and phase, and poles and zeros of complex functions. The topics covered in this chapter are very elementary or introductory, but provide the basis on which all other issues dealt with in this book are constructed.

### 2.1 Function and Sequence

In discrete signal analysis, a sequence  $\{x(n)\}$  (or simply  $x(n)$  to denote a function  $x$  of  $n$ ) is taken instead of a function  $s(t)$ . Here  $n$  denotes an integer, whereas  $t$  is a continuous variable. Given  $s(t)$  as a function of a time variable  $t$ , a discrete sequence  $x(n)$  can be derived by sampling at specified times. Suppose that a function  $s(t)$  is sampled after every time interval  $T_s$ (s). The discrete sequence  $x(n)$  can then be defined as  $x(n) = T_s \cdot s(t)|_{t=nT_s}$ . Here  $T_s$  is called the sampling period; equivalently,  $F_s = 1/T_s$  is called the sampling frequency (Hz). Thus, a discrete signal  $x(n)$  is a sequence defined over a set of integers  $n$  in analogy to a continuous function  $s(t)$  given on an interval of the real numbers  $t$ .

### 2.2 Generating Functions and $z$ -Transforms

#### 2.2.1 Convolution and Generating Function

Each discrete sequence determines a generating function [1], a polynomial (or a formal power series) on a formal variable  $X$  defined as  $A(X) = \sum_n a(n)X^n$ .

Such functions are useful when multiplying sequences together. In particular, a fundamental operation between two discrete sequences is convolution.

The convolution of two sequences is defined in accord with the product of the two corresponding polynomials. Suppose that  $a(n)$  and  $b(n)$  are the two discrete sequences. For each sequence, introduce its generating function [1]:

$$A(X) = \sum_m a(m)X^m \quad (2.1)$$

$$B(X) = \sum_n b(n)X^n. \quad (2.2)$$

Their product is given as  $C(X) = A(X)B(X)$  where

$$\begin{aligned} C(X) &= \left[ \sum_m a(m)X^m \right] \left[ \sum_n b(n)X^n \right] \\ &= \sum_m \sum_n a(m)b(n)X^{m+n} \\ &= \sum_p \sum_m a(m)b(p-m)X^p \\ &= \sum_p c(p)X^p. \end{aligned} \quad (2.3)$$

By comparing the coefficients for a given exponent of  $X$ , the sequence  $c(n)$  that is to be called the convolution of  $a(n)$  and  $b(n)$  is then determined to be

$$\begin{aligned} c(n) &= a * b(n) \\ &= \sum_m a(m)b(n-m) \\ &= \sum_m a(n-m)b(m) \\ &= b * a(n). \end{aligned} \quad (2.4)$$

For example, sequence  $\{1, 0, -1\}$  can be written as

$$\{1, 0, -1\} = \{1, 1\} * \{1, -1\}. \quad (2.5)$$

Thus, the outcome above indicates that the convolution operation between a pair of sequences can be interpreted as a process to evaluate the coefficients of the polynomial that is defined by the product of the pair of polynomials.

### 2.2.2 Correlation and Generating Function

Other fundamental operations among sequences are correlations. Again, take sequences  $a(n)$  and  $b(n)$ . One such correlation, the cross-correlation sequence  $r_c(n)$  between  $a(n)$  and  $b(n)$ , is defined as

$$\begin{aligned} r_c(n) &= a \otimes b(n) \\ &= \sum_m a(m)b(m-n). \end{aligned} \quad (2.6)$$

Note that for each summation term in the correlation formula stated above, the relation  $n = m - (m - n)$  holds for each pair of indices, whereas in Eq. 2.4 on the preceding page, namely in the convolution formula,  $n = (n - m) + m$ . The cross-correlation can also be defined in terms of the corresponding generating functions. The generating function for the correlation sequence  $r_c(n)$  can be written as

$$\begin{aligned} R_c(X) &= \sum_n r_c(n)X^n \\ &= A(X)B^*(X) \end{aligned} \quad (2.7)$$

where a second type of generating function,  $B^*(X) = \sum_n b(n)X^{-n}$ , has been introduced. Note here that the cross-correlation includes terms for both positive and negative integers  $n$ . With this in mind, a relation between the two operations, correlation and convolution, holds, specifically

$$c(n) = a * b(n) = a \otimes \hat{b}(n) \quad (2.8)$$

or

$$r_c(n) = a \otimes b(n) = a * \hat{b}(n) \quad (2.9)$$

where  $\hat{b}$  denotes the sequence  $b(-n)$ . For example,

$$\begin{aligned} \{-1, 0, 1\} &= \{1, 1\} \otimes \{1, -1\} \\ &= \{1, 1\} * \{-1, 1\} \end{aligned} \quad (2.10)$$

In the special case when the two sequences of the correlation are the same, the correlation sequence is called an auto-correlation sequence, and is written

$$r_a(n) = a \otimes a(n). \quad (2.11)$$

Its generating function becomes

$$R_a(X) = A(X)A^*(X). \quad (2.12)$$

Note that the auto-correlation sequence is symmetric, that is, it is an even sequence of  $n$ .

### 2.2.3 ***z**-Transformation of Discrete Sequence*

In the previous sub-sections, a formal variable  $X$  was introduced to define the generating function of a discrete sequence. In this sub-section,  $X$  is re-defined to be a point on the complex plane, such as  $X = z^{-1}$ . Substituting this complex variable  $z^{-1}$  for  $X$ , the generating function of a discrete sequence is rewritten as

$$A(z^{-1}) = \sum_n a(n)z^{-n}. \quad (2.13)$$

Here  $A(z^{-1})$  is called the *z*-transform of the sequence  $a(n)$  instead of the generating function.

Suppose a system that gives output sequences  $y_1(n)$  and  $y_2(n)$  corresponding to the input sequences  $x_1(n)$  and  $x_2(n)$ , respectively. The system is called a linear system if the output signal  $y_1(n) + y_2(n)$  corresponds to the input signal  $x_1(n) + x_2(n)$ . The impulse response is defined as the response of the linear system to the unit-pulse sequence where  $u(n) = 0$  for all integers  $n$  except zero, for which  $u(0) = 1$ . Suppose that the impulse response of a linear system is  $h(n)$ . The *z*-transform of the impulse response,  $H(z^{-1}) = \sum_n h(n)z^{-n}$ , is called the transfer function of the linear system. The transfer function or the impulse response uniquely characterizes the linear system.

A linear system is called a causal system, if the impulse response is a right-hand sequence such that  $h(n) = 0$  for  $n < 0$ . In addition, if the sum of the squared absolute value of  $h(n)$ , converges, that is  $\sum_{n=0}^{\infty} |h(n)|^2 < +\infty$ , the system is called causal and stable. Note that the transfer function  $H(z^{-1})$  is not defined on the entire *z*-plane, even if the system is causal and stable. Points in the *z*-plane at which the transfer function is undefined are singular points called poles of the transfer function.

### 2.2.4 ***Convergence of z-Transformation***

Consider the causal and stable power sequence  $a(n) = a^n$  for  $n \geq 0$  where in general  $a$  is complex. The *z*-transform of the sequence

$$\begin{aligned} A(z^{-1}) &= \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} a^n z^{-n} \\ &= \frac{1}{1 - az^{-1}} \end{aligned} \quad (2.14)$$

converges to the function  $A(z^{-1})$  in the limit for  $N \rightarrow +\infty$  if  $|az^{-1}| < 1$ . That is, the region determined by criterion  $|z| > |a|$  is called the region of convergence. The point  $z = a$  is a pole of the function  $A(z^{-1})$ . The equation above also indicates that a causal and stable sequence  $a^n$  for  $|a| < 1$  can be generated from a generating function  $A(z^{-1})$ .

However, with  $|a| > 1$ , the same generating function determines a non-causal and stable sequence. In other words, a region of convergence can also be defined even when the series is not causal. Rewrite the function  $A(z^{-1})$  as

$$A(z^{-1}) = \frac{-z/a}{1 - z/a} = B(z). \quad (2.15)$$

The generating function now denoted by  $B(z)$  can be expanded as a series

$$B(z) = \frac{-z}{a} \sum_{n=0}^{\infty} \left(\frac{z}{a}\right)^n \quad (2.16)$$

for  $|z| < |a|$ . The equation above indicates that a non-causal and stable sequence can be generated from the function  $B(z)$  in place of the causal but non-stable sequence. The regions of convergence for the causal and non-causal stable sequences are complementary; however, note that the generating functions are the same except at  $z = a$ . Consequently, the poles are located within the unit disc with the radius of unity centered at the origin of the  $z$ -plane for the causal and stable sequence, but distributed outside the unit circle for the non-causal and stable sequence.

### 2.2.5 Inverse $z$ -Transformation

As stated above, a discrete sequence can be derived from its  $z$ -transform. This fact explains why the  $z$ -transform can be interpreted as the generating function of the corresponding sequence. This sub-section introduces a formula that yields the corresponding sequence.

Consider a causal or non-causal stable sequence  $a(n)$  for which the  $z$ -transform is  $A(z^{-1})$ . The sequence  $a(n)$  can be written as

$$a(n) = \frac{1}{2\pi i} \oint_{c_R} A(z^{-1}) z^{n-1} dz \quad (2.17)$$

where  $A(z^{-1}) = \sum_n a(n)z^{-n}$ . The equation above is called the representation of a sequence by the inverse  $z$ -transform. Here, the contour  $c_R$  is taken to be a circle with radius  $R$  centered at the origin  $z = 0$ ;  $R$  is such that all poles of  $A(z^{-1})$  are located inside the disc. The proof of Eq. 2.17 on the previous page can be developed by recalling the contour integration of the  $n^{\text{th}}$  power of a complex variable,

$$\oint_{c_R} z^n dz = 2\pi i \delta(n+1) \quad (2.18)$$

where  $\delta(n+1) = 1$  for  $n = -1$ , otherwise 0. Hence,

$$\begin{aligned} & \frac{1}{2\pi i} \oint_{c_R} A(z^{-1}) z^{n-1} dz \\ &= \frac{1}{2\pi i} \sum_m a(m) \oint_{c_R} z^{-m+n-1} dz \\ &= \sum_m a(m) \delta(m-n) \\ &= a(n). \end{aligned} \quad (2.19)$$

The inverse  $z$ -transformation formula above also can be proved for the closed form of the function  $A(z^{-1})$  by recalling the complex integration following the residue theorem,

$$f(a) = \frac{1}{2\pi i} \oint_{c_R} \frac{f(z)}{z-a} dz \quad (2.20)$$

where none of the poles of  $f(z)$  lie on the contour path  $c_R$ , the disc of radius  $R$  centered at  $z = 0$ , or contained inside  $c_R$ , whereas  $z = a$  is located inside the disc [2, 3].

## 2.3 Zeros, Poles, and Unstable Poles

As described in the previous section, unstable poles are distributed outside the unit disc on the  $z$ -plane. It might be interesting to see in what context unstable poles appear in waveform analysis.

### 2.3.1 Zeros of Polynomial and Poles of Series

A polynomial such as

$$f(z) = \sum_{n=0}^{N-1} a_n z^n \quad (2.21)$$

has  $N - 1$  roots in the complex domain. That is, the polynomial above can be factorized such that [2].

$$f(z) = K(z - z_1)(z - z_2) \cdots (z - z_{N-1}). \quad (2.22)$$

Here  $K$  is a constant, and  $z_i$  is a root (i.e., a zero) of the polynomial, because  $f(z_i) = 0$ . The factorization theorem states that there are no poles for a polynomial of finite order.

Rather than the  $z$ -transform, consider instead the power series representation of a function

$$g(z) = \sum_{n=0}^{\infty} b^n z^n \quad (2.23)$$

where  $b$  is a real number. The series above can be rewritten as

$$g(z) = \sum_{n=0}^{\infty} b^n z^n = \frac{1}{1 - bz} \quad (2.24)$$

provided  $|bz| < 1$ . Note that the pole of the power series,  $z_p = b^{-1}$ , is located outside the convergence region.

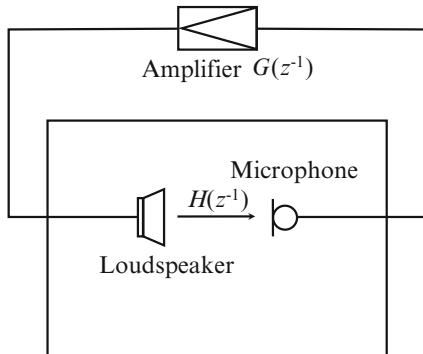
Conversely, the polynomial can be represented

$$g_N(z) = \frac{1 - b^N z^N}{1 - bz} \quad (2.25)$$

by truncating the series  $g(z)$  at  $n = N - 1$ . Here, there are  $N - 1$  zeros given by  $z_k = b^{-1} e^{i2\pi k/N}$  for  $k = 1, 2, \dots, N - 1$ , all located on the circle centered at  $z = 0$  with radius  $b^{-1}$ . The other zero of the numerator  $z_0 = b^{-1}$  associated with  $k = 0$  is canceled by the zero of the denominator. Consequently, the polynomial above has only zeros, but no poles.

This fact indicates that the poles cannot be observed in a mathematical sense from the  $z$ -transforms of sequences as long as the length of the sequence is finite. However, a virtual pole at  $z_p = b^{-1}$  could be estimated in the convergent region instead, if the signal length  $N$  is long enough. In other words, a pole for a sequence of finite length cannot be seen from the mathematical point of view by computational analysis, but only the effects (or traces) of the poles can be observed. Therefore, the pole locations can be estimated, even if only the finite-length sequence is obtained from a sampled sequence. Here, the estimated pole locations indicate the singularities in the mathematical sense that could be observed if the sequence length was infinitely long.

**Fig. 2.1** Model of public address system, from [4] (Fig. 15.10)



### 2.3.2 Unstable Pole

As stated above, the traces of the poles are observed even for a finite length of sequence. Therefore, the effects are problematic from an engineering point of view, when the virtual poles are unstable. In communication systems, this phenomenon is associated with howling or singing that sometimes occurs over a public address system because of sudden changes in the acoustic echo path between pairs of microphones and loudspeaker systems.

Figure 2.1 shows a simplified scheme of a public address system. This system constitutes a closed loop because of a recursive sound path. The transfer function for such a recursive system is written as

$$\begin{aligned}
 L(z^{-1}) &= H(z^{-1}) \frac{1}{1 - G(z^{-1})H(z^{-1})} \\
 &= H(z^{-1})F(z^{-1}) \\
 &= H(z^{-1}) \frac{1}{E(z^{-1})}
 \end{aligned} \tag{2.26}$$

where  $E(z^{-1}) = 1 - G(z^{-1})H(z^{-1})$  assuming  $|G(z^{-1})H(z^{-1})| < 1$ . Note that the transfer function itself holds even if the assumption breaks down. This results from analytic continuation of the transfer function in the complex plane modulo singularities of the function [2].

Nevertheless, as described in Sect. 2.2.3 on page 16, the system is stable if  $\sum_{n=0}^{\infty} |f(n)|^2 < +\infty$ . Here,  $f(n)$  is the impulse response corresponding to  $F(z^{-1})$  for the closed loop shown in Fig. 2.1. System stability is conditioned on whether the poles of the transfer function are located inside the unit disc. In other words, all zeros of  $E(z^{-1})$  must be located inside the unit disc for the stable system. This is equivalent to the condition that  $|G(z^{-1})H(z^{-1})| < 1$  when  $z = e^{j\omega}$ . Here  $e^{j\omega}$  determines the unit circle of the complex domain.

Consider the following very simplified example of a single-zero system

$$H(z^{-1}) = 1 - az^{-1} \quad (2.27)$$

and

$$G(z^{-1}) = b \quad (2.28)$$

where  $0 < b < 1$ . The transfer function  $L(z^{-1})$  becomes

$$\begin{aligned} L(z^{-1}) &= \frac{1 - az^{-1}}{1 - b(1 - az^{-1})} \\ &= \frac{1 - az^{-1}}{1 - b} \cdot \frac{1}{1 - \alpha z^{-1}} \end{aligned} \quad (2.29)$$

where

$$\alpha = \frac{-ab}{1 - b}. \quad (2.30)$$

The impulse response for the closed loop  $f(n)$  can be written as

$$f(n) = \left( \frac{-ab}{1 - b} \right)^n \frac{1}{1 - b} \quad (2.31)$$

where  $n$  is a nonnegative integer. The impulse response  $f(n)$  is stable provided  $|\frac{-ab}{1-b}| < 1$ . Note that  $\frac{-ab}{1-b}$  determines the pole location of  $F(z^{-1})$ . The above example thus illustrates a system that is stable as long as the pole of the transfer function is located inside the unit disc. This is equivalent to  $|b(1 - az^{-1})| < 1$  when  $z = e^{j\Omega}$ . If a pole of the transfer function moves out of the unit disc, the system becomes unstable, thus initiating the phenomenon associated with howling. The pole moves in accordance with the open-loop condition described by  $H(z^{-1})$  and/or the loop gain.

An ideally stable system that produces no howling is made by inserting a linear system having transfer function

$$G_i(z^{-1}) = \frac{-1}{H(z^{-1})} = -H^{-1}(z^{-1}). \quad (2.32)$$

A system with transfer function of the form  $H^{-1}(z^{-1})$  is called the inverse system for  $H(z^{-1})$ . If the ideal inverse system is inserted as a feedback system, as illustrated in Fig. 2.1 on the preceding page, the transfer function of the recursive system can be written as

$$\begin{aligned} L(z^{-1}) &= H(z^{-1}) \frac{1}{1 + G(z^{-1})H^{-1}(z^{-1})H(z^{-1})} \\ &= (1 - az^{-1}) \frac{1}{1 + b} \end{aligned} \quad (2.33)$$

where  $G(z^{-1}) = b > 0$ . The result above shows that the impulse response for the system is stable; specifically, there are no poles as long as  $b > 0$ . This type of ideal system is commonly called a negative feedback system within the context of classical linear systems [5]. However, the inverse system  $H^{-1}(z^{-1})$  is not realizable in general as it is stable only if all zeros of the transfer function (to be inverted) are located inside the unit disc [4].

## 2.4 Fourier Transform

In this section, the Fourier transform is introduced into the  $z$ -transformation on the unit circle, which below is parameterized as  $e^{i\Omega}$ .

### 2.4.1 Fourier Transform of Discrete Sequence

Recall the  $z$ -transform  $A(z^{-1})$  for a discrete sequence  $a(n)$ . Substituting  $z = e^{i\Omega}$

$$A(z^{-1})|_{z=e^{i\Omega}} = A(e^{-i\Omega}) = \sum_n a(n)e^{-i\Omega n} \quad (2.34)$$

where the phase angle  $\Omega$  denotes the normalized angular frequency equal to  $\omega T_s$ , and  $T_s$  is the sampling period of the sequence. Equation 2.34 is called the Fourier transform or the continuous spectrum of the discrete sequence  $a(n)$ . A particular instance arises when the sequence corresponds to the impulse response of a linear system, for then the Fourier transform is called the frequency response or (characteristics) of the linear system of interest.

According to the definition of the Fourier transform above,  $A(e^{-i\Omega})$  is a periodic function of period  $2\pi$  with respect to  $\Omega$ . Note that the normalized angular frequency  $2\pi$  corresponds to the sampling (angular) frequency given by  $2\pi/T_s$ . Recalling the definition of the generating function of a discrete sequence, the Fourier transform also determines the generating function for the discrete sequence. In other words, the discrete sequence itself becomes the coefficients of the Fourier series expansion for the continuous periodic function  $A(e^{-i\Omega})$ . Consequently, the sequence  $a(n)$  can be represented using  $A(e^{-i\Omega})$  such that

$$a(n) = \frac{1}{2\pi} \int_0^{2\pi} A(e^{-i\Omega}) e^{i\Omega n} d\Omega. \quad (2.35)$$

This is called the discrete sequence representation of the continuous spectrum by the inverse Fourier transform. Here, note the orthogonality of the complex function  $e^{i\Omega n}$  such that

$$\frac{1}{2\pi} \int_0^{2\pi} e^{i\Omega(n-m)} d\Omega = \delta(n-m). \quad (2.36)$$

Suppose that the signal  $a(n)$  has a finite length  $N$ , so that the Fourier transform might be written as

$$A(e^{-i\Omega}) = \sum_{n=0}^{N-1} a(n) e^{-i\Omega n}. \quad (2.37)$$

The inverse Fourier transform can be derived for finite-length sequence  $a(n)$  when  $n = 0, \dots, N-1$ ; nevertheless, it might be interesting to see how the sequence is extended under the inverse Fourier transform with  $|n| > N-1$ . Substituting Eq. 2.37 for  $A(e^{-i\Omega})$  into Eq. 2.35 yields

$$\begin{aligned} \hat{a}(n) &= \frac{1}{2\pi} \int_0^{2\pi} \left( \sum_{m=0}^{N-1} a(m) e^{-i\Omega m} \right) e^{i\Omega n} d\Omega \\ &= \sum_{m=0}^{N-1} \frac{a(m)}{2\pi} \int_0^{2\pi} e^{i\Omega(n-m)} d\Omega \\ &= \sum_{m=0}^{N-1} a(m) \delta(n-m) \\ &= 0 \end{aligned} \quad (2.38)$$

for  $|n| > N-1$ . Thus, the discrete sequence representation by the Fourier transform implicitly assumes that the extended sequence  $a(n)$  vanishes outside the finite interval over which the sequence is given. For signal analysis, this is an important constraint placed on sequences based on their Fourier transforms.

### 2.4.2 Fourier Transform of Real Causal Sequence

Consider a real sequence  $a(n)$ . Its Fourier transform is given by a complex function of the angular frequency  $\Omega$  that is normalized from 0 to  $2\pi$ . Taking the real and imaginary parts of the Fourier transform,

$$\Re[A(e^{-i\Omega})] = A_r(e^{-i\Omega}) = \sum_{n=0}^{N-1} a(n) \cos \Omega n \quad (2.39)$$

$$\Im[A(e^{-i\Omega})] = A_i(e^{-i\Omega}) = - \sum_{n=0}^{N-1} a(n) \sin \Omega n \quad (2.40)$$

are obtained, respectively. Thus the real (imaginary) part is an even (odd) function of  $\Omega$ .

However, the Fourier transform is not always a complex function. Suppose  $a_e(n)$  is an even sequence defined by

$$a_e(n) = a_e(-n). \quad (2.41)$$

Then, the Fourier transform  $A_e(e^{-i\Omega})$  becomes

$$A_e(e^{-i\Omega}) = \sum_{n=0}^{N-1} a_e(n) \cos \Omega n. \quad (2.42)$$

That is, the Fourier transform is a real function for a real and even sequence. In contrast, suppose  $a_o(n)$  is an odd sequence defined by

$$a_o(n) = -a_o(-n), \quad (2.43)$$

and  $a_o(0) = 0$ . Then, the Fourier transform  $A_o(e^{-i\Omega})$  becomes

$$A_o(e^{-i\Omega}) = -i \sum_{n=0}^{N-1} a_o(n) \sin \Omega n. \quad (2.44)$$

These results dictate the relationship between the real and imaginary parts of the Fourier transform with the respective even and odd parts of a real causal sequence, the one being derived from the other.

Consider a real causal sequence as in the following:

$$x_c(n) = \begin{cases} x_c(n) & n \geq 0 \\ 0 & n < 0. \end{cases} \quad (2.45)$$

Such sequences can be expressed as the sum of even and odd sequences such that

$$x_c(n) = x_e(n) + x_o(n) \quad (2.46)$$

where  $x_e(n)$  is an even sequence and  $x_o(n)$  is an odd sequence. Namely, one can define the even and odd parts explicitly here by

$$x_e(n) = (x_c(n) + x_c(-n))/2 \quad (2.47)$$

$$x_o(n) = (x_c(n) - x_c(-n))/2. \quad (2.48)$$

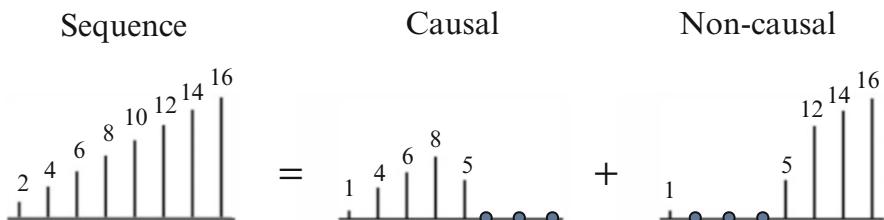
The real causal sequence can be represented by the inverse Fourier transform as

$$\begin{aligned} x_c(n) &= \frac{1}{2\pi} \int_0^{2\pi} X_c(e^{-i\Omega}) e^{i\Omega n} d\Omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} X_{cr} \cos \Omega n d\Omega \\ &\quad + \frac{1}{2\pi} \int_0^{2\pi} -X_{ci} \sin \Omega n d\Omega \\ &= x_e(n) + x_o(n). \end{aligned} \quad (2.49)$$

Specifically, the even (odd) sequence is obtained from the real (imaginary) part of the Fourier transform. Consequently, the even and odd sequences of a real causal sequence can be separated from each other. In turn, the results stated above indicate that the real and imaginary parts of the Fourier transform for a real causal sequence can also be separated from each other. The decomposition into even and odd sequences is possible for a non-causal sequence as illustrated in Fig. 2.2. It explains why the Fourier transform of a sequence is a complex function.

### 2.4.3 Fourier Transform for an Analytic Sequence

Sampled sequences are real; however, complex sequences are also frequently used in signal analysis. Recalling the complex representations of the trigonometric functions,



**Fig. 2.2** Decomposition of sequence (period = 8) into causal and non-causal sequences from Fig. 14.3 in [4]

$$\cos \Omega_0 n = \frac{1}{2} (e^{i\Omega_0 n} + e^{-i\Omega_0 n}) \quad (2.50)$$

$$\sin \Omega_0 n = \frac{1}{2i} (e^{i\Omega_0 n} - e^{-i\Omega_0 n}). \quad (2.51)$$

This representation enables real sinusoidal sequences to be interpreted as complex exponentials with positive and negative frequencies. Its Fourier transform of a real sinusoidal sequence is simply  $\frac{1}{2}\delta(\Omega \pm \Omega_0)$ . Thus, if only the positive (or negative) frequency component is taken, the respective complex sinusoidal sequence  $e^{\pm i\Omega_0 n}$  is reconstructed. This type of representation of a sequence is called the analytic representation of a real sequence.

The analytic form  $z(n)$  of a real sequence  $x(n)$  is obtained by setting

$$Z(e^{-i\Omega}) = \begin{cases} 2X(e^{-i\Omega}), & 0 < \Omega < \pi, \\ X(e^{-i\Omega}), & \Omega = 0, \pi \\ 0, & \pi < \Omega < 2\pi \end{cases} \quad (2.52)$$

where  $X(e^{-i\Omega})$  denotes the Fourier transform of  $x(n)$ . The inverse Fourier transform of  $Z(e^{-i\Omega})$  confirms that the real part of  $z(n)$  is the same as  $x(n)$ . Moreover, the inverse Fourier transform becomes  $z(n) = x(n) + iy(n)$  where

$$y(n) = \frac{1}{\pi} \int_0^\pi (X_r(e^{-i\Omega}) \sin n\Omega + X_i(e^{-i\Omega}) \cos n\Omega) d\Omega. \quad (2.53)$$

The analytic representation is sometime called the causal spectral representation.

In general, an analytic sequence takes a complex form. Thus, it can also be represented in polar form using magnitude and phase angle

$$z(n) = x(n) + iy(n) = |z(n)|e^{i\theta(n)} \quad (2.54)$$

where

$$|z(n)|^2 = x^2(n) + y^2(n) \quad (2.55)$$

and

$$\theta(n) = \tan^{-1} \frac{y(n)}{x(n)}. \quad (2.56)$$

Here,  $|z(n)|$  is called the instantaneous magnitude or the instantaneous envelope, and  $\theta(n)$  is called the instantaneous phase. Consequently, a real sequence can be expressed as

$$x(n) = \Re[z(n)] = |z(n)| \cos \theta(n) \quad (2.57)$$

using the instantaneous magnitude (envelope) and phase. The instantaneous magnitude or envelope represents temporal dynamics of sequences, whereas the instantaneous phase highlight spectral fine structure.

#### 2.4.4 Exponential Windowing and z-Transform

Recalling the  $z$ -transform of a sequence  $a(n)$ , i.e.,  $A(z^{-1}) = \sum_n a(n)z^{-n}$ . Putting  $z = r e^{i\Omega}$  where  $r = |z|$ , and substituting for  $z$  in  $A(z^{-1})$ , then

$$\begin{aligned} A(r^{-1}e^{-i\Omega}) &= \sum_n a(n)r^{-n}e^{-i\Omega n} \\ &= \sum_n a(n)w(n)e^{-i\Omega n} \end{aligned} \quad (2.58)$$

is derived, where  $w(n) = r^{-n} = e^{-n \log_e r}$ . Thus, the  $z$ -transform can be interpreted as the Fourier transform of the windowed sequence  $w(n)a(n)$ . For the equation above,  $r$  denotes the radius of the circle on which the  $z$ -transform is obtained. This windowing method is useful for the pole/zero analysis of the transfer function [6].

## References

1. R. Nelson, *Probability, Stochastic Processes, and Queuing Theory* (Springer, 1995)
2. S. Lang, *Complex Analysis* (Springer, 1999)
3. M. Tohyama, T. Koike, *Fundamentals of Acoustic Signal Processing* (Academic Press, 1998)
4. M. Tohyama, *Sound and Signals* (Springer, 2011)
5. H.S. Black, Stabilized feedback amplifiers. *Bell Syst. Tech. J.* **13**(1), 1–18 (1934)
6. M. Tohyama, R.H. Lyon, Zeros of a transfer function in a multi-degree-of-freedom system. *J. Acoust. Soc. Am.* **86**(5), 1854–1863 (1989)

# Chapter 3

## Temporal and Spectral Characteristics of Discrete Sequence

The temporal nature of a sequence is observed in narrow-band envelopes or frame-wise spectra. The envelopes represent the temporal changes in macroscopic structures, whereas the frame-wise spectral properties can be interpreted as related to the time-dependent fine structure of a sequence. The local behavior of phase spectral records or group delay is crucial in the construction of the envelopes. Inspired by the methods used for group delay, triangular window functions are introduced to divide a sequence into short frames. Consequently, the differential (averaging) properties of the spectral functions are formulated in terms of triangular with a positive (negative) slope windowing. Moreover, the short-term auto-correlations are obtained keeping its non-negative Fourier transform. Taking the causal part of the short-term auto-correlation sequence yields the complex power spectral function, the magnitude of which gives the temporal change in the power spectral envelope.

### 3.1 Magnitude and Phase of Spectrum

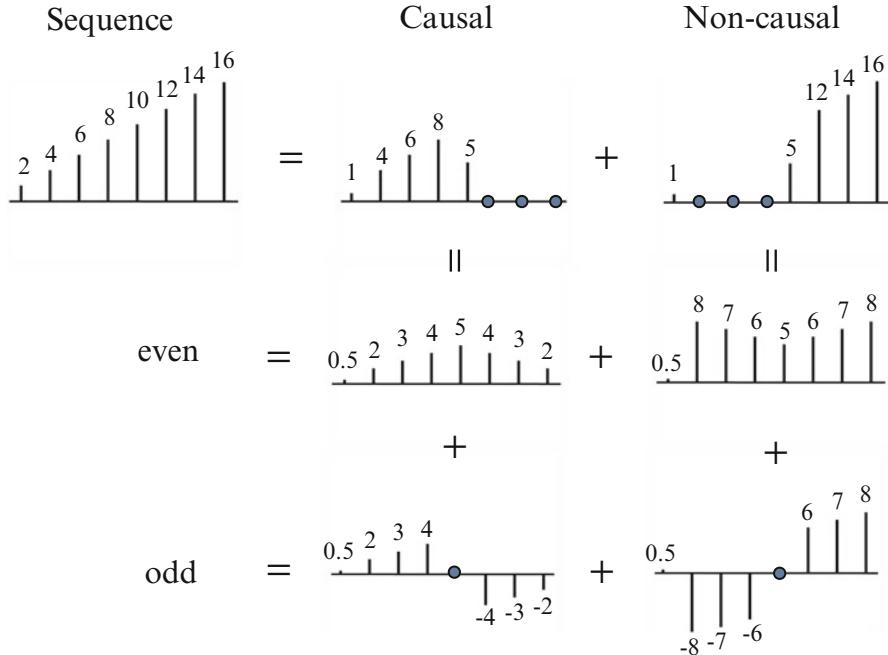
#### 3.1.1 *Magnitude and Phase Spectrum of Real Sequence*

As well as its real and imaginary components, complex numbers can be represented in polar form using the magnitude and phase. Similar to the properties of real and imaginary parts of a complex spectrum, the magnitude and phase of a spectrum are respectively the even and odd functions of a real sequence. Consider a real even sequence  $x_e(n)$ , and a real odd sequence  $x_o(n)$ . The spectrum of  $x_e(n)$ ,  $X_e(e^{-i\Omega})$  is a real even function, whereas, for the real odd sequence, the spectrum  $X_o(e^{-i\Omega})$  is pure imaginary. A real causal sequence can be expressed as a summation of even and odd sequences as illustrated in Fig. 3.1 on the following page. Thus, the

---

The original version of this chapter was revised.

An erratum to this chapter can be found at DOI [10.1007/978-4-431-54424-1\\_10](https://doi.org/10.1007/978-4-431-54424-1_10)



**Fig. 3.1** Decomposition of sequence (period = 8) into odd and even sequences from Fig. 14.3 in [7]

Fourier transform is also given by  $X_e(e^{-i\Omega}) + X_o(e^{-i\Omega})$  for  $x_c(n) = x_e(n) + x_o(n)$ . However, note that the spectral magnitude is written as

$$|X_c(e^{-i\Omega})| = \sqrt{X_e^2(e^{-i\Omega}) + (\Im[X_o(e^{-i\Omega})])^2} \quad (3.1)$$

and the phase spectrum becomes

$$\theta_c(e^{-i\Omega}) = \tan^{-1} \frac{\Im[X_o(e^{-i\Omega})]}{X_e(e^{-i\Omega})}. \quad (3.2)$$

### 3.1.2 Power Spectral Density and Auto-correlation Sequence

The square of the magnitude of the spectrum is called the power spectral density function of its sequence. Here one half of the power spectral density function can be interpreted as the power of a sinusoidal component. Recall the auto-correlation sequence  $r(n)$ ; then, introducing the  $z$ -transform  $R(z^{-1})$  into  $r(n)$  yields

$$R(z^{-1}) = X(z^{-1})X(z), \quad (3.3)$$

which can be obtained for a real sequence  $x(n)$ . Therefore, the auto-correlation sequence  $r(n)$  for  $x(n)$  can be expressed using the power spectral density (or the square of the magnitude spectrum) of  $x(n)$ , in particular,

$$\begin{aligned} R(e^{-i\Omega}) &= X(e^{-i\Omega})X(e^{i\Omega}) \\ &= |X(e^{-i\Omega})|^2. \end{aligned} \quad (3.4)$$

This result says that in general a sequence is not expressible in terms of its power spectral density with null-phase spectrum. The auto-correlation sequence is an even sequence, because the power spectral density is an even function.

The auto-correlation, however, is useful for period analysis of a sequence. An important and informative signal of a periodic sequence is its fundamental period. The human ear, for example, is very sensitive to changes in the fundamental period. However, how humans sense pitch in the varieties of sound is still an open research question [1].

In general, the pitch is represented by a sinusoidal frequency that gives the same pitch as that for the test sequence. The sinusoidal frequency corresponding to the pitch, is closely related to the fundamental frequency, that is, the inverse of the fundamental period. Therefore, period analysis is the first step in the analysis of the signal signature of a periodic sequence.

Suppose a periodic sequence  $x(n)$  of period  $N$  is given; that is, elements of the sequence satisfy the relation

$$x(n) = x(n + pN) \quad (3.5)$$

where  $p$  is an integer. Thus, a periodic sequence can be written as

$$x(n) = x_0 * u(n) \quad (3.6)$$

where  $x_0(n)$  denotes a single cycle of the periodic sequence, and  $u(n)$  is a periodic unit-pulse train with period  $N$  mathematically written in the form

$$u(n) = \sum_p \delta(n - pN). \quad (3.7)$$

Taking the  $z$ -transform of  $u(n)$  gives

$$U(z^{-1}) = \delta(z^{-N} - 1). \quad (3.8)$$

Therefore, the  $z$ -transform of a periodic sequence becomes

$$X(z^{-1}) = X_0(z^{-1})\delta(z^{-N} - 1) \quad (3.9)$$

where  $X_0(z^{-1})$  denotes the  $z$ -transform of  $x_0(n)$ . The above result gives the power or squared magnitude spectrum as a line-spectral sequence

$$|X(e^{-i\Omega})|^2 \Delta\Omega = P_{0s}(e^{-i\Omega}) \quad (3.10)$$

where  $e^{-i\Omega N} = 1$  and  $\Delta\Omega = 2\pi/N$ . Consequently, the auto-correlation sequence of the periodic sequence can be expressed as

$$\begin{aligned} r(n) &= \frac{1}{2\pi} \int_0^{2\pi} |X_0(e^{-i\Omega})|^2 \delta(e^{-i\Omega N} - 1) e^{i\Omega n} d\Omega \\ &= \sum_{k=0}^{N-1} P_0(e^{-i\frac{2\pi k}{N}}) \cos \frac{2\pi k n}{N} \end{aligned} \quad (3.11)$$

where  $P_0(e^{-i\Omega})$  denotes the power spectral density of a single cycle of the sequence, and  $P_{0s}(e^{-i\Omega}) = P_0(e^{-i\frac{2\pi k}{N}})$  represents the squared magnitude spectrum sampled at  $e^{-i\Omega N} = 1$  from the power spectral density function.

The result above indicates that the auto-correlation sequence is also periodic with period  $N$ ; in addition, the auto-correlation sequence takes its maximum at  $n = 0$ , which is periodically repeated at every interval of  $N$ . In contrast, the maximum of a single cycle of sequence  $x_0(n)$  is taken at  $n = n_0$  depending on the phase spectral property. This explains why the auto-correlation sequence is used for period analysis; explicitly, the auto-correlation sequence composed of zero-phase spectral components takes its maximum always at  $n = 0$ , and thus helps to estimate the fundamental period more easily.

The initial value of the auto-correlation sequence  $r(0)$  reduces to the total sum of the sampled squared magnitude spectra expressed in terms of its density over a single cycle of the sequence. In general, the total squared sum of a sequence is expressed using its power spectral density such that

$$\sum_{n=0}^{N-1} x^2(n) = \frac{1}{2\pi} \int_0^{2\pi} |X(e^{-i\Omega})|^2 d\Omega. \quad (3.12)$$

The auto-correlation sequence is usually normalized by  $r(0)$  so that the maximum can be normalized to unity.

The auto correlation sequence is a basic tool in the period analysis of a sequence. In particular, the fundamental period of a periodic sequence can be estimated by the auto-correlation analysis, even if the sequence does not contain the fundamental; that is, the sequence is composed of only harmonics without the fundamental. The pitch perception for such a sequence is called the pitch of the missing

fundamental [1]. The fundamental can still be heard from within the pitch for the sequence of the missing fundamental. This fact implies that the perception of pitch might be an attribute mainly given in the time domain, rather than the frequency domain.

The fundamental is a basic component of a musical tone; the pitch is, in principle, independent of the musical instrument producing the tones. However, as shown by the relationship between the auto-correlation and the power spectral density, the auto-correlation is strongly dependent on the power spectral property of the sound produced. Therefore, the spectral-free estimation of the fundamental is difficult depending on the auto-correlation sequence [2].

### ***3.1.3 Flat Magnitude Spectral Function and Its Phase Correlation***

The auto-correlation sequence was constructed from only the power spectral density independent of the phase spectral function. In contrast, a sequence constructed with only the phase information, discarding the magnitude spectral information, is also possible. The power or magnitude spectral property has been the main focus in signal analysis, whereas the phase spectral characteristics have received much less attention. This is because the power or magnitude spectral characteristic is the most significant factor governing human perception of speech under normal conditions in a wide range of situations.

However, in audio applications, so-called all-pass filters such as reverberators [3] are for historical reasons quite important tools. In addition, it is still interesting to uncover how the phase spectral characteristics makes up and enriches the perceptual signals of a sequence [2, 4–7]. Now, taking the Fourier transform, instead of the auto-correlation, of a real squared sequence gives

$$P_{xx}(e^{-i\Omega}) = \sum_{n=0}^{N-1} x^2(n)e^{-i\Omega n}. \quad (3.13)$$

Thus, rather than being related to the power spectral density, the spectrum of a squared sequence is related to the envelope of the sequence. The envelope of a sequence is defined formally by the instantaneous magnitude; however, other definitions are also possible, in particular, from a perceptual or qualitative point of view [8]. Taking the lower frequency components of the squared sequence is a possible way to derive the envelope and its spectrum. The spectral property of the envelope is as an important signal signature of a sequence as the spectrum of the sequence itself [2, 9, 10]. Interestingly, envelope spectral information remains in the phase spectral properties even after discarding the magnitude spectral information of the sequence.

The Fourier transform of a squared sequence can be rewritten using the Fourier transform of the original sequence  $x(n)$ . Recall the expression for the generating function for a sequence that is given as a convolution of two sequences; the generating function of sequence  $c(n) = a * b(n)$  is written as  $C(X) = A(X)B(X)$ . Similarly, if the sequence is written as a product instead of the convolution,  $c(n) = a(n)b(n)$  then the generating function becomes  $C(X) = A * B(X)$  where the convolution is extended over continuous variables. Therefore, the Fourier transform of the real squared (self-product) sequence can be written as

$$\begin{aligned} P_{xx}(e^{-i\Omega}) &= X * X(e^{-i\Omega}) \\ &= \int_0^{2\pi} X(e^{-i\Omega'}) X(e^{-i(\Omega-\Omega')}) d\Omega' \\ &= X \otimes X^*(e^{-i\Omega}). \end{aligned} \quad (3.14)$$

The equation above indicates that the spectrum of the squared sequence is expressed as a spectral correlation. This fact implies the phase spectral properties might be important in the representation of the envelope rather than the magnitude spectral characteristics.

Consider the Fourier transform of sequence  $x(n)$ ,  $X(\Omega) = |X(\Omega)|e^{i\phi(\Omega)}$ ; here the variable is expressed simply by  $\Omega$  instead of  $e^{-i\Omega}$ . The spectrum above can be rewritten as

$$P_{xx}(\Omega) = \int_0^{2\pi} |X(\Omega')| |X^*(\Omega' - \Omega)| e^{i(\phi(\Omega') - \phi(\Omega' - \Omega))} d\Omega'. \quad (3.15)$$

By discarding the magnitude spectrum, or assuming that the magnitude spectrum is constant (flat magnitude spectrum), the equation above is reduced to

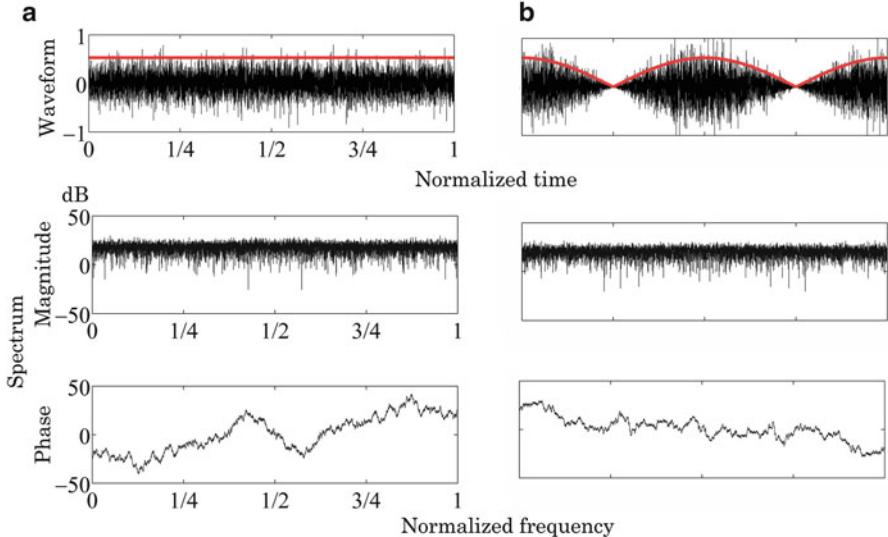
$$E_{xx}(\Delta\Omega) = C \int_0^{2\pi} e^{-i\tau(\Omega')\Delta\Omega} d\Omega' \quad (3.16)$$

where  $C$  is a constant and  $\tau(\Omega)$  denotes the group delay defined as

$$\tau(\Omega) = \frac{-d\phi(\Omega)}{d\Omega} \quad (3.17)$$

or, expanding as a Taylor series to first-order,

$$\phi(\Omega) - \phi(\Omega - \Delta\Omega) \cong -\tau(\Omega)\Delta\Omega. \quad (3.18)$$



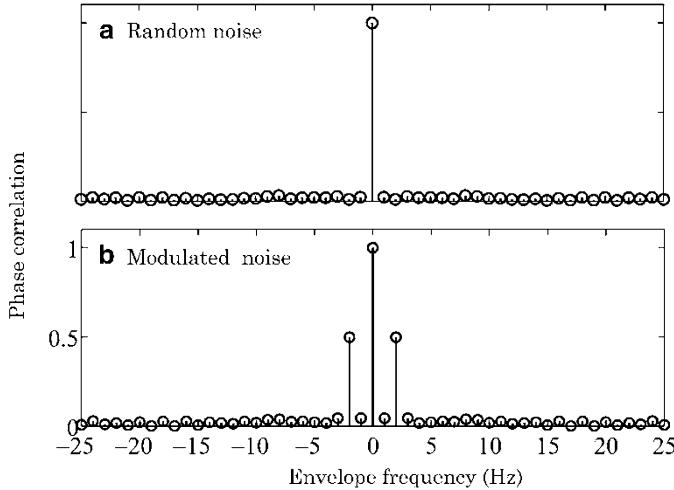
**Fig. 3.2** Random noise (a) and amplitude-modulated noise with a sinusoidal function (b); *top*: waveform, *middle*: magnitude spectrum, *bottom*: phase spectrum. All frequencies including the envelope frequency are normalized from Fig. 5 of [6]

The result above implies that the spectral property of the envelope (or the squared sequence) is hidden in the local behavior of the phase spectrum such as the group delay of the sequence, even if the magnitude spectrum is discarded or assumed flat.

The envelope spectrum might be related to the phase correlation, that is, the spectral correlation after discarding the magnitude spectrum. An amplitude-modulated random noise might be a good example to estimate the envelope frequency following a phase correlation analysis [6, 7]. Figure 3.2 shows examples of random noise (a) without modulation and (b) with a sinusoidal modulation. The sequence can be expressed as  $x_b(n) = e(n)x_a(n)$  where  $x_a(n)$  denotes the sequence associated with the random noise and  $e(n)$  that for the sinusoidal modulation envelope of angular frequency  $\Omega_e$ . It is clear to see (or listen to) the differences in the waveforms between these two sequences; however, the magnitude or the phase recorded samples, might not be helpful in characterizing the difference in the two sequences.

By estimating the envelope frequency, the phase correlation analysis could possibly discriminate between the modulated noise and the random noise. Reformulating the phase correlation defined by Eq. 3.16 on the preceding page in terms of statistical averaging gives

$$R_{ph}(\Omega) = E \left[ e^{i(\phi(\Omega') - \phi(\Omega' - \Omega))} \right] \quad (3.19)$$



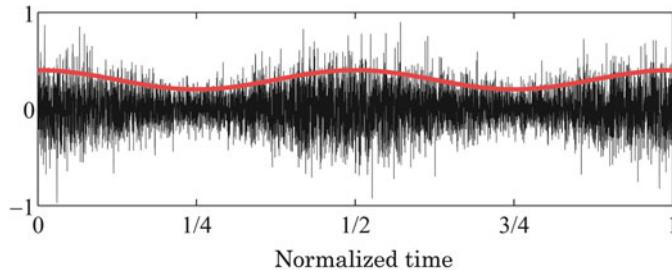
**Fig. 3.3** Phase correlation samples for the random noise (a) and modulated noise (b) from Fig. 6 of [6]

where  $E[ \cdot ]$  denotes the ensemble average. Figure 3.3 shows numerical samples of the phase correlation analysis for the random-noise samples shown in Fig. 3.2 on the previous page. The vertical axis in the figure shows the absolute of the phase correlation  $R_{ph}(\Omega)$ , whereas the horizontal represents the angular-frequency shift for the correlation analysis and thus indicates the envelope frequency. Clearly, there is no definite correlation in the random-phase spectral components for the random noise, as shown in Fig. 3.3a. In contrast, the envelope angular frequency  $2\Omega_e$  can be read off from the correlation for the modulated noise, as shown in panel (b). This result can be interpreted as indicating that every spectral pair separated by interval  $2\Omega_e$  is composed of in-phase components in the modulated noise, although the spectral components of the random noise are un-correlated random variables, as shown in panel (a). The envelope can be reconstructed from the inverse Fourier transform of the phase spectrum in Fig. 3.2b on the previous page, as shown in Fig. 3.4 on the facing page.

## 3.2 Frame-Wise Fourier Transform and Filter Bank

### 3.2.1 Frame-Wise Fourier Transform

Phase spectral information determines the envelope. This implies that signal signatures are not necessarily represented only by the power spectrum or the auto-correlation. On the other hand, by listening, the difference can be clearly discerned between random and modulated noise sequences. What is the main signal signature

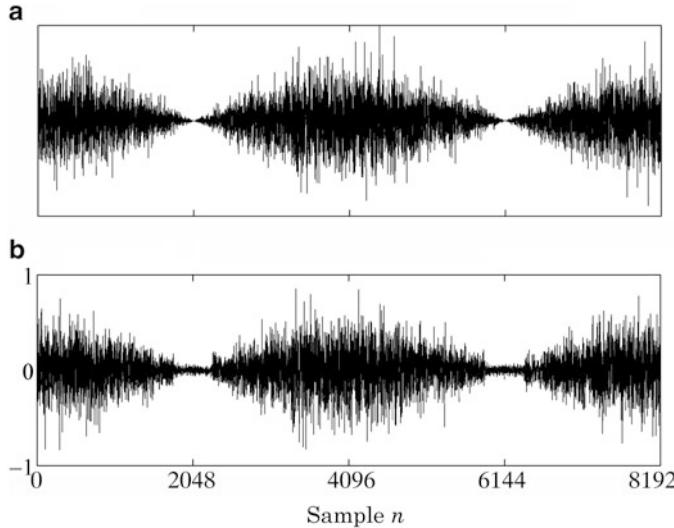


**Fig. 3.4** Envelope reconstruction from the phase spectrum shown in Fig. 3.2b on page 35 with the random-noise magnitude spectrum from Fig. 7 of [6]

essentially reconstructed from the phase spectral information? This might be a deep question delving into the perception of sound; but clearly, if sequences can be discerned by listening, the envelope seems one of the possible factors at present from a perceptual point of view. The envelope is an attribute of a waveform in the time domain, although, temporal changes in the waveform, or equivalently temporal dynamics in its sequence, contributes to the signal signature rather than a temporal average over the entire sequence.

A time waveform given in a time interval can be mutually transformed between the time and frequency domains by the Fourier transformation. Here, the independent variables are *time* for the time domain and *frequency* for the frequency domain. In contrast, in signal analysis, the time-frequency region is from a perceptual point of view introduced between the time and frequency domains using the frame-wise Fourier transformation. The frame-wise Fourier transform is defined as the Fourier transform performed over a very short interval (frame), which is taken from the entire interval for a given sequence; thus, every frame-wise Fourier transform might represent the temporal change of the spectrum. The frame-wise Fourier transform is conventionally expressed as  $X_l(e^{-i\Omega})$  for the  $l$ -th frame of a sequence  $x_l(n)$ . Here the frame number  $l$  is interpreted as a variable representing the temporal change as well as  $\Omega$  giving the angular frequency.

Figure 3.5 on the following page presents an example of frame-wise Fourier transforms for the modulated random noise shown in Fig. 3.2 on page 35. The waveforms in the figure are the modulated random noise (a) and reconstructed by frame-wise inverse Fourier transform using only the magnitude spectrum after discarding the records of phase in every frame (b), respectively. Given  $N$  as the length of the frame, the frame-wise Fourier transforms are taken for every  $N/2$  samples, and thus the reconstructed waveform is obtained by concatenating all the frame-wise inverse Fourier transforms, subject to triangle weighing applied for all concatenating frames resulting from the  $N/2$  overlapping frames. Interestingly, the envelope is almost perfectly reconstructed even after discarding the phase spectral information [6, 7]. Comparing both results for envelope recovery, that is, the phase spectrum for the entire record length shown in Fig. 3.4 and the magnitude spectrum for every frame-wise record shown in Fig. 3.5 on the following page, the inference is



**Fig. 3.5** Reconstructed waveform for modulated random noise by frame-wise Fourier transforms using the magnitude spectrum with the phase of wide-band random noise (frame-length is  $N = 256$ ). (a) Modulated random noise (original). (b) Reconstructed waveform

that the temporal characteristics governing the entire waveform can be expressed by the phase spectrum for the entire sequence, or the temporal (frame-wise) change in the records of magnitude of the frame-wise Fourier transforms [2,5,6]. The samples of envelope recovery suggest that hearing perception works from phase effects on a waveform.

The division of a sequence into short time frames is also called time windowing. Windowing can be implemented by applying a weighing sequence  $w(n)$  to an observed record  $x(n)$ . The frame-wise sequence made after windowing can be formally expressed using the windowing sequence in the manner

$$x_l(n) = x(n)w_l(n) \quad (3.20)$$

where  $x_l(n)$  denotes the frame-wise sequence for the  $l$ -th frame, and  $w_l(n)$  is the weighing sequence for the  $l$ -th frame. Recalling the Fourier transform of a sequence given by the product of two sequences (Eq. 3.14 on page 34), then, by analogy, the Fourier transform of the frame-wise sequence defined by Eq. 3.20 becomes

$$X_l(e^{-i\Omega}) = X * W_l(e^{-i\Omega}) \quad (3.21)$$

where  $X(e^{-i\Omega})$  denotes the Fourier transform for the entire sequence  $x(n)$ , and  $W_l(e^{-i\Omega})$  denotes the Fourier transform of  $w_l(n)$ . The above equation indicates that the frame-wise spectrum can be derived from the spectrum of the entire sequence by the convolution of the spectral functions, provided the windowing sequence is known.

### 3.2.2 Sub-band Filters

The frame-wise Fourier transformation can also be introduced by dividing the frequency into short frequency intervals using a filtering function  $H_k(e^{-i\Omega})$  instead of the time-windowing function  $w_l(n)$ . The  $k$ -th frequency interval divided by the  $k$ -th filtering function is called the  $k$ -th sub-band for the entire frequency band.

Suppose the entire spectrum  $X(e^{-i\Omega})$ , and thus the sub-band spectrum, is given by

$$X_k(e^{-i\Omega}) = X(e^{-i\Omega}) \cdot H_k(e^{-i\Omega}) \quad (3.22)$$

for the  $k$ -th sub-band. Therefore, taking the inverse Fourier transform for the sub-band spectrum, the sub-band time-sequence  $x_k(n)$  can be obtained as

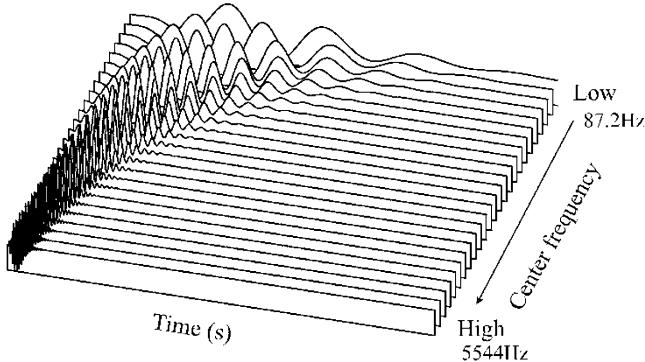
$$x_k(n) = x * h_k(n) \quad (3.23)$$

where  $h_k(n)$  denotes the impulse response for the  $k$ -th sub-band filtering function. The sub-band sequence is obtained from the entire band sequence following application of the filtering function.

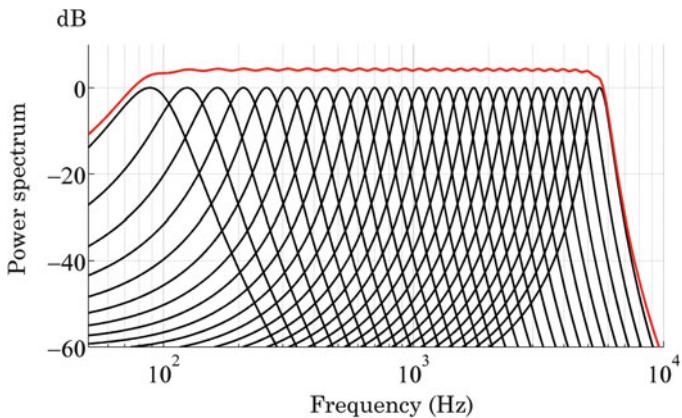
Exploring the hidden signals in a sequence from a perceptual point of view is customarily performed by time windowing and/or spectral filtering. The set of sub-band filters is called a filter bank [11]. A 1/4 octave-band filter is a typical example to mimic the hearing function that accounts for masking properties [1, 12–15]. A single octave band denotes the frequency interval that sets the ratio  $f_H/f_L = 2$ , where  $f_L$  and  $f_H$  give the lowest and highest frequencies in the interval. Following the definition, a 1/4 octave-band has a frequency interval for  $f_H/f_L = 2^{1/4}$  or  $f_L = f_c \cdot 2^{-1/8}$  and  $f_H = f_c \cdot 2^{1/8}$  where  $f_c$  denotes the center frequency of the band. Consequently, the logarithmic sub-band interval becomes wider as the center frequency increases.

### 3.2.3 Perfect Reconstruction Filters

Figure 3.6 on the following page displays examples of impulse response data for sub-band filters [16]. Note that the duration of the impulse response data shortens (is more brief) as the center frequency increases. This implies that the temporal or



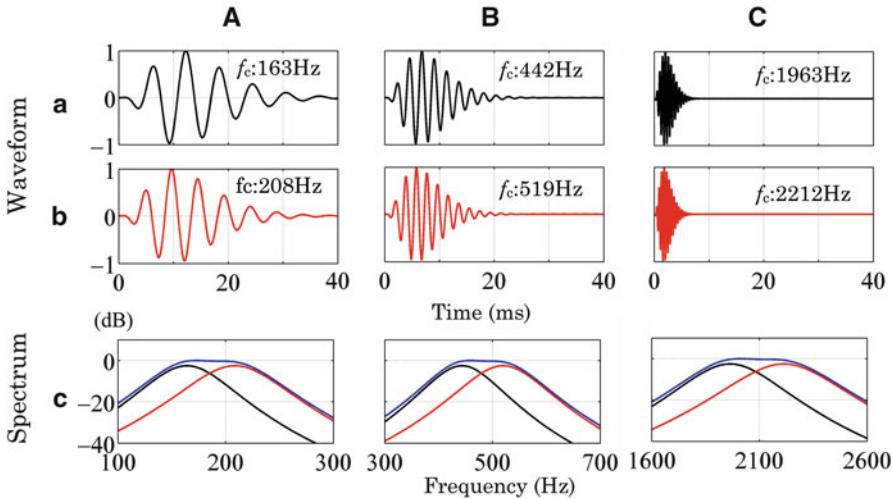
**Fig. 3.6** Example of impulse response records for 1/4 octave-band filter bank [16]



**Fig. 3.7** Power spectral characteristics for sub-band responses shown in Fig. 3.6 including whole-of-band power spectrum (red line)

timing signals of a sequence might be mainly sensed by the sub-band sequences through the higher frequency rather than the lower frequency bands, if such a sub-band filter bank underpins hearing models. In contrast, Fig. 3.7 illustrates the power spectral responses of the sub-band filters. The frequency range becomes wider as the center frequency goes higher.

An important design condition for filter banks is that the filtering process yields no spectral loss. However, other conditions for the perfect reconstruction filter are difficult to realize [11]. A perfect reconstruction filter would mean that the whole-of-band impulse response, which is directly obtained without the filter bank, can be reconstructed even by summing over the sub-band impulse responses. As seen in Figs. 3.6 and 3.8 on the facing page, the sub-band spectral responses overlap each other outside the main frequency range among adjacent sub-bands. Even if the overlap creates few severe distortions in the power spectral response, the original



**Fig. 3.8** Spectral responses (blue) between adjacent sub-bands(red and black)

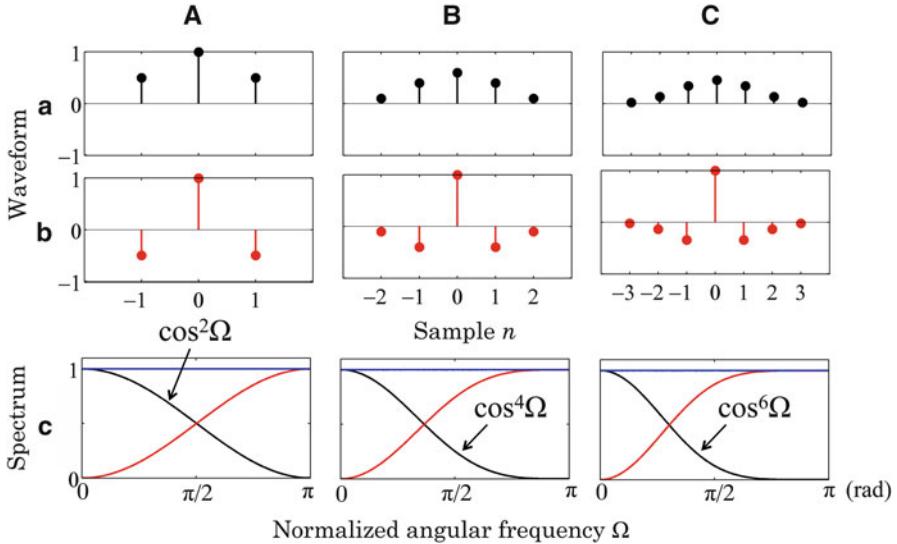
sequence is not always reconstructed. This is because reconstruction of the sequence is sensitive to the phase spectral condition for the overlapped frequency responses. Perfect reconstruction removing the time delay can be expected if the following relation holds:

$$h_0(n) = \sum_{k=1}^K a_k h_k(n) \quad (3.24)$$

where  $h_0(n)$  denotes the whole of band response,  $h_k(n)$  is the impulse response for the  $k$ -th sub-band filter, and  $a_k$  a real number for the  $k$ -th sub-band. The formulation above can be interpreted as the construction of the original response as a linear combination of sub-band impulse responses, if the filter bank provides perfect reconstruction.

Figure 3.9 on the following page shows a simple example of the perfect reconstruction filter that divides a frequency band of interest into low- and high-pass sub-bands. The basis of impulse response of the low- and high-pass filters are given by  $h_{L0}(n) = 1, 2, 1$  and  $h_{H0}(n) = -1, 2, -1$ , respectively. The pair of these sub-band filters realize the perfect reconstruction. The theory and design methods for a perfect reconstruction filtering system are described in detail in [11].

Figure 3.10 on page 43 displays examples of sub-band waveforms of speech. The signal signature extracted from the whole-of-band speech waveform changes as the center frequency increases, and also when the frequency range of the sub-band widens. Similar to Fig. 3.6 on the preceding page, the temporal feature of speech appears as the frequency range widens. In contrast, in the low-frequency range, the periodic characteristics are mostly represented by the center frequencies.



**Fig. 3.9** Examples of perfect reconstruction

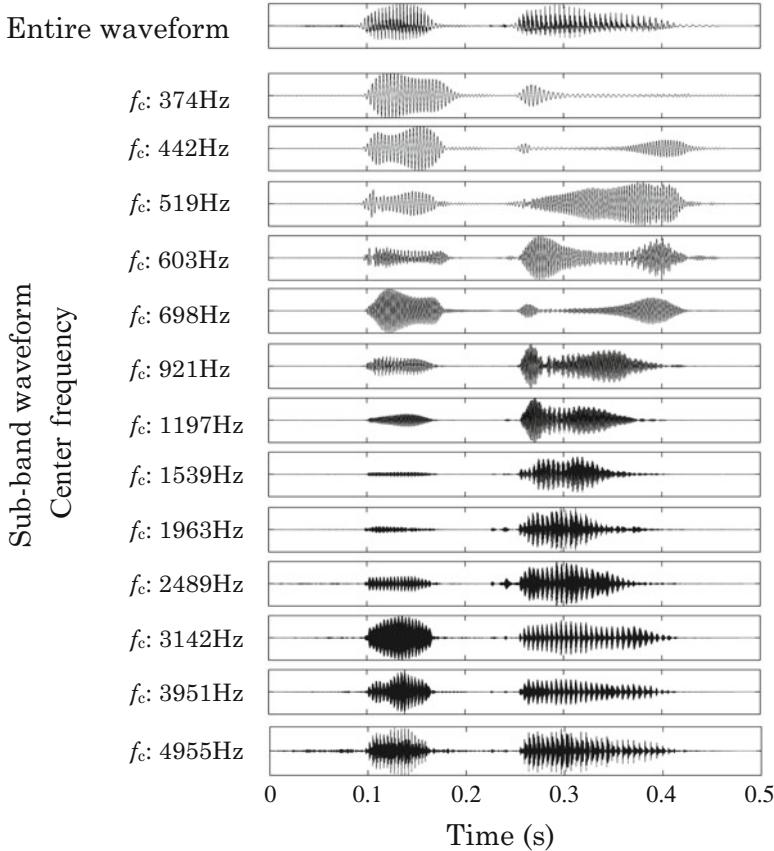
In addition, another important temporal feature can be observed in the envelopes of sub-band waveforms. The envelopes convey speech information such as intelligibility, and interestingly, even the microscopic temporal features of the waveform such as the pitch or the fundamental frequency can also be estimated from the sub-band envelopes [1, 9, 13, 14, 17].

### 3.3 Modulation Envelope and Group Delay

#### 3.3.1 Sinusoidal Modulation and Envelope

As described in Eq. 3.16 on page 34, the envelope spectrum is related to the group delay. The relationship between the envelope and the group delay can be intuitively understood by sinusoidal-amplitude modulation rather than from formulations such as Eq. 3.16 on page 34. Consider the three sinusoidal components  $\Omega_c \pm \Delta\Omega$  and  $\Omega_c$ . Setting magnitudes to unity for the central component and  $1/2$  for the side components, then the initial phases are  $\mp\Delta\phi$  for the sides and null at the center  $\Omega_c$ . The amplitude modulated sequence is then written as

$$\begin{aligned} x(n) &= (1 + \cos \Delta\Omega(n - \tau_{\Omega_c})) \cdot \sin \Omega_c n \\ &= e_M(n) \cdot e_c(n) \end{aligned} \quad (3.25)$$



**Fig. 3.10** Sub-band waveforms of speech observed through the filter bank shown in Fig. 3.6 on page 40

where  $\tau_{\Omega_c}$  denotes the group delay (in the units of the samples) at  $\Omega_c$ , that is,  $\tau_{\Omega_c} = -(-\Delta\phi/\Delta\Omega)$ , and  $e_M(n) = 1 + \cos \Delta\Omega(n - \tau_{\Omega_c}) \geq 0$ . Thus,  $e_M(n)$  is called the envelope of the amplitude modulation, whereas  $e_c(n)$  is the carrier.

The example above indicates that the group delay at the carrier frequency gives the delay of the envelope. Thus, the envelope is a temporal and macroscopic signature of a sequence, which is essentially related to the spectral fine structure. That is, the spectral frequency spacing around the central frequency determines the envelope frequencies, and the phase differences yield the delays of the envelopes.

It is interesting to see the peak of the envelope depending on the group delay [2]. Suppose an amplitude modulated sequence is written as

$$x(n) = e(n) \sin \Omega_c n$$

$$e(n) = K + \sum_{k=1}^K \cos(k\Delta\Omega n + \Delta\phi_k). \quad (3.26)$$

Here,  $\Delta\Omega$  is the angular frequency separation (spacing) such as  $\Omega_{\pm 1} = \Omega_c \pm \Delta\Omega, \dots, \Omega_{\pm K} = \Omega_c \pm K\Delta\Omega$ ,  $2\pi/\Delta\Omega$  gives the period of the envelope, and  $\Delta\phi_k$  denotes the phase difference of the  $k$ -th side component from the central component; the magnitudes are  $1/2$  for all side components, and thus the central component has magnitude  $K$ , the number of pairs of side components that make the envelope. If  $\Delta\phi_k$  is 0, independent of the envelope frequencies, then the peak of the envelope has maximum  $2K$  if

$$n = m \frac{2\pi}{\Delta\Omega}, \quad (3.27)$$

where  $m$  denotes an integer. In addition, if  $\Delta\phi_k$  is proportional to  $k$  so that  $\Delta\phi_k = k\Delta\phi$ , then by substituting  $\Delta\phi_k = k\Delta\phi$  for  $\Delta\phi_k$  in Eq. 3.26,

$$\begin{aligned} e(n) &= K + \sum_{k=1}^K \cos k\Delta\Omega \left( n - \frac{-\Delta\phi}{\Delta\Omega} \right) \\ &= \left( K + \sum_{k=1}^K \cos k\Delta\Omega (n - \tau_{\Omega_c}) \right). \end{aligned} \quad (3.28)$$

Thus, the peak of the envelope is at

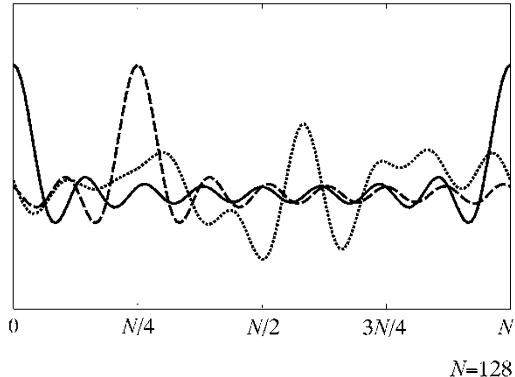
$$n = m \frac{2\pi}{\Delta\Omega} + \tau_{\Omega_c}. \quad (3.29)$$

The result above indicates that the location of peaks are delayed by the group delay given by  $\tau_{\Omega_c}$  at the center frequency. Then, if  $\Delta\phi = -l\pi/N$  and  $\Delta\Omega = \pi/N$ , the envelope is delayed by  $l$  samples from that under the 0-phase condition. The phase accumulation  $k\Delta\phi$  supposed above is called the linear phase.

In contrast, if the phase difference  $\Delta\phi_k$  changes irregularly between adjacent envelope frequencies as the phase difference for a pair of sinusoidal components in a random noise record, then the envelope has no distinguished peaks [2]. Figure 3.11 presents some samples of envelopes. The difference due to the phase conditions (or group delays) is clearly understood [2].

The example above gives an intuitive interpretation of the envelope and group delay for a modulated sequence that is composed of line-spectral components. In addition, the group delay is expressed in accord with the linear phase approximation, although the group delay is essentially defined on the continuous spectral function assumed to be differentiable with respect to the frequency. However, the group delay need not be well defined by the linear phase approximation. This fact implies that

**Fig. 3.11** Samples of envelopes under different group delay conditions where  $K = 8$ , period = 128, and  $\Delta\Omega = 2\pi/128$ , solid line: 0-phase condition (group delay = 0), broken line: linear phase (group delay) = 32, and dotted line: random phase



the envelope itself does not always provide good information because of the phase condition of the sinusoidal components.

Consider once more the three sinusoidal components but with phases  $\Delta\phi, 0, \Delta\phi$  added for the sinusoidal angular frequencies,  $\Omega_c - \Delta\Omega, \Omega_c, \Omega_c + \Delta\Omega$ . The constructed sequence is written as

$$\begin{aligned} x(n) &= a(n) \sin \Omega_c n + b(n) \cos \Omega_c n \\ &= c(n) \sin(\Omega_c n + \theta(n)) \end{aligned} \quad (3.30)$$

where

$$a(n) = 1 + \cos \Delta\phi \cos \Delta\Omega n \quad (3.31)$$

$$b(n) = \sin \Delta\phi \cos \Delta\Omega n \quad (3.32)$$

$$c(n) = \sqrt{a^2(n) + b^2(n)} \quad (3.33)$$

$$\theta(n) = \tan^{-1} \frac{b(n)}{a(n)}. \quad (3.34)$$

Note that the phase difference no longer yields a simple delay for the expected envelope  $c(n)$ . In addition, if the phase difference approaches  $\pi/2$ , then the envelope almost disappears [18]. That is, the expected envelope becomes

$$c(n) = \sqrt{1 + \cos^2 \Delta\Omega n} \quad (3.35)$$

when  $\Delta\phi = \pi/2$ [7,18]. Compare the two envelopes  $e_M(n)$  and  $c(n)$  of Eqs. 3.25 on page 42 and 3.35, respectively. Troughs making zeros are seen on the time axis  $n$  of the envelope represented by  $e_M(n)$ . In contrast,  $c(n)$  has no zeros anywhere on the time axis. This difference indicates that  $c(n)$  conveys less modulation information than  $e_M(n)$ .

Express the modulated envelope in the general form

$$e(n) = 1 + m \cos(\Omega_e n + \phi_e) \quad (3.36)$$

where  $\Omega_e$  and  $\phi_e$  are the angular frequency and the initial phase, respectively. Here,  $m(0 \leq m \leq 1)$  is called the modulation index [9]. Modulation of the envelopes becomes stronger as the modulation index approaches unity, whereas the modulated envelope is almost flat by losing the modulation signature as the modulation index approaches zero.

The envelope represents one of the temporal signatures of a time-sequence made by a group of clustered sinusoidal components around a central frequency, in which the phase characteristics can be approximated by a linear phase. Consequently, a delay in the envelope is expected from the group delay, and thus the fact that the envelope almost disappears because of this phase effect is interpreted as stemming from an ill-defined group delay.

In an acoustic system where a wave propagates from a source to another position (supposed to be an observation point) with separation  $r$ , the time delay in catching the wave from the source is given by

$$\tau_p = \frac{-\phi}{\omega} = \frac{r}{c_p} = r \frac{k}{\omega} \quad (3.37)$$

where  $\omega$  is the angular frequency of the sinusoidal wave,  $-\phi$  the phase difference between source and receiver,  $c_p$  the phase velocity, and  $k = \omega/c_p$  is called the wavenumber. If  $c_p$  depends on the frequency, then the system or the medium filling the system is called dispersive.

From an alternative perspective, the group delay is, a local property of the phase. The group delay between the waves at the source and observation positions is written as

$$\tau_g = \frac{-\partial\phi}{\partial\omega} = r \frac{\partial k}{\partial\omega} = \frac{r}{c_g} \quad (3.38)$$

where  $c_g = \partial\omega/\partial k$  is called the group velocity corresponding to the propagation speed of the envelope or the group of sinusoidal components in the system. If the system is not dispersive, then the two speeds are equal to each other [19, 20].

### 3.3.2 Group Delay for Real Discrete Sequence

The group delay is given by the first derivative of the phase. However, the group delay can be obtained for a real discrete sequence without directly performing differential calculus [21].

The phase is defined by

$$\phi(e^{-i\Omega}) = \tan^{-1} \frac{X_i(e^{-i\Omega})}{X_r(e^{-i\Omega})} \quad (3.39)$$

where  $X_r(e^{-i\Omega})$  and  $X_i(e^{-i\Omega})$  denote the real and imaginary parts of the Fourier transform  $X(e^{-i\Omega})$  for a real and discrete sequence  $x(n)$ . Thus, the derivative of  $\phi(e^{-i\Omega})$  with respect to  $\Omega$  is

$$\frac{-d}{d\Omega} \phi(e^{-i\Omega}) = \frac{X_i X'_r - X'_i X_r}{|X|^2}. \quad (3.40)$$

Recalling the definition for the Fourier transform of  $x(n)$ , then

$$\begin{aligned} X'_r(\Omega) &= - \sum_{n=0}^{N-1} n x(n) \sin \Omega n \\ &= Y_i(\Omega) \end{aligned} \quad (3.41)$$

$$\begin{aligned} X'_i(\Omega) &= - \sum_{n=0}^{N-1} n x(n) \cos \Omega n \\ &= -Y_r(\Omega) \end{aligned} \quad (3.42)$$

where  $Y(e^{-i\Omega})$  denotes the Fourier transform of the sequence  $y(n) = nx(n)$ ; explicitly,

$$Y(e^{-i\Omega}) = -X'_i(e^{-i\Omega}) + iX'_r(e^{-i\Omega}). \quad (3.43)$$

Consequently, the group delay can be expressed without using the derivatives such that

$$\tau(\Omega) = \frac{X_i Y_i + X_r Y_r}{|X|^2}. \quad (3.44)$$

The relations represented by Eqs. 3.41 and 3.42 indicate that windowing effects caused by applying the triangular time window can be interpreted as taking the spectral derivatives in the frequency domain. Consequently, the spectral differences between the adjacent frequencies are emphasized by applying the triangular time window on the original sequence.

### 3.3.3 Triangle Windows and Spectral Derivatives

In the previous subsection, the group delay was derived based on the relation between the triangular window and spectral derivatives without performing differential calculus. In addition to the group delay, if the power spectrum is taken for the windowed sequence  $y(n) = nx(n)$ , then

$$\begin{aligned} P_y(e^{-i\Omega}) &= Y_r^2(e^{-i\Omega}) + Y_i^2(e^{-i\Omega}) \\ &= X_r'^2(e^{-i\Omega}) + X_i'^2(e^{-i\Omega}) \\ &= P_{x'}(e^{-i\Omega}) \end{aligned} \quad (3.45)$$

is obtained following Eqs. 3.41 on the previous page and 3.42 on the previous page. Thus, the power spectrum for the windowed sequence  $y(n)$  is equal to the power spectrum for the derivatives of the original spectrum [22]. The temporal window with positive slope is a simplified model for the forgetting function of the ear. Although this function must not be represented simply by the temporal windowing effect, temporal windowing reveals a microscopic structure originating from the derivatives of the spectrum as well as the group delay.

In addition to the triangular window  $w(n)$  with positive slope, one with negative slope can be defined by reversing the time axis. However, spectral differential effects are no longer expected using the negative slope triangular window. For a simple and intuitive understanding of the windowing effect, consider an even sequence of negative-slope triangular windows:

$$\overline{w^-}(n) = \begin{cases} N - 1 + n & -(N - 1) \leq n < 0 \\ N - 1 - n & 0 \leq n \leq N - 1. \end{cases} \quad (3.46)$$

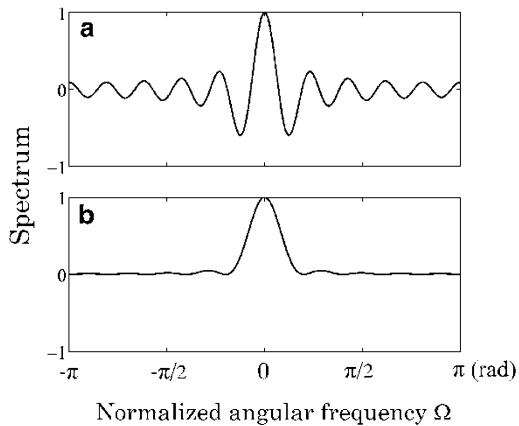
Its Fourier transform is a real non-negative function. In contrast, consider the positive-slope triangular windows in an even window sequence such that

$$\overline{w^+}(n) = \begin{cases} -n & -(N - 1) \leq n < 0 \\ n & 0 \leq n \leq N - 1. \end{cases} \quad (3.47)$$

The Fourier transform is a real function, but alternates with both positive and negative values as illustrated in Fig. 3.12 on the facing page.

The spectral function of a windowed sequence can be expressed as the convolution of spectral functions for the window and the original sequences. The differential effect resulting from the positive-slope triangular window sequence is due to the alternating signs of its spectral functions. In contrast, spectral averaging effects are expected in negative-slope windowing because of the non-negative property of its Fourier transform.

**Fig. 3.12** Fourier transforms for symmetric triangular sequence with positive (a) and negative (b) slope



Schematics of the Fourier transforms of a sequence after triangular windowing with positive (a) and negative (b) slope are illustrated in Fig. 3.13 on the following page.

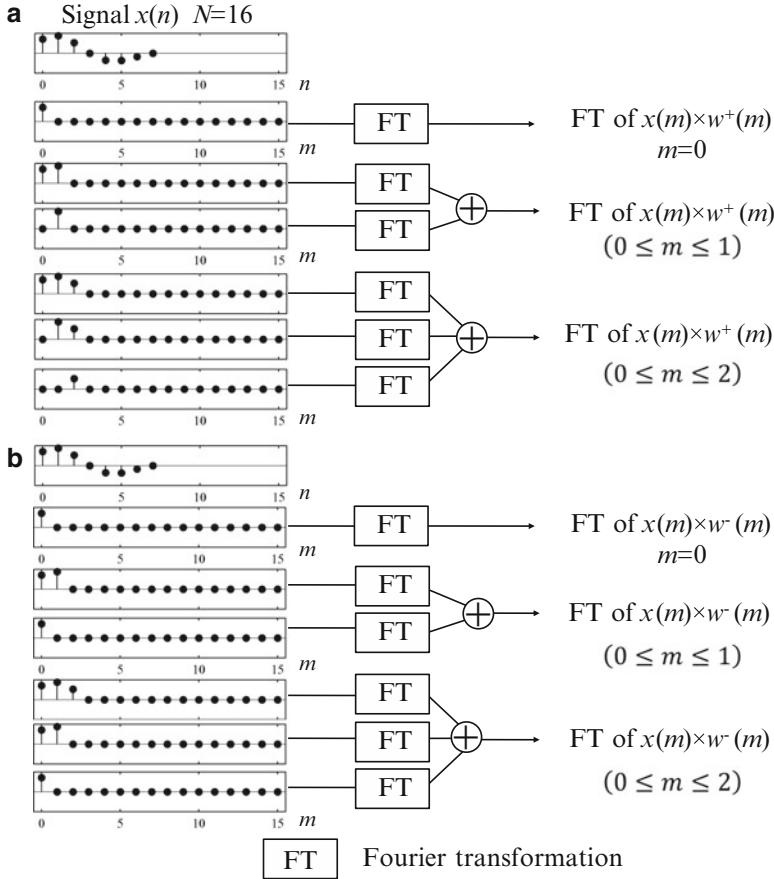
## 3.4 Triangular Windowing and Auto-correlation Sequence

### 3.4.1 Short-Term Auto-correlation Sequence

As described in Sect. 3.2 on page 36, temporal signals of sequences can be conveyed in the power spectral properties expressed by the frame-wise Fourier transform as well as the phase spectrum determined using a long record length [6]. The power spectrum obtained by the frame-wise Fourier transform is also called the short-term power spectrum. The short-term power spectral properties of a sequence, in addition, determine the short-term (or frame-wise) auto-correlation sequence that shows the periodic structure of the sequence in the time domain, as described in Sect. 3.1.2 on page 30. In particular, the importance of the auto-correlation sequence is intensively investigated in [23] from a perceptual point of view.

The auto-correlation sequence can be estimated from the inverse Fourier transform of the power spectrum of a given sequence. The inverse relation can be used to estimate the power spectrum as expressed by Eq. 3.4 on page 31. Take a frame-wise real causal sequence  $x(n)$  of record length  $N$  obtained from a complete sequence. The auto-correlation sequence of the frame-wise sequence, that is, the short-term auto-correlation can be defined as

$$r(n) = \sum_{m=0}^{N-1} x(m)x(m-n). \quad (3.48)$$



**Fig. 3.13** Schematics of frame-wise Fourier transforms and triangular windowing with positive (a) and negative (b) slope

The record length of the auto-correlation sequence is  $2N - 1$  and the correlation has an even sequence form for  $-(N - 1) \leq n \leq N - 1$ . Therefore, the Fourier transform of the auto-correlation sequence is a real non-negative function that gives the power spectral density function of the sequence.

However, estimating the short-term auto-correlation sequence sometimes poses a problem. The auto-correlation analysis is originally related to a method for random sequence analysis. In principle, the auto-correlation sequence  $r(n)$  defined for a frame-wise sequence is only a single sample of the record that can be expected to give an estimate of the auto-correlation sequence, if the ensemble average can be taken for the sample. The averaging is partially performed even in a sample of the auto-correlation by taking the summation over  $m$  of the products of  $x(m)x(m-n)$ . However, the number of the products that can be taken from a short-term (frame-wise) record of sequence decreases as the absolute time-lag  $|n|$  increases. That is,

the effect of averaging that might decrease the variances between samples weakens as time-lag  $|n|$  becomes large.

In contrast, if a long sequence can be taken to get a sample of the correlation, including long time-lag components, a problem still remains in estimating the short-term power spectrum. If only the short time-lag components are taken from the auto-correlation sequence to get an estimate of the short-term power spectrum, then its Fourier transform need not necessarily be non-negative in general; the Fourier transform would not be an estimator of the power spectrum in this case.

A practical way to avoid the difficulty is to apply a triangular window on the auto-correlation or a sample of a real sequence of interest. If the components of early time-lags are taken from the auto-correlation sequence, it can be formulated as in Eq. 3.20 on page 38 by applying a windowing sequence on the auto-correlation sequence. The Fourier transform of the windowed auto-correlation sequence can then be expressed as in Eq. 3.21 on page 38. Suppose that the windowing sequence itself is the auto-correlation sequence of a real causal sequence. Then the Fourier transform of the windowing sequence must be real and non-negative. This is because, following Eq. 3.21 on page 38, the Fourier transform of the windowed auto-correlation sequence is expressed as the convolution in the frequency domain between the two Fourier transforms of the windowing sequence and the auto-correlation sequence before windowing. The convolution of real non-negative sequences (or functions) again yields real non-negative sequences (or functions). Therefore, windowing using, for example, a triangular window provides an estimate of the short-term power spectrum from the auto-correlation sequence including even long time-lag components.

The triangular window described in Sect. 3.3.3 on page 48 is an example of an auto-correlation sequence, if  $w^-(n)$  and  $w^-(-n)$  are combined to make an even sequence. Consider a real sequence  $x(n)$  such that

$$x(n) = \begin{cases} 1, & 0 \leq n \leq N-2, \\ 0, & n = N-1 \end{cases} \quad (3.49)$$

for  $0 \leq n \leq N-1$ . The auto-correlation sequence of  $x(n)$  above is expressed as

$$r(n) = N - 1 \pm \quad (3.50)$$

for  $n = 0, \pm 1, \dots, \pm(N-1)$ , and it can be rewritten as an even sequence

$$\begin{aligned} r(n) &= w^-(-n) + w^-(n) - w^-(0) \\ &= \overline{w^-(n)} = w_e(n). \end{aligned} \quad (3.51)$$

Here, the Fourier transform of  $r(n)$  is a real non-negative sequence. Therefore, the Fourier transform of the triangular windowing sequence  $w_e(n)$  is also a real non-negative sequence. Consequently, the even triangular sequence is a candidate as a windowing sequence for which a real non-negative Fourier transform might be derived even after windowing of the auto-correlation sequence. That is, the short-term power spectrum can be estimated by triangular-windowing the auto-correlation sequence including long-lag components.

A triangular window can be applied also to a frame-wise sequence. The variance in the estimates of the correlation sequence increases as the time-lag is extended. The triangular window is a candidate in pre-processing for correlation analysis to decrease the effects of variance on the estimates of the auto-correlation sequence.

### 3.4.2 One-Side Auto-correlation Sequence and Envelope of Power Spectrum

The effects of triangular windowing on auto-correlation sequences in power spectral estimations can be understood by taking the one-side auto-correlation of the original. An analytic sequence is defined by constructing the causal spectral function. In contrast, a real causal sequence yields an analytic spectral function that normally is defined as the Fourier transform. The magnitude of the complex spectral function can then be interpreted as the spectral envelope in the frequency domain corresponding to the instantaneous magnitude of the sequence in the time domain, and similarly, the phase spectral function corresponds to the instantaneous phase in the time domain.

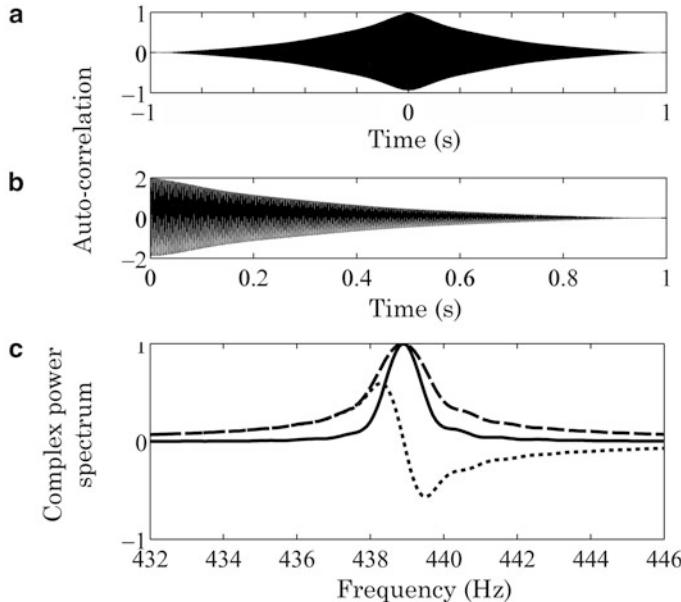
Consider an auto-correlation sequence  $r(n)$  that is a non-causal even sequence in the time domain. Defining a causal (one-side) sequence as

$$\hat{r}(n) = \begin{cases} 2r(n), & n > 0, \\ r(0), & n = 0, \\ 0, & n < 0, \end{cases} \quad (3.52)$$

its Fourier transform becomes

$$\hat{R}(e^{-i\Omega}) = R(e^{-i\Omega}) + i\Im[\hat{R}(e^{-i\Omega})]. \quad (3.53)$$

Here, the real part  $R(e^{-i\Omega})$  is equal to the power spectral density function that must be derived from the original non-causal auto-correlation sequence. In other words, the causal auto-correlation yields the analytic expression of the power spectral density function in complex function form. Consequently, the magnitude

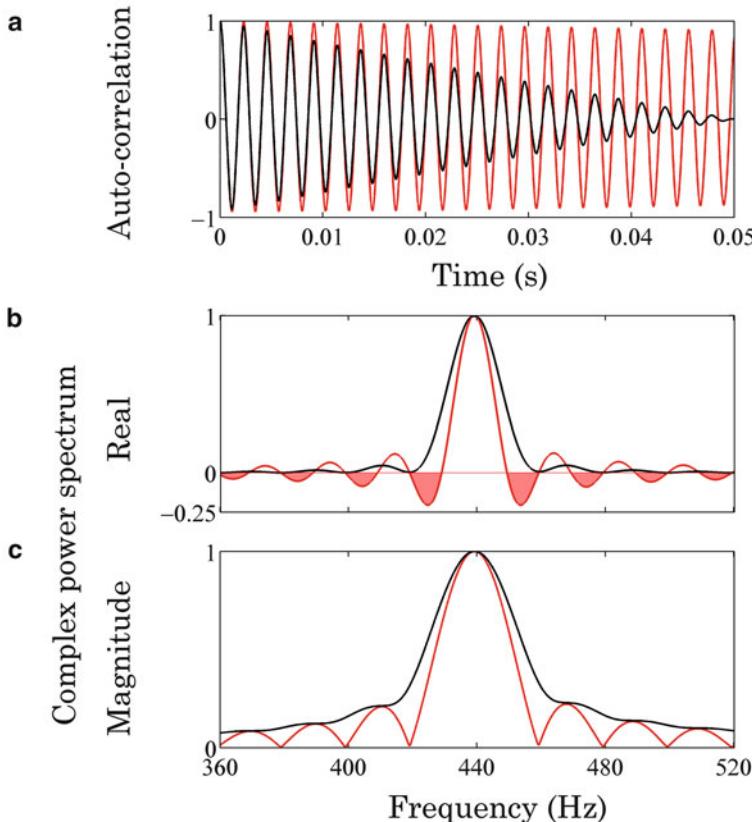


**Fig. 3.14** Causal auto-correlation and complex power spectral density; (a) top: sample of auto-correlation sequence, (b) middle: causal sequence obtained from top panel, (c) bottom: Fourier transforms and magnitude; solid line: real part, dotted: imaginary, broken: magnitude of the complex Fourier transform that is obtained using auto-correlation displayed in middle panel

$|\hat{R}(e^{-j\omega})|$  gives the envelope of the power spectral density function. That is, the macroscopic property of the spectrum can be estimated by suitably averaging the magnitude of the complex power spectrum (instant envelope in the frequency domain) instead of the ordinary power spectral function representing the squared envelope of magnitude spectrum in the frequency domain.

Figure 3.14 presents samples of auto-correlation sequences and complex power spectral density function. The top panel shows an auto-correlation sequence, the middle panel the causal sequence obtained from the auto-correlation, and bottom panel displays the real (solid line) and imaginary (dotted line) parts of the complex power spectrum, and its magnitude (broken line). The solid line is obtained either from the original (top panel)  $r(n)$  or the causal (middle panel) sequence  $\hat{r}(n)$  taken from the original following Eq. 3.52 on the preceding page. In addition, the envelope of the power spectral density function can be obtained, as indicated by the broken line in the bottom panel, by taking the magnitude of the Fourier transform of the causal auto-correlation displayed in the middle panel.

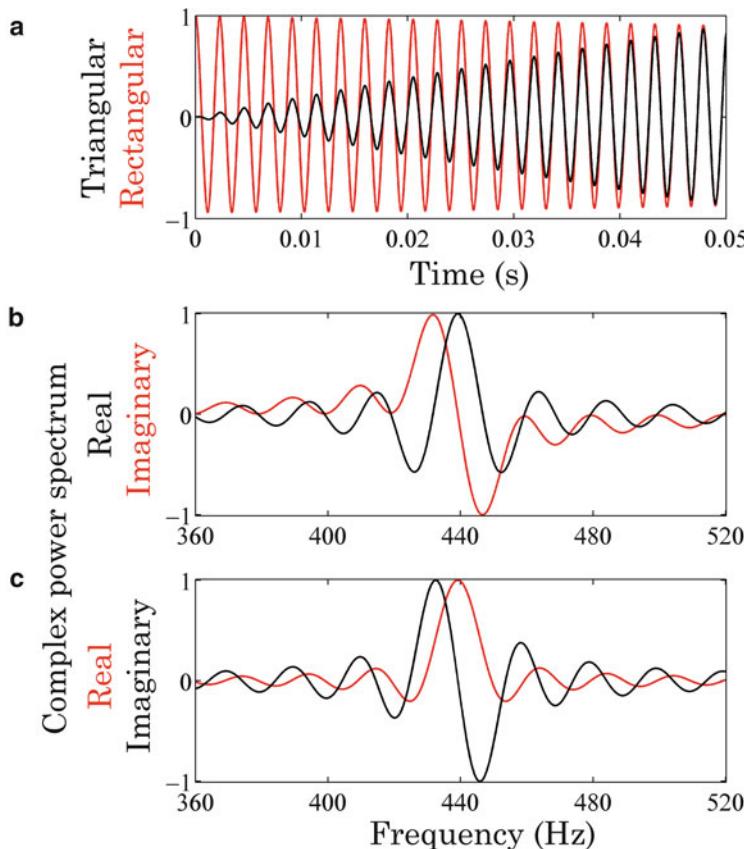
If the triangular window is applied to the causal auto-correlation shown in the middle panel of Fig. 3.14, then a sample of a short-term power spectral density (the real part of the Fourier transform) is obtained as displayed in Fig. 3.15 .



**Fig. 3.15** Sample of short-term power spectral density; (a) normalized short-term causal auto-correlation taken from the auto-correlation shown in Fig. 3.14 on the previous page using triangular window (black) and rectangular window (red), (b) short-term power spectral density using triangular (black) and rectangular (red) windows, (c) magnitude of a complex power spectral density displayed similarly to (a) and (b)

Comparing the triangular (black line) and rectangular (red line) windows confirms that the triangular window yields a non-negative Fourier transform. In addition, the magnitude of the complex power spectral density function also shows up differences due to windowing methods.

Figure 3.16 on the facing page illustrates the effects of triangular windowing with positive slope. The spectral differential effects on the spectrum of the auto-correlation sequence showed in Fig. 3.15 can be confirmed.



**Fig. 3.16** Short-term spectrum; (a) normalized short-term causal auto-correlation using positive-slope triangular window (black) and rectangular window (red), (b) real part of short-term spectrum using triangular window (black) and imaginary part for rectangular (red) window, (c) similarly imaginary for triangular and real for rectangular window

## References

1. R. Meddis, M.J. Hewitt, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *J. Acoust. Soc. Am.* **89**(6), 2866–2882 (1991)
2. M.R. Schroeder, *Computer Speech* (Springer, 1999)
3. M.R. Schroeder, Improved quasi-stereophony and colorless artificial reverberation. *J. Acoust. Soc. Am.* **33**, 1061–1064 (1961)
4. M.R. Schroeder, H.W. Strube, Flat-spectrum speech. *J. Acoust. Soc. Am.* **79**(5), 1580–1583 (1986)
5. A.V. Oppenheim, J.S. Lim, The importance of phase in signals. *Proc. IEEE* **69**(5), 529–541 (1981)
6. M. Kazama, S. Gotoh, M. Tohyama, T. Houtgast, On the significance of phase in the short term Fourier spectrum for speech intelligibility. *J. Acoust. Soc. Am.* **127**(3), 1432–1439 (2010)
7. M. Tohyama, *Sound and Signals* (Springer, 2011)

8. W.A. Yost, R. Patterson, The role of envelope in processing iterated rippled noise. *J. Acoust. Soc. Am.* **104**(4), 2349–2361 (1998)
9. T. Houtgast, H.J.M. Steeneken, R. Plomp, A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* **77**(3), 1069–1077 (1985)
10. M.R. Schroeder, Modulation transfer functions: definition and measurement. *Acustica* **49**(3), 179–182 (1981)
11. G. Strang, T. Nguyen, *Wavelets and Filter Banks* (Wellesley Cambridge Press, 1997)
12. B.C.J. Moore, *An Introduction to the Psychology of Hearing*, 4th edn (Academic Press, 1997)
13. R. Meddis, M.J. Hewitt, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity. *J. Acoust. Soc. Am.* **89**(6), 2883–2894 (1991)
14. R. Meddis, L. O'Mard, A unitary model of pitch perception. *J. Acoust. Soc. Am.* **102**(3), 1811–1820 (1997)
15. M.R. Schroeder, Models of hearing. *Proc. IEEE* **63**(9), 1332–1350 (1975)
16. R.D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, M. Allerhand, Complex sounds and auditory images. *Audit. Physiol. Percep.* **83**, 429–446 (Pergamon Press, 1992)
17. R. Drullman, Temporal envelope and fine structure cues for speech intelligibility. *J. Acoust. Soc. Am.* **97**(1), 585–592 (1995)
18. W.M. Hartmann, *Signals, Sound, and Sensation* (Springer, 1997)
19. R.H. Lyon, *Machinery Noise and Diagnostics* (Butterworth, 2000)
20. M. Tohyama, T. Koike, *Fundamentals of Acoustic Signal Processing* (Academic Press, 1998)
21. A.V. Oppenheim, R.W. Shafer, *Digital Signal Processing* (Prentice-Hall, 1975)
22. Y. Takahashi, M. Tohyama, Y. Yamasaki, Cumulative spectral analysis for transient decaying signals in a transmission system including a feedback loop. *J. Audio Eng. Soc.* **54**(7/8), 620–629 (2006)
23. Y. Ando, *Auditory and Visual Sensation* (Springer, 2009)

# Chapter 4

## Temporal and Spectral Enhancement by Sound Path

A mathematical formulation is given of the temporal and spectral characteristics of sound, including sound path information such as transmission in a room. The auto-correlation analysis of sound followed by a single reflection produces a condition where the direct sound is enhanced by the reflection without spectral degradation. Frame-wise auto-correlation analysis reveals some of the significant source signatures from reverberant speech in which the original speech is embedded. This partly explains why intelligible speech can be delivered under adverse reverberation. The early portion of the auto-correlation sequences for impulse responses between the source and receiving positions yields macroscopic power spectral features, such as the power spectral envelopes, that are not highly sensitive to the reverberation condition and thus are preserved even in a reverberation field. In addition, the spectral energy analysis of temporal envelopes of speech reveals that some of the significant narrow-band envelopes could be enhanced even in the coherent region close to the sound source in a room.

### 4.1 Steady-State Response of Sound Path in Rooms and Early Reflection

#### 4.1.1 Impulse Response and Frequency Characteristics

Listening to sounds in a room or similar confined spaces is a common experience for all. However, in general, environmental conditions deform sound signals as these travel diverse paths from source to receiver. The sound paths or transmission between two spatial positions, as for example in a good concert hall [1], can help in the enjoyment of music, but sometimes, a multitude of sound paths can make speech difficult to perceive such as in a large auditorium [2]. Such path characteristics of sound in confined spaces can, in principle, be represented by an impulse response

between the locations of source and receiver, assuming that the sound response is linear along the transmission path.

The direct sound, that is, sound propagating directly from source to receiver, is normally followed by a series of reflected sounds, as intuitively imagined assuming a geometric wave picture of reflections from walls and other prominent surfaces [3]. Thus, the impulse response is composed of direct and reflected sounds followed by reverberations. From a perceptual point of view, sound reflection can be broken down into those early reflections that fuse in with the direct sound and the distinct echoes that follow. Moreover, reverberation following the early reflections and echoes, continually but almost randomly arrive at the receiving point, and thus, the latter part of the impulse response seems as if it might be composed of continuously attenuating sound waves. Consequently, the total energy of the impulse response that is given by the square sum of the response is mostly determined by reverberation in general, except when the receiver is positioned very close to the sound source.

The Fourier transform of the impulse response gives the steady-state frequency characteristics including magnitude and phase, or real and imaginary parts. The reverberation constructing the latter portion of the impulse response can be interpreted as being random much like wide-band noise. Therefore, the frequency characteristics (real and imaginary parts) can also be treated as random. This explains why statistical approaches have been used in sound field analysis, in particular, room acoustics [4, 5]. If spatial averages are then taken with respect to the positions of both source and receiver throughout the room, the power spectral density function can be expressed as a superposition of the power spectral density for each eigenfrequency of the sound field in the room. Thus, the effects of the eigenfrequencies of the environmental linear system can still be seen in the power spectral response, even after spatial averages are taken.

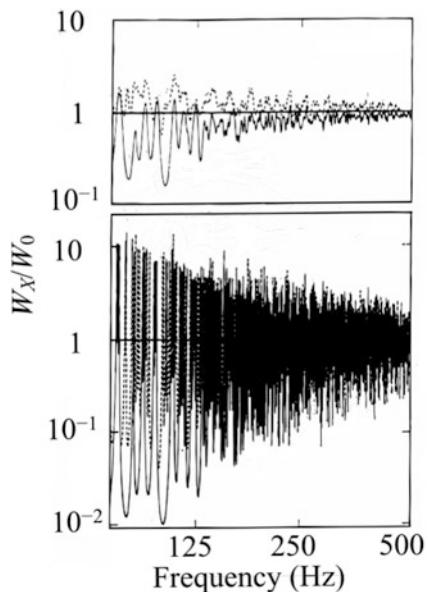
The sound power output of a source, which is called the power response of a source, depends on its position in a room. By taking the space average with respect to the sound source position in the room, the sound power output becomes

$$\frac{\langle W_X(\omega) \rangle}{W_0} \cong \frac{4\pi c^3}{V} 2\delta \sum_N \frac{1}{(\omega^2 - \omega_N^2)^2 + 4\omega^2\delta^2} \quad (4.1)$$

which is unique to the room where the sound source is located and indicates the eigenfrequencies are also significant factors for the sound power output [3]. Here  $W_0$ ,  $\delta$ ,  $\omega_N$ ,  $V$ , and  $c$  denote, respectively, the sound power output in a free field, the decaying factor, the eigen-angular-frequency, the volume of the room and the speed of sound in the room.

Figure 4.1 on the facing page is an example of a numerical calculation of the sound power output from a point source of sinusoidal waves in a rectangular reverberation room [6] using Eq. 4.1. The sound power output in a room might greatly differ from that in a free field. If the frequency of sound radiating from the source is the same as an eigenfrequency of the room, the sound power output becomes large, whereas it attenuates for sounds of different frequencies. The eigenfrequencies are

**Fig. 4.1** Sound power output from source averaged over the source position, calculated using Eq. 4.1 on the preceding page.  $W_0$  is the sound power output of a point source in a free field (W),  $T_R$  is the reverberation time for the reverberation room in which sound source is located (s), room volume is  $200 \text{ m}^3$ , and the ratios of the room dimensions are  $1 : 2^{1/3} : 4^{1/3}$ . (a)  $T_R = 1 \text{ s}$ ; (b)  $T_R = 16 \text{ s}$ , dashed lines: axial, tangential and oblique modes, solid lines: oblique modes only; taken from [6] (Fig. 1) and [3] (13.1)



sparsely found at low frequencies, in particular in a small room. Consequently, the frequency dependence of the variance in the power response appears very large because of the randomness of the distribution of eigenfrequencies, if the sound source position is averaged throughout the room.

With regard to the randomness of the frequency domain or temporal region, it seems unlikely for sound signals such as intelligible speech to be circulated without deformation from the source to listeners in a room. However, despite ambient conditions, our ears are well structured so that speech can be recognizable, even if the frequency characteristics of the transmission path seem random [7]. This capability is key in exploring sound perception derived from speech and/or music.

#### 4.1.2 Direct and Early Reflected Sounds

The steady-state frequency characteristics within a room are mostly determined by reverberation perceived by the listener in the later part of the impulse response. Nevertheless, early reflections following the direct sound are quite important from a perceptual point of view [1]. The implication is that the effect of early reflections might instead be noticeable in the temporal domain rather than the frequency plane. Two classical parameters have been commonly used for sound field analysis in rooms under steady-state conditions; one is the energy ratio  $K$  between direct and other sounds. This ratio is sensitive to the distance from source to receiver in the

room, assuming that direct sound reaches the receiver as a spherical wave [3]. Therefore, the distance that gives  $K = 1$  is called the critical distance in the sound field.

The other parameter, denoted  $D$ , is a modified energy ratio from a perceptual point of view. Here,  $D$ , standing for definition (related to clarity of sound) or Deutlichkeit in German, is defined as the energy ratio of the direct sound, including reflections arriving within about 50 or 30 ms afterwards (denoted  $D_{50}$  and  $D_{30}$ ), with respect to the total sound [8,9]. The modification is based on the empirical fact that reflected sounds arriving at the receiver shortly after the direct sound are helpful in enhancing the energy in direct sound, particularly speech. Thus, late reflections such as echoes arriving at the receiver 50 ms after the direct sound are perceived separately from direct sound. This fact suggests a superposition model for direct sound and its first or very early reflections.

#### 4.1.3 Auto-correlation of Direct Sound Followed by a Single Reflection

The envelope of a time sequence represents the temporal change in the energy. Suppose that the direct and reflected sounds are superposed. Denoting the direct sound as  $s_d(n)$ , the reflected sound is written as  $s_r(n) = s_d(n-m)$  where  $m$  denotes the delay time of the reflected sound. The squared average of the sum of the direct and reflected sounds depends on the cross-correlation between the two components, or the auto-correlation sequence of the direct one. That is,

$$\begin{aligned} & E[(s_d(n) + s_d(n-m))^2] \\ &= E[s_d^2(n)] + E[s_d^2(n-m)] + 2r_d(m) \end{aligned} \quad (4.2)$$

where  $E[*$ ] denotes the ensemble average, and  $r_d(m)$  the auto-correlation sequence of the direct sound. If the auto-correlation is positive (negative), then the reflected sound is positively (negatively) added to the direct sound.

However, the auto-correlation sequence of sound strongly depends on the audio frequency. It indicates that the criterion for superposition varies due to frequency. Suppose that the  $i$ -th sub-band component of direct sound is expressed as

$$\begin{aligned} s_{d_i}(n) &= e_i(n) \cdot c_i(n) \\ &\cong e_i(n) \cdot \cos \Omega_i(n) \end{aligned} \quad (4.3)$$

where  $e_i(n)$  and  $c_i(n)$  denote the envelope and carrier in the  $i$ -th band respectively, and  $\Omega_i$  the center (or representative) angular frequency of the narrow band [3, 10]. The auto-correlation sequence for the sub-band component of direct sound can be written as

$$\begin{aligned} r_i(m) &= E[s_{d_i}(n) \cdot s_{d_i}(n - m)] \\ &= C r_{e_i}(m) \cdot \cos \Omega_i(m), \end{aligned} \quad (4.4)$$

where  $C$  is a constant and  $r_{e_i}(m)$  denotes the auto-correlation sequence of the envelope

$$r_{e_i}(m) = E[e_i(n) \cdot e_i(n - m)]. \quad (4.5)$$

Following the formulation above, it can be surmised that if the auto-correlation of the envelope becomes sufficiently small in every frequency band, then the superposition might be non-negative independent of the frequency bands. Here, note that the auto-correlation of the narrow-band envelope can be replaced by the envelope of the auto-correlation for the narrow-band component of the sequence.

The auto-correlation of a sequence is determined by the power spectrum of the sequence. The envelope spectrum of speech is mostly distributed below about 20 Hz, and the most significant spectral component can be seen around 4 or 5 Hz [2]. Therefore, the time lag in which the auto-correlation decreases to around zero is expected to be about 50–60 ms. This time interval might provide the basis for the introduction of  $D$ ; namely, it is quite likely that reflected sound might be separately perceived from direct sound, if the time delay exceeds 50 ms, because the auto-correlation becomes quite low between the corresponding envelopes of these two.

Interestingly, [1] defines  $\tau_e$  as the time lag required to reduce the envelope of the auto-correlation to 1/10 from its initial value (unity for auto-correlation sequences), and consequently indicates that  $\tau_e$  is the most preferable delay time of a single reflected sound. Namely, the preferable acoustic condition of sound fields depends on the sound source characteristics, because  $\tau_e$  depends on the property of the source itself. A time-lag approaching 0 for the envelope correlation seems the appropriate critical condition of a single reflection, because the time delay yields the non-negative superposition to the direct sound, subject to that the reflected sound might not be separately perceived from the direct sound.

#### **4.1.4 Temporal and Spectral Representation of Transmission Sound**

Frame-wise Fourier transforms represent temporal characteristics of sound. This is true of sound propagating within a room. Taking the frame-wise Fourier transforms of such sounds, it is quite unlikely because of short frame lengths that the steady-state response dominates the frequency characteristics. Thus, the characteristics for transient sound should be expressed in both the temporal and frequency domains.

Suppose that real sequence  $x(n)$  provides the signal fed into a room at  $t = 0$ . Denoting the impulse response between the source and receivers by  $h(n)$ , the transient response  $y(n)$  at the receiver can be written as

$$\begin{aligned} y(n) &= \sum_{m=0}^n h(m)x(n-m) \\ &= \sum_{m=0}^n h(m)e^{-i\Omega m} \frac{1}{2\pi} \int_0^{2\pi} X(e^{-i\Omega}, n) e^{i\Omega m} d\Omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} X(e^{-i\Omega}, n) H(e^{-i\Omega}, n) e^{i\Omega n} d\Omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} Y(e^{-i\Omega}, n) e^{i\Omega n} d\Omega \end{aligned} \quad (4.6)$$

where

$$x(n) = \frac{1}{2\pi} \int_0^{2\pi} X(e^{-i\Omega}, n) e^{i\Omega n} d\Omega \quad (4.7)$$

$$X(e^{-i\Omega}, n) = \sum_{m=0}^n x(m) e^{-i\Omega m}. \quad (4.8)$$

The results above indicate that the temporal and spectral response to an input sequence is not determined solely by the impulse response. The power spectral property can be formulated;

$$|Y(e^{-i\Omega}, n)|^2 = |X(e^{-i\Omega}, n)|^2 \cdot |H(e^{-i\Omega}, n)|^2. \quad (4.9)$$

In other words, the auto-correlation sequence of the transient response can be expressed as

$$r_y(m, n) = r_x * r_h(m, n) \quad (4.10)$$

where  $m$  denotes the time shift over which the auto-correlation sequence is taken. Note that the auto-correlation sequence changes in time corresponding to the temporal response.

## 4.2 Source Effects on Sound Path

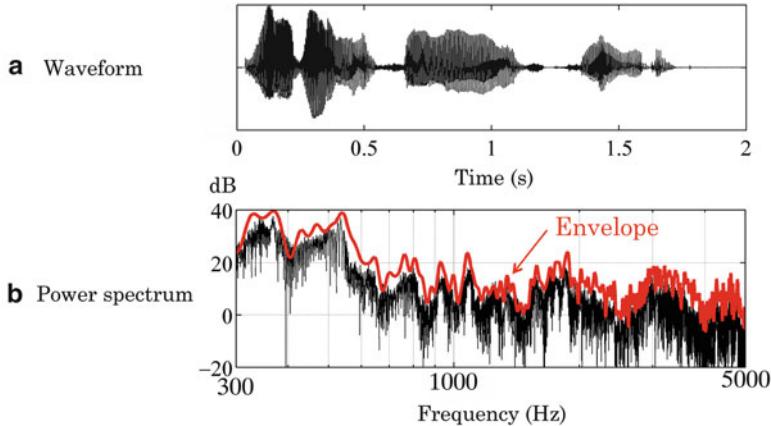
### 4.2.1 Frame-Wise (Short-Term) Auto-correlation of Transmission Sound

The power spectral response of a linear system, such as the sound field in a room, contains both source and transmission path information. However, decomposing the power spectral response into the frequency characteristics of the source and transmission path is a difficult task. Called blind source separation, much research has been conducted on this aspect and is summarized in [11]. If the sound source is similar to white noise, namely its auto-correlation sequence can be approximated as  $r_x(n) \cong C \cdot \delta(n)$  with  $C$  a constant, then it is quite likely that the power spectral information of the transmission path might be obtained from the power spectral response without the source effect. Alternatively, if the sound source is a sinusoidal wave, namely, its auto-correlation sequence is given by  $r_x(n) \cong C \cos \Omega_0 n$ , then it is unlikely to obtain an estimate of the power spectral characteristics of the transmission path from the response.

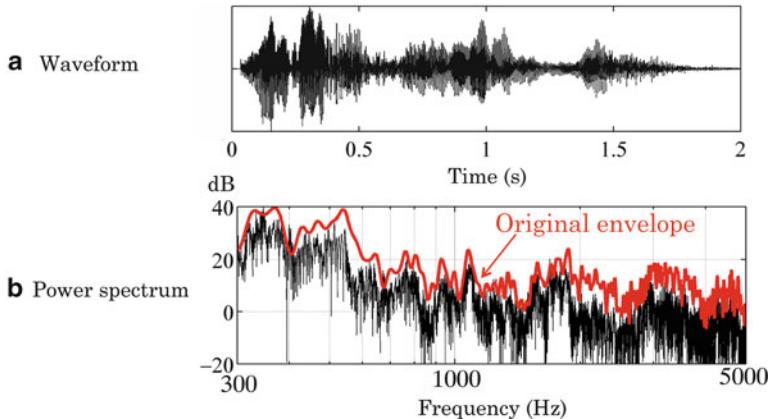
In contrast, the fact that the intelligibility of speech might be preserved even in a reverberating room implies that the human ear seems particularly good at separating the source of sound from the background noise of daily life. The frame-wise response might give a possible interpretation of this fact. A point in a room receives direct sound from a source followed by its early reflections in the present short frame and reverberations from the past direct sound in previous frames [7]. It is quite likely, however, that the auto-correlation of the frame-wise speech sample might be approximated by  $r_{rev-sp} \cong D_{30}r_{sp} + (1 - D_{30})r_{rev}$  subject to  $r_{sp} \cong 1$  and  $r_{rev} \cong 0$  where  $r_{sp}$  and  $r_{rev}$  denote the auto-correlation of the direct and reverberation components of the speech sample, respectively. That is, the frame-wise auto-correlation could be dominated by the direct sound component. As the frame-length becomes long, the received sound is a mixture of the successive frame-wise responses in which the direct sound might be embedded within the reverberation.

Figure 4.2 on the following page shows a sample of a speech waveform (a), the power spectral density function (b) with the power spectral envelope. Figure 4.3 on the following page displays the response for the same speech sample as in Fig. 4.2 on the following page but recorded within a reverberant room. Panel (a) is the recorded waveform, whereas (b) presents the power spectrum similar to (b) in Fig. 4.2 on the following page. Comparing the waveforms in Figs. 4.2 on the following page and 4.3 on the following page, the effects of transmission path on speech samples are noticeable; hence, it seems difficult to estimate the source characteristics by separating the path effects from the response.

Nevertheless, some of the source signatures can be seen such as the power spectral envelope, even in the reverberant responses. This fact suggests some of the source signals hidden in the response might appear if the frame-wise (short-term) auto-correlation or power spectral analysis is performed. This is because



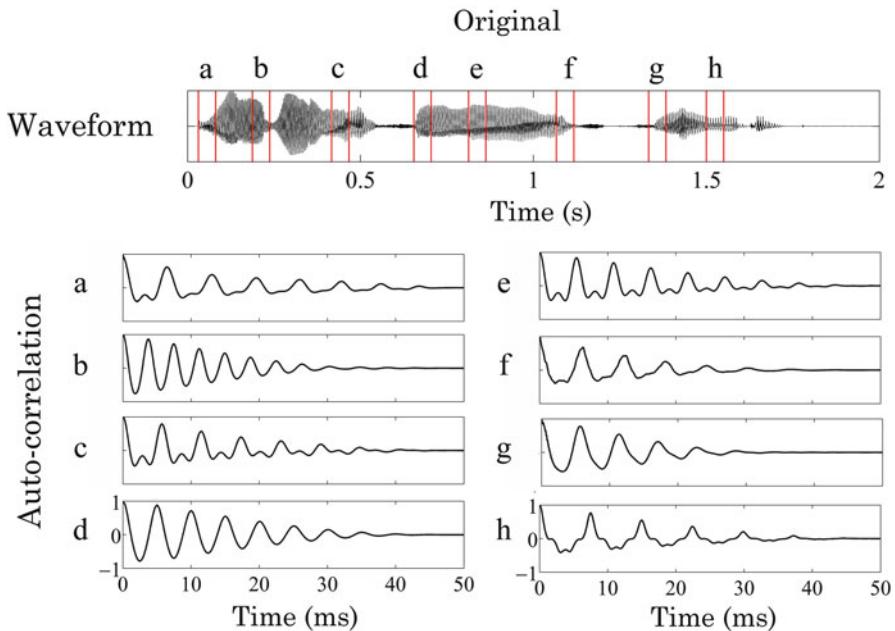
**Fig. 4.2** Sample of a speech waveform (**a**) and its power spectral density function (**b**) with spectral envelope (red line)



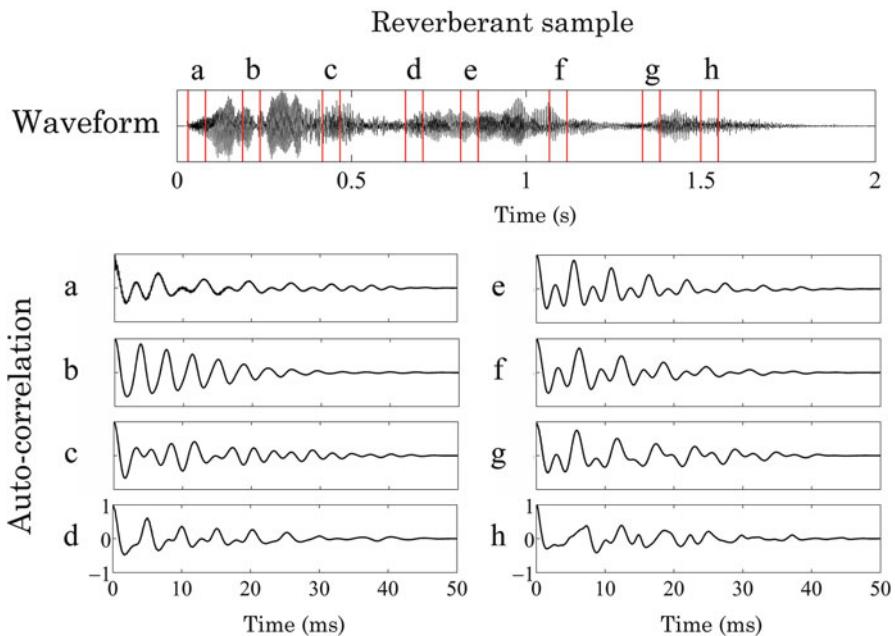
**Fig. 4.3** Response to speech sample shown in Fig. 4.2 in a reverberation room

the frame-wise approach is helpful in smoothing the response so that the spectral envelope might appear. Figure 4.4 on the facing page displays examples of frame-wise (short-term) auto-correlation sequences where the triangular window sequence is applied to each frame-wise speech sequence. These frame-wise auto-correlations represent the temporal characteristics of the speech sample. Similarly, Fig. 4.5 on the facing page shows samples of frame-wise auto-correlation sequences taken over from the reverberant speech by Fig. 4.3.

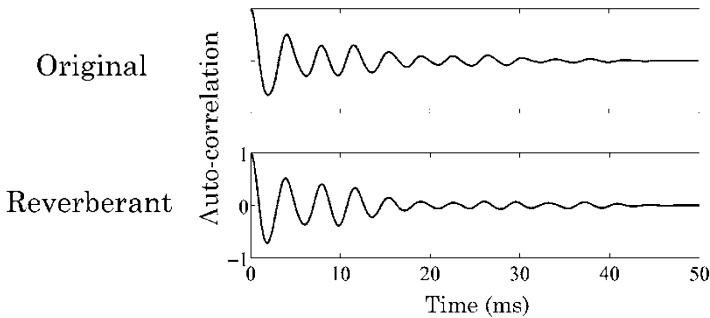
Note that the frame-wise responses possibly exhibit source signatures in the auto-correlation responses, which could be identified in Fig. 4.2 where the speech sample is recorded in an anechoic room without reverberation. Figure 4.6 on page 66 illustrates the short-term auto-correlations, which are obtained using



**Fig. 4.4** Samples of frame-wise auto-correlation sequences of speech shown in Fig. 4.2 on the preceding page



**Fig. 4.5** Frame-wise auto-correlation samples for reverberation speech shown in Fig. 4.3 on the preceding page where frames are taken similarly to Fig. 4.4



**Fig. 4.6** Initial portion of auto-correlations for speech and reverberant samples

triangular windowing with a long-term auto-correlation sequence. The power spectral envelopes are mostly identical to each other. This may explain why intelligible speech can be delivered even under adverse reverberant conditions.

#### 4.2.2 *Source Effect and Distance to Receiver*

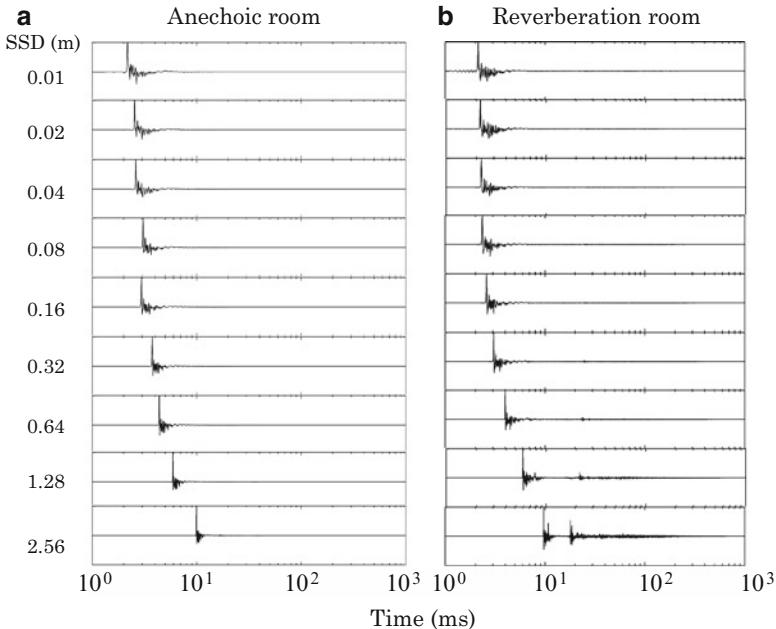
The signal signatures of sound emanating from a sound source are assumed to be independent of path information, in general; however, the characteristics of the transmission path are also changed by sound-source conditions. The energy in direct sound changes because of the distance between source and receivers, specifically, energy decreases inversely proportion to the square of the distance, assuming the wave propagates radially from source to receiver.

Figure 4.7 on the facing page illustrates samples of impulse responses measured in anechoic (a) and reverberation (b) rooms, respectively [12, 13] as shown in Fig. 4.8 on the facing page.

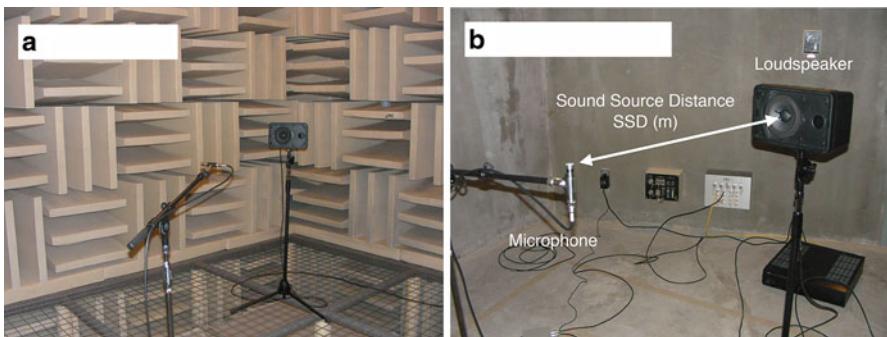
Here, the distance is a parameter that changes the path-information. The sound source used for the measurements in the figure is a commercially available small loudspeaker system. Rows in both columns present the impulse response for a specified distance. Intuitively, the initial delay portion, which is inevitable for wave propagation from source to receivers, becomes longer. In addition, note that the magnitude range along the vertical axis of each graph is normalized.

In column (a), there are no significant changes in the responses, even as the receiver moves further from source. This is because reflected sounds mostly contribute to changes in the response; there are no noticeable reflections in the anechoic room. Therefore, the impulse responses indicate only the characteristics of the sound source, that is, the loudspeaker system itself. In contrast, column (b) displays significant changes because of conditions along the transmission paths.

Such effects of distance on the impulse responses under reverberation conditions can be expressed using the energy ratio  $D$ . Figure 4.9 on page 68 shows  $D_{30}$  for

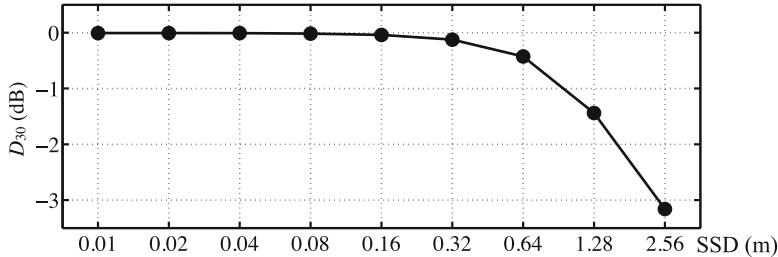


**Fig. 4.7** Samples of impulse responses (initial portion) from a loudspeaker system in an anechoic room (a) and a reverberation room (b) with varying distance between source and receiver



**Fig. 4.8** Experimental setup. (a) Anechoic room. (b) Reverberation room

results from column (b) in Fig. 4.7. Clearly,  $D_{30}$  remains almost constant where direct sound contributes mainly to the field; however,  $D_{30}$  rapidly decreases outside this range. As described in Sect. 4.1.2 on page 59, the critical distance  $K = 1$  indicates the range of the region controlled by direct sound. This region is called the coherent region of the sound field [3, 14]. The results above for  $D_{30}$  imply that sound perception in a room changes significantly if the distance exceeds the range of the



**Fig. 4.9** Energy ratio  $D_{30}$  for responses in Fig. 4.7 on the previous page

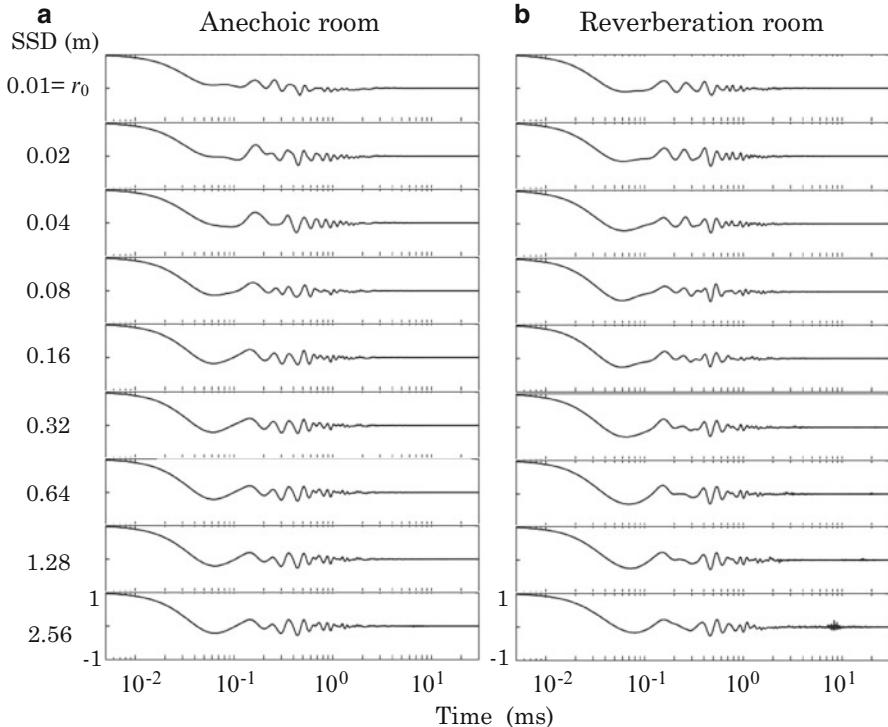
coherent region. There, the source information is largely deformed by reflections and reverberation in the transmission path.

#### 4.2.3 Auto-correlation of Impulse Response and Distance from Source

As shown in Fig. 4.9,  $D_{30}$  remains almost constant in the region very close to the sound source. However, it might be possible for a listener to hear differences due to distance, which cannot be suitably detected by  $D_{30}$  values. The short-term auto-correlation sequences that display distance effects on the received sound including those arising from transmission path.

Figure 4.10 on the facing page shows the distance dependence of the auto-correlation sequence governing shorter time-lags of less than 30 ms for each of the impulse response in Fig. 4.7 on the previous page under anechoic (a) or reverberation (b) room condition. Each auto-correlation sequence is obtained by applying the triangular window for the long-term auto-correlation sequence. As intuitively expected, there are no significant differences between the anechoic and reverberation room conditions in the region very close ( $SSD \leq 0.16$  m) to the sound source. This indicates that the differences in the auto-correlation sequences are mostly due to sound-source effects without including the path condition. As the distance lengthens beyond 0.16 m, the auto-correlation differences because of distance increase more rapidly under reverberant than anechoic conditions, even if the distance is within the region in which  $D_{30}$  remains constant. This increase is mainly due to both the early reflected and reverberant sounds propagating in the reverberation room [1].

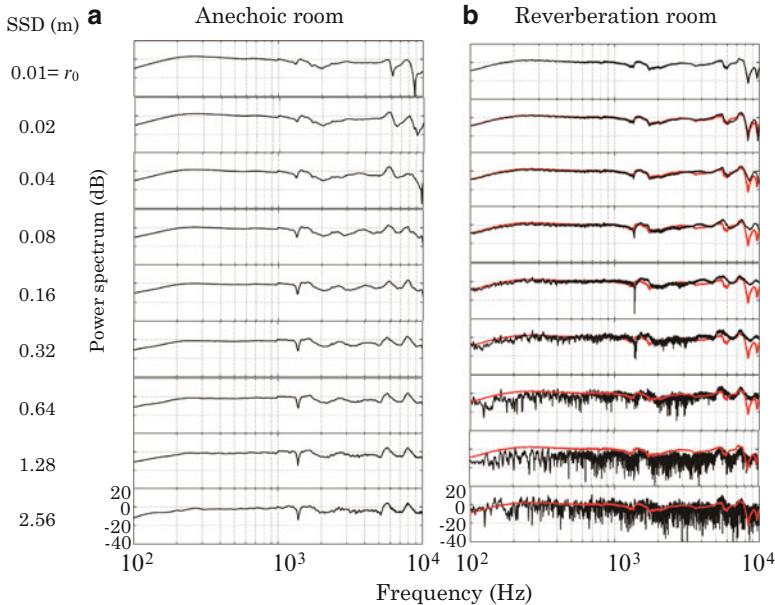
The changes in the auto-correlation sequences could be also understood by power spectral characteristics. Figure 4.11 on page 70 shows the power spectral functions at each distance. In the anechoic room condition, the power spectral responses are mostly identical independent of the distances. In contrast, the frequency characteristics significantly change in the reverberation room as the distance becomes large; however, the response at  $r_0 = 0.01$  m looks like the spectral envelope at every



**Fig. 4.10** Distance dependence of the short-term auto-correlation sequences for impulse response in Fig. 4.7 on page 67 under anechoic (a) or reverberation (b) room conditions. Each auto-correlation sequence is obtained by applying the triangular window for a long-term sequence

distance. Figure 4.12 on page 71 illustrates the frequency response for the early reflections in the reverberation room. The power spectral envelopes of the responses are also mostly identical independent of the distances. It means the spectral changes due to the distances could be interpreted as the variations of the spectral fine structure rather than the macroscopic nature like the spectral envelope. In other words, spectral envelope-like feature might be preserved even when the distance becomes long from the source. It explains again why frame-wise auto-correlation analysis of the reverberation speech could reveal the original signature of the speech sample even under the reverberation effects due to the sound path.

The results might be summarized as follows. The change in the auto-correlation is mainly due to spectral effects inside the coherent range; thus, the loudspeaker effect is the origin of spectral change, as well as changes in the anechoic room. In contrast, as the sound receiving position moves from the source to outside the coherent range in the reverberation room, the change in the auto-correlation sequence is mostly due to temporal effects, rather than the spectral effect that arise under anechoic room conditions. The temporal effects might mark a separation into



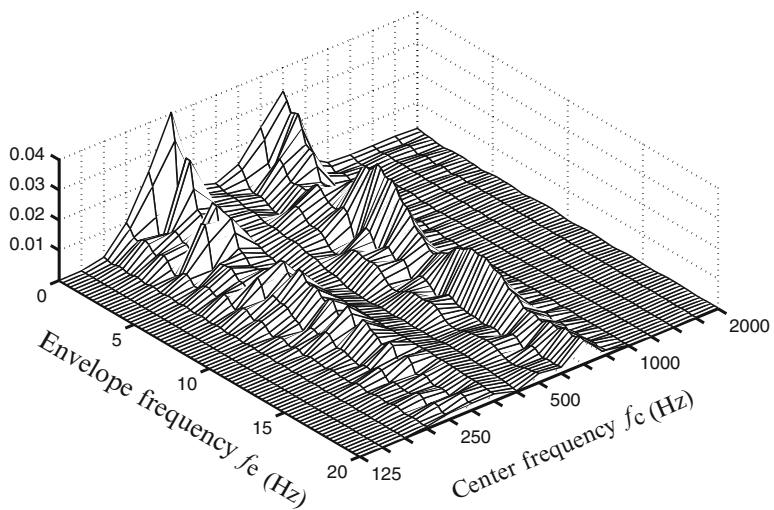
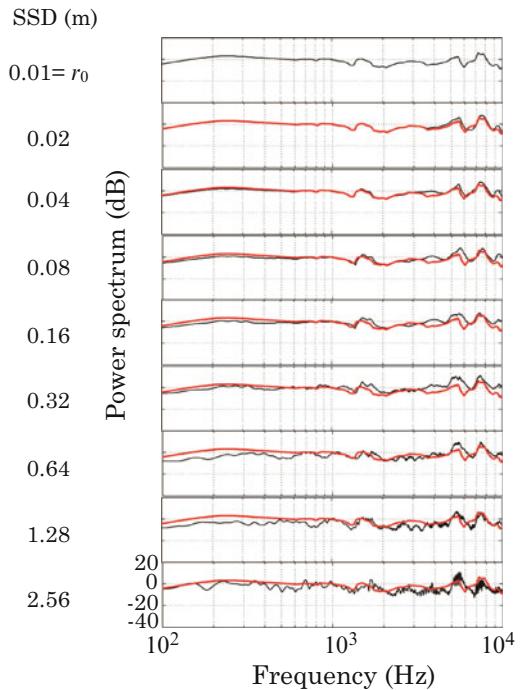
**Fig. 4.11** Power spectral responses in reverberation for impulse response in Fig. 4.7 on page 67 under anechoic (**a**) or reverberation (**b**) room conditions where red line denotes power spectral response at  $r_0 = 0.01$  m

direct sound with echoes and into solely reverberation. In other words, temporal effects cannot be expected under anechoic room conditions.

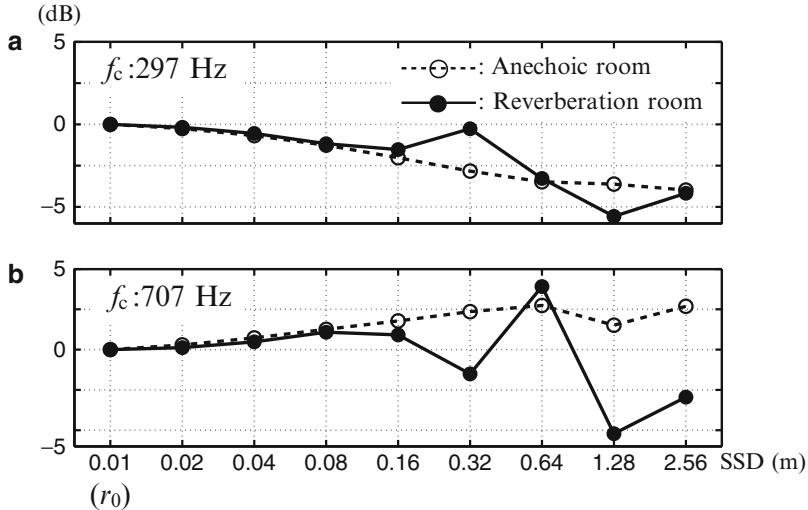
#### 4.2.4 Envelope Energy and Distance from Sound Source

A possible approach to the temporal and spectral analysis of sound including signal dynamics is narrow-band envelope analysis. Figure 4.13 on the facing page illustrates the spectral energy analysis of narrow-band envelopes of a speech sample taken at certain distances from the source [15]. Two representative frequency bands can be seen, specifically, around 300 and 700 Hz. The sound path, in general, might give a positive effect on the speech material, when the receiving position is located within the coherent region, whose range is shorter than 0.16 m in this experiment carried out in the reverberation room. In particular, the early reflected sound advances the speech, so that the power spectral property of speech could be preserved. Interestingly, such a positive effect that depends on the frequency band can be seen in the energy of envelopes, which represents the temporal dynamics of speech.

**Fig. 4.12** Power spectral responses in early reflection for auto-correlation in Fig. 4.10 on page 69 under reverberation room conditions where red line denotes power spectral response at  $r_0 = 0.01$  m [13]



**Fig. 4.13** Example of narrow-band envelope spectral analysis of female speech in a reverberation room from Fig. 3 in [15]

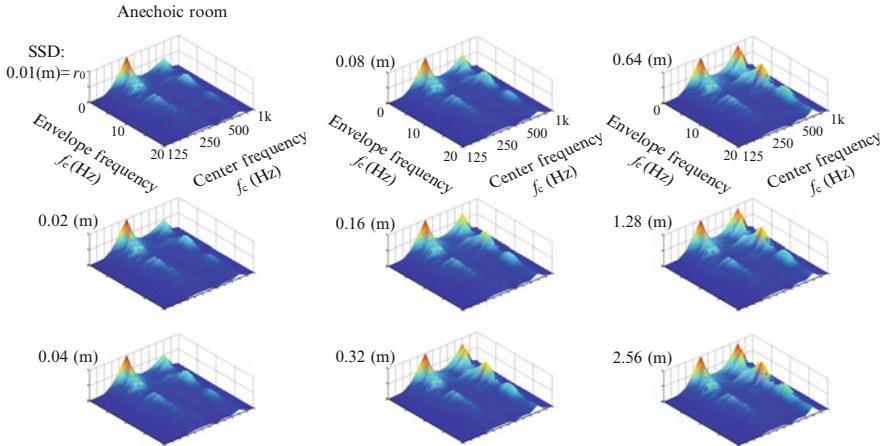


**Fig. 4.14** Spectral energy analysis of narrow-band envelopes from Fig. 4 in [15]

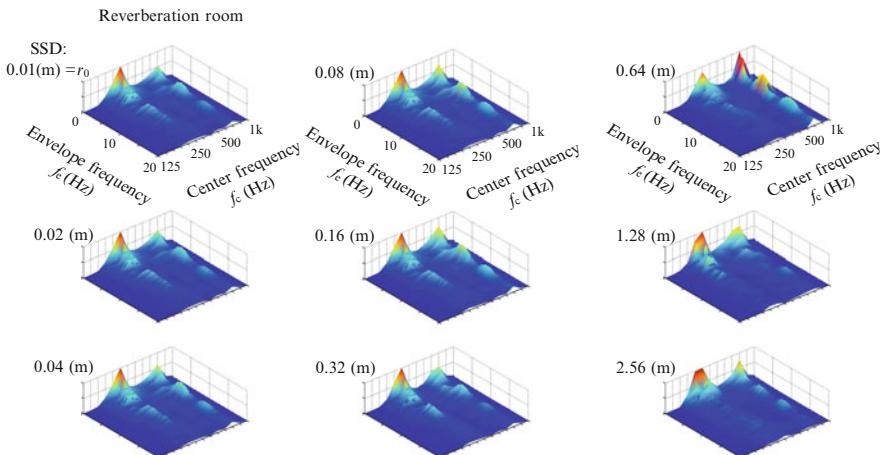
Figure 4.13 on the previous page also presents the spectral components of the narrow-band envelopes. The frequency ranges of the envelopes are mostly within 20 Hz. If the envelope energy increases (decreases), the temporal dynamics is strong (weak). Figure 4.14 shows the energy within 20 Hz normalized with respect to distance  $r_0$ .

It would be quite intriguing to see if the temporal dynamics of speech are enhanced within  $\text{SSD} \leq 0.16$  m for 700 Hz band; however, such temporal information would be weakened as the distance exceeds 0.16 m. In contrast, the energy for 300 Hz does not increase markedly even within 20 Hz. This trend is quite similar to that for the anechoic room. The positive effect is also supported by a subjective evaluation of speech enhancement [15].

The result suggests that the effect of the sound path might be positive in the coherent range at least in the 700 Hz band, but it could be negative outside the range. Envelope energy analysis would be a possible approach to estimate the frequency-dependent enhancement of the temporal signal dynamics arising from sound paths involved in early reflections and reverberation. Figures 4.15 on the facing page and 4.16 on the facing page display the envelope-spectral analysis as a function of a distance from the sound source in the anechoic and reverberation rooms in which the magnitudes are normalized by the maximum in every distance [13]. It is quite intriguing to see the average spectrum over the spectral plane at every distance. Figure 4.17 on page 74 presents the average of the magnitudes of envelope-modulation spectra at each distance. Comparing the results of anechoic

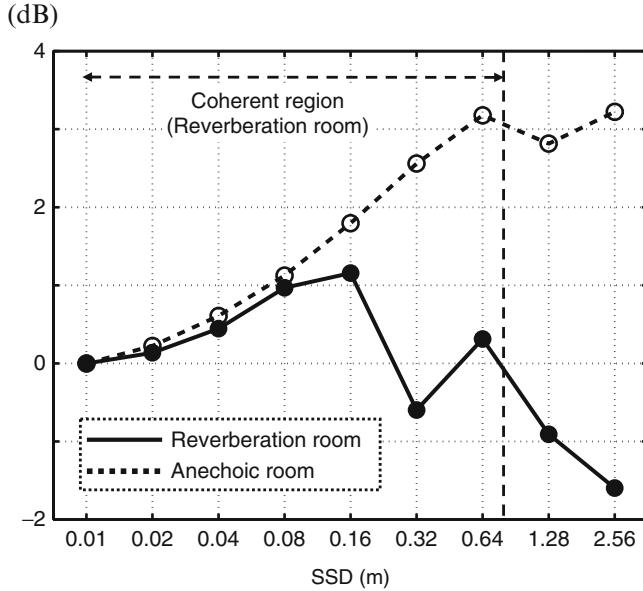


**Fig. 4.15** Narrow-band envelope spectral analysis of female speech as a function of distance from the source in the anechoic room where magnitudes are normalized by maximum at every distance from Fig. 5.4 in [13]



**Fig. 4.16** Similar to Fig. 4.15 but for the reverberation room from Fig. 5.3 in [13]

and reverberation rooms, the differences in the effects of reflections on the speech envelopes between the two conditions are intuitively understood. In particular, the reflection waves enhance the modulation signature of speech in the coherent region even in the reverberation room as well as the anechoic room.



**Fig. 4.17** Envelope spectral energy averaged over envelope-spectral plane at every distance where the *broken line* and *open plots* indicate the anechoic room (Fig. 4.15 on the previous page) and the *solid line* and *plots* show the reverberation room (Fig. 4.16 on the previous page) from Fig. 5.7 in [15]

## 4.3 Decay of Reverberation

### 4.3.1 Reverberation Decay Curve of Sound Waves in Rooms

Reverberation of sound waves are typical phenomena present in sound fields in rooms. Suppose that a white-noise source  $x(n)$  stops propagating at  $n = 0$  after the sound field reaches a steady state in the room. The sound pressure response  $y(n)$  observed at a position in the room is expressed as

$$y(n) = \sum_{m=n}^{N-1} h(m)x(n-m) \quad (4.11)$$

after the noise source has stopped. Here,  $x(n) = 0$  for  $n \geq 0$  and  $h(n)$  denotes the impulse response with record length  $N$  between the source and receiver in the room. By taking the ensemble average of  $y^2(n)$  due to the random nature of the noise source, the reverberation decay curve

$$R(n) = E[y^2(n)] = E_0 \sum_{m=n}^{N-1} h^2(m) \quad (4.12)$$

is obtained, where

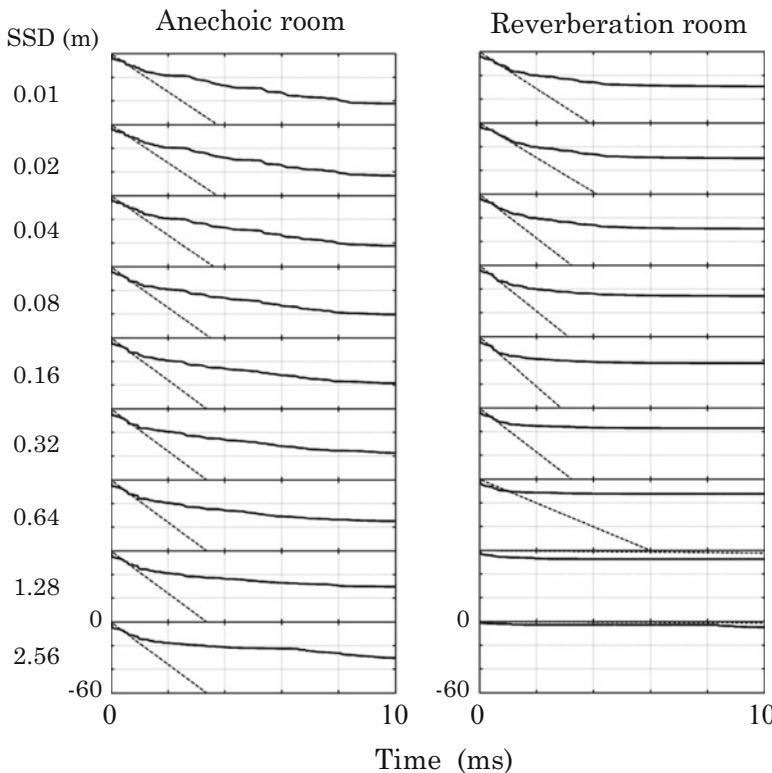
$$E[x(m)x(m-n)] = E_0 \delta(n) \quad (4.13)$$

is being assumed [16]. A detailed description of reverberation decay curves in rooms can be found in [3]. In general, the reverberation decay curves are measured at every sub-band, such as the 1/3-octave bands, by various pairs of sound sources and observation locations by taking spatial averages within the room.

### 4.3.2 Decay Curves and Distance from Sound Source

The reverberation decay curve depends also on the distance between source and receiver, in particular, the initial portion of the decay curve is sensitive to this distance. This effect of distance on the decay curve could be due to the direct sound from the source rather than the room condition.

Figure 4.18 shows the dependence of the initial decay curves in the reverberation room using Eq. 4.12 on the preceding page on distance from the source obtained from the impulse responses as displayed in Fig. 4.7 on page 67. The initial decay



**Fig. 4.18** Initial decay curves using Eq. 4.12 on the preceding page for impulse response displayed in Fig. 4.7 on page 67 where decay curves in anechoic room are shown for reference; *broken lines* represent estimates for the initial decay rate

rate (or the slope of the initial decay) remarkably decreases as the distance becomes longer than the range controlled by the direct sound, as mentioned in Sect. 4.2.2 on page 66. Such a slow decay rate is due to reverberation. In contrast, a rapid decay rate over short distances is primarily due to the direct sound.

## References

1. Y. Ando, *Auditory and Visual Sensation* (Springer, 2009)
2. T. Houtgast, H.J.M. Steeneken, R. Plomp, A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* **77**(3), 1069–1077 (1985)
3. M. Tohyama, *Sound and Signals* (Springer, 2011)
4. M.R. Schroeder, Statistical parameters of the frequency response curves in large rooms. *J. Audio Eng. Soc.* **35**(5), 299–306 (1987)
5. R.H. Lyon, Statistical analysis of power injection and responses in structures and rooms. *J. Acoust. Soc. Am.* **45**(3), 545–565 (1969)
6. M. Tohyama, A. Imai, H. Tachibana, The relative variances in sound power measurements using reverberation rooms. *J. Sound Vib.* **128**(1), 57–69 (1989)
7. M.R. Schroeder, *Fractals, Chaos, Power Laws* (W.H. Freeman and Company, 1991)
8. R. Thiele, Richtungsverteilung und Zeitfolge der Schallrueckwuerfe in Raeumen. *Acustica* **3**, 291–302 (1953)
9. T.J. Schlutz, Acoustics of the Concert Hall. *Spectrum IEEE*, 56–67 (June 1965)
10. K. Yoshida, M. Kazama, M. Tohyama, Pitch and speech-rate conversion using envelope modulation modeling. *ICASSP I SP-P04.04* 425–428 (2003)
11. S. Makino (ed), T.-W. Lee, H. Sawada, *Blind Speech Separation* (Springer, 2007)
12. Y. Takahashi, M. Tohyama, Y. Yamasaki, Phase response of transfer functions and coherent field in a reverberation room. *Electron. Commun. Jpn.* **90**(4 Part 3), 1–8 (2007)
13. Y. Hara, Sound perception and temporal-spectral characteristics in sound field near sound source (in Japanese with English abstract). Ph.D. Thesis, Kogakuin University (2014), p. 126
14. P.M. Morse, R.H. Bolt, Sound waves in rooms. *Rev. Mod. Phys.* **16**, 69–150 (1944)
15. Y. Hara, Y. Takahashi, K. Miyoshi, Narrow-band envelope spectral analysis close to sound source (in Japanese). *J. Inst. Elect. Inf. Comm. Eng. Jpn.* **J97-A**(3), 221–223 (2014)
16. M.R. Schroeder, New method of measuring reverberation time. *J. Acoust. Soc. Am.* **37**(3), 409–412 (1965)

# Chapter 5

## Modulation and Periodic Properties of Temporal Envelope

Temporal envelopes are representatives of frequency-dependent signatures of sequences that exhibit the temporal dynamics of signals. This chapter formulates the modulation index, which characterizes the spectral magnitudes of the envelope frequencies normalized. Speech intelligibility is estimated by the modulation index of the narrow-band envelopes. An intriguing question is whether the magnitude or phase spectrum is dominant in synthesizing intelligible speech. Interestingly, it depends on the frame-length. Thus the phase is dominant when the frame length is very long or extremely short. A time-reversed speech sample is used as a good example demonstrating the phase effect on intelligibility. The envelopes also provide estimates of the fundamental period of a periodic sequence. This chapter introduces the period estimation of a periodic scheme when the fundamental of a signal is missing.

### 5.1 Modulation Spectrum of Envelope

#### 5.1.1 *Magnitude Spectrum and Modulation Index for Speech Envelope*

Speech intelligibility can be estimated using narrow-band envelopes rather than the narrow-band carriers [1–3]. Actually, if the narrow-band envelopes are taken from intelligible speech samples, intelligibility is preserved almost independent of the narrow-band carriers [4, 5]. In detail, if a sequence is produced for each narrow band, such as

$$x_j(n) = e_j(n)c_j(n) \quad (5.1)$$

the narrow-band carrier  $c_j(n)$  can be substituted by a sinusoidal sequence with a frequency equal to the  $j$ -th narrow-band center frequency, or even by a corresponding narrow-band noise, to synthesize intelligible speech. This fact indicates that intelligibility, the prime signature of speech, is conveyed in the temporal and frequency-band dependent dynamical characteristics of sequences rather than the spectral fine structure [3]. Of course, speech quality is improved by choosing narrow-band carriers, for example, sinusoidal carriers that are determined by the maximal-magnitude spectral components of the frequency spectrum in every short frame [5, 6].

The modulation index is a good estimator for speech intelligibility [2, 3, 6]. Consider a modulated signal for a sinusoidal sequence  $A_e \cos(\Omega_e n + \phi_e)$ . The modulated sequence is given by  $x(n) = A_e \cos(\Omega_e n + \phi_e) \cdot c(n)$  where  $c(n)$  is a random carrier. Take the ensemble average for the square of the sequence,

$$\begin{aligned} E[x^2(n)] &= A_e^2 \cos^2(\Omega_e n + \phi_e) \cdot E[c^2(n)] \\ &= \frac{A_e^2}{2} (1 + \cos 2(\Omega_e n + \phi_e)) \end{aligned} \quad (5.2)$$

where the average is normalized using  $E[c^2(n)] = 1$ . Here,  $2\Omega_e$  denotes the angular frequency of the squared envelope. In addition, this squared envelope is modified so that

$$E_m(n) = \frac{1}{2} (1 + m \cos 2(\Omega_e n + \phi_e)) \quad (5.3)$$

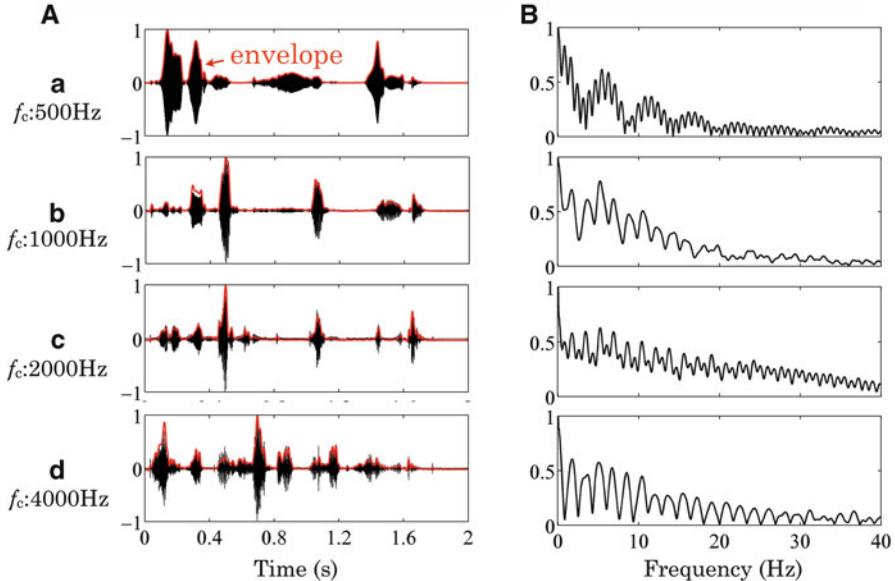
with  $m$  the modulation index [2]. This index can be generalized to envelopes composed of multiple frequencies such as

$$x(n) = \sum_i A_{e_i} \cos(\Omega_{e_i} n + \phi_{e_i}) \cdot c_i(n). \quad (5.4)$$

The ensemble average for the squared envelope can be written as

$$E_m(n) = \frac{K}{2} \left( 1 + \sum_i m_i \cos 2(\Omega_{e_i} n + \phi_{e_i}) \right), \quad (5.5)$$

where  $m_i = A_{e_i}^2 / K$  determines the modulation index for the  $i$ -th envelope frequency component, and  $K = \sum_i A_{e_i}^2$ . Figure 5.1 on the facing page displays a sample of the envelope frequencies [2]. The modulation index can be understood as the normalized magnitude spectrum of the envelope [2]. That is, intelligible speech can be delivered to an entire audience even in a large hall if the magnitude spectrum (or the modulation index) for each narrow-band envelope is not degraded over the transmission path.



**Fig. 5.1** Sample of envelope frequencies [2]. (A) Narrow band speech waveform. (B) Squared envelope spectrum

### 5.1.2 Phase Spectrum, Cross-Correlation, and Modulation Index for Envelopes

The dissimilarity in the envelope compared with the original also could be a candidate for estimating intelligibility of speech along with the modulation index. One way to measure this dissimilarity is analyzing the cross-correlation between a pair of squared envelopes [3,6]. Suppose that the squared envelopes are written

$$\begin{aligned} E_{Am}(n) &= 1 + \sum_i m_{Ai} \cos 2(\Omega_{eAi} n + \phi_{eAi}), \\ E_{Bm}(n) &= 1 + \sum_i m_{Bi} \cos 2(\Omega_{eBi} n + \phi_{eBi}), \end{aligned} \quad (5.6)$$

for envelopes  $A$  and  $B$ , respectively. The cross-correlation coefficient between the two envelopes is defined as

$$r_{AB} = \frac{\text{E}[X_A \cdot X_B]}{\sqrt{\text{E}[X_A^2] \cdot \text{E}[X_B^2]}} \quad (5.7)$$

where

$$X_A = E_{Am}(n) - \mu_A \quad (5.8)$$

$$X_B = E_{Bm}(n) - \mu_B \quad (5.9)$$

and  $\mu_A$  ( $\mu_B$ ) denotes the average of the envelope  $E_{Am}$  ( $E_{Bm}$ ). Here, note that the envelope correlation is not determined solely by the modulation index or the magnitude spectrum. The envelope phase spectrum is also key to the envelope correlation being 0 if the phase difference between the two envelopes is  $\pi/2$ .

Consider a sum of two squared envelopes such as

$$E_{sm} = A^2 \cos^2(\Omega_e n) + A^2 \cos^2(\Omega_e n + \phi_e) \quad (5.10)$$

where there is a phase difference  $\phi_e$  between the two modulating sinusoids. The sum above can be rewritten

$$E_{sm} = A^2 (1 + \cos \phi_e \cos(2\Omega_e n + \phi_e)). \quad (5.11)$$

This expression indicates that the phase difference between the two envelopes changes not only the envelope correlation but also the modulation index.

### 5.1.3 Complex Modulation Transfer Function

The modulation index determines the magnitude spectral property of the envelope, whereas the phase spectral difference between the two envelopes is key in estimating the cross-correlation coefficients between the two. However, if the phase difference between the two envelopes is considered as that between input and output sequences of a linear system, the corresponding changes in the magnitude and phase spectral characteristics can be determined using the impulse response of the linear system, subject to assuming the carrier for the input sequence random noise [7]. Suppose that a modulated input sequence is given,

$$x(n) = \cos \Omega_e n \cdot r(n) \quad (5.12)$$

where  $r(n)$  denotes random noise, and

$$\mathbb{E}[r(n)] = 0 \quad (5.13)$$

$$\mathbb{E}[r(m)r(m-n)] = N\delta(n) \quad (5.14)$$

$$\mathbb{E}[r^2(n)] = N. \quad (5.15)$$

The output sequence from the linear system is written as

$$y(n) = \sum_m h(m) \cos \Omega_e (n-m) r(n-m) \quad (5.16)$$

where  $h(n)$  denotes the impulse response of the linear system. Therefore, taking the ensemble average of the squared output sequence,

$$\begin{aligned} \text{E}[y^2(n)] &= N \sum_m h^2(m) \frac{1 + \cos 2\Omega_e(n-m)}{2} \\ &= \Re \left[ \frac{N}{2} \sum_m h^2(m) (1 + m_c e^{i2\Omega_e n}) \right] \end{aligned} \quad (5.17)$$

is obtained where

$$m_c(2\Omega_e) = \frac{\sum_m h^2(m) e^{-i2\Omega_e m}}{\sum_m h^2(m)} \quad (5.18)$$

called the complex modulation transfer function, the magnitude of which yields the modulation index [7]. Note that the complex modulation transfer function determines both the modulation index and the phase change of the respective envelopes for magnitude and phase.

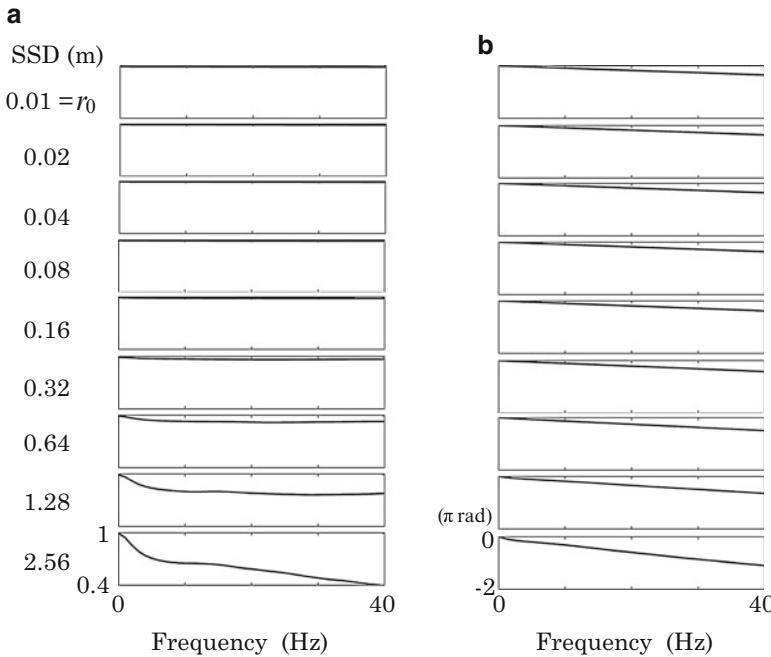
The Fourier transform of a squared impulse response determines the complex spectral correlation (or auto-convolution of the complex spectral function) for the frequency characteristics of the linear system of interest. The auto-correlation functions or sequences can be defined over both temporal and frequency domains. The auto-correlation sequence over the time domain gives the power spectral density function over the frequency domain, whereas the auto-correlation function (or sequence) over the frequency domain yields the envelope (or instantaneous magnitude) of the squared time sequence.

### 5.1.4 Temporal and Spatial Range of Direct Sound Envelope

As illustrated in Fig. 5.1 on page 79, which displays a sample of the envelope frequencies and the modulation index [2], the most noticeable envelope frequency is around 4–5 Hz. Thus, the time delay that produces phase difference  $\pi/2$  for a 5 Hz component is 50 ms that corresponds to the classical parameter so-called  $D_{50}$  giving the energy ratio between the direct sound followed by early reflection within 50 ms and the total sound.

Figure 5.2 on the following page presents a sample of the complex modulation transfer function for the impulse response illustrated by Fig. 5.3 on page 83. This result indicates that the source effect (the spatial range of direct sound) on the modulation index of the envelope can also be estimated by  $D_{30}$  (or  $D_{50}$ ) as shown in Fig. 5.4 on page 83.

Interestingly, the phase characteristics in Fig. 5.2 on the following page indicate mostly the linear phase even when the distance becomes far from the source position in the reverberation field. It means the entire envelope of the envelope of the source



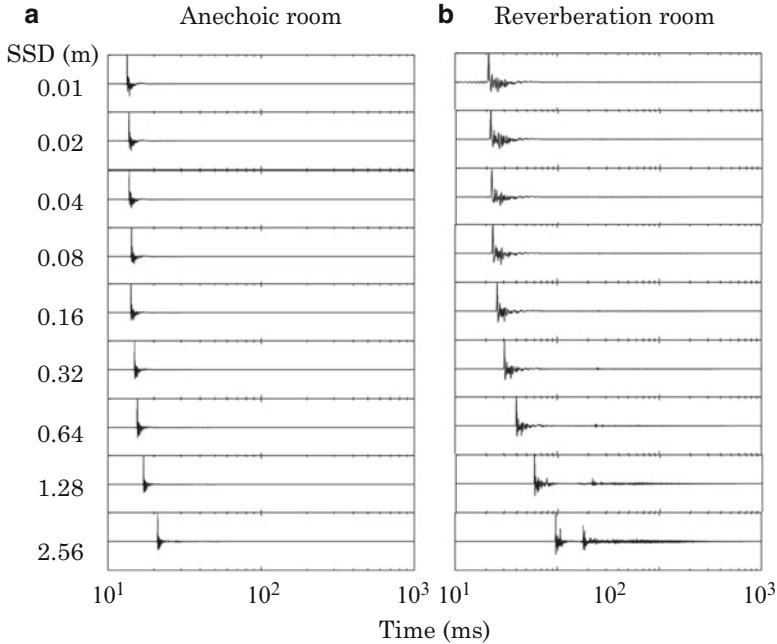
**Fig. 5.2** Sample of a complex modulation transfer function in reverberation room: magnitude and phase spectral responses at receivers. **(a)** Magnitude (modulation index). **(b)** Phase

signal is mostly preserved in the reverberation field, even when the magnitude of the complex modulation transfer function greatly changes in the field. Figure 5.5 on page 84 presents a sample of speech envelopes in the reverberation field in which the entire waveform of the envelope is mostly maintained even under the reverberation condition. It also explains that recognizable speech could be circulated in the reverberation field [8].

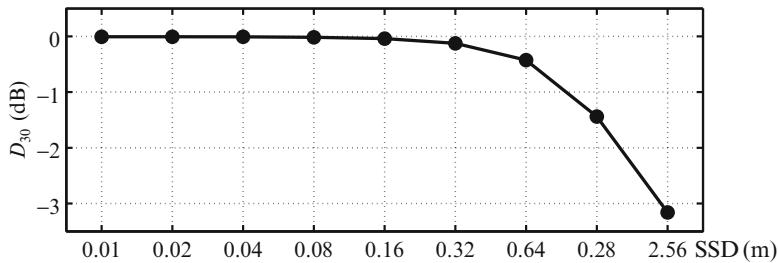
## 5.2 Narrow-Band Envelopes and Speech Intelligibility

### 5.2.1 Envelope Recovery and Speech Intelligibility

As mentioned in Sect. 5.1.1 on page 77, speech intelligibility can be estimated by analyzing narrow-band envelopes [1–3]. In particular, the cross-correlation coefficients between two envelopes defined by Eq. 5.7 on page 79 is a good measure for estimating the intelligibility for synthesized speech. Figure 5.6 on page 85 shows, according to Eq. 5.7 on page 79, the recovery of narrow-band envelopes of synthesized speech that is reconstructed on a frame-wise basis using the original

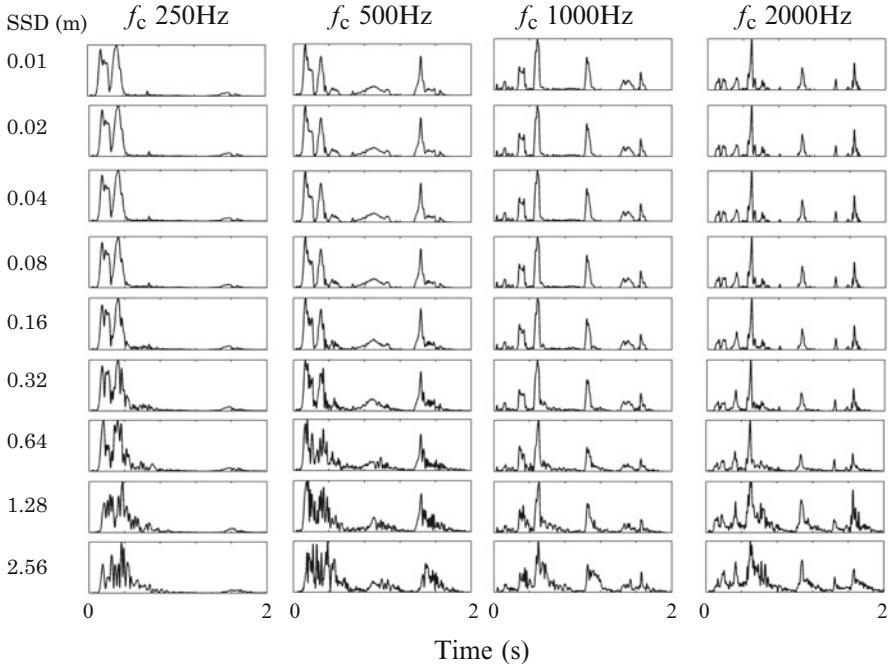


**Fig. 5.3** Samples of impulse responses from a loudspeaker system in an anechoic room (a) and a reverberation room (b) with varying distance between source and receiver



**Fig. 5.4** Energy ratio  $D_{30}$  for responses in Fig. 5.3

magnitude (phase) spectrum with a random phase (magnitude) one [3, 6]. The schematic for speech reconstruction is outlined by Fig. 5.7 on page 85 [3, 6]. Figure 5.6 on page 85 demonstrates speech intelligibility mainly follows the cross-correlation coefficients for the narrow-band envelopes between the original and synthesized speech samples in every band. Interestingly, the phase, rather than the magnitude, spectral property is dominant in preserving intelligibility under very short or very long frame-length conditions [3, 6, 9, 10]. In particular, the phase dominance under very short frame lengths implies that the periodic nature of the zero-crossing points in the time domain (sign changes of the waveform),



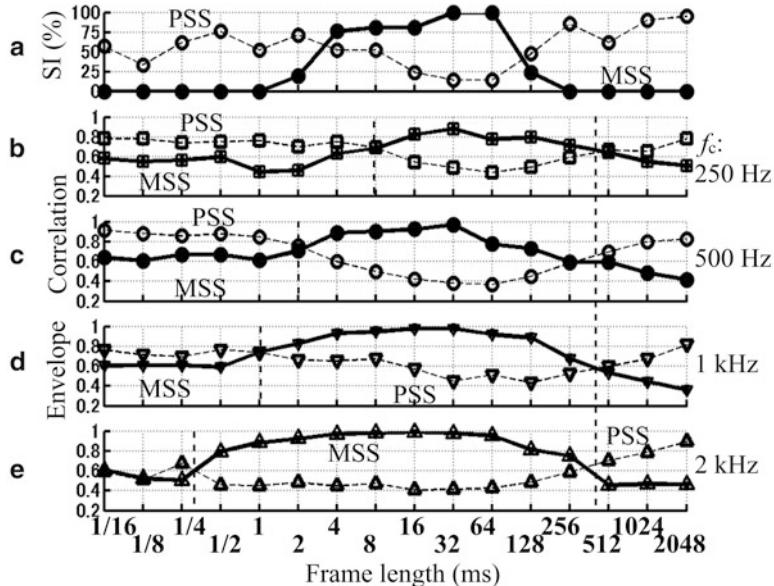
**Fig. 5.5** Sample of speech envelopes in reverberation field

which is essential in recovering the modulation property, can be preserved by the phase spectral property. Detailed analysis of the results on speech intelligibility is described in [3, 6].

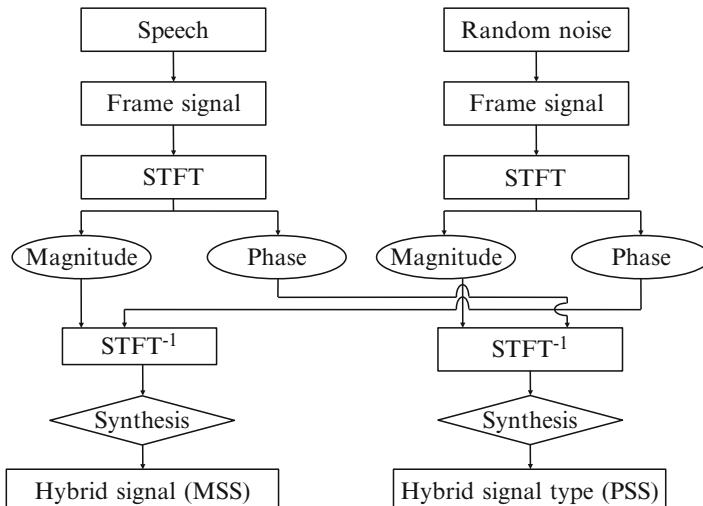
### 5.2.2 Phase Effect on Speech Intelligibility for Time-Reversed Speech

Effects of phase on speech intelligibility or recovery of narrow-band envelopes as described above are very noticeable, in particular, for the very short or long frame lengths. The phase effects demonstrated in Fig. 5.6 on the facing page are due to randomizing the phase spectra even while retaining the original magnitude spectra of the speech sample.

Moreover, regarding speech intelligibility, another type of phase effect arising from opposite signs in the phase condition of the original is also noteworthy. Consider a sequence  $x(n)$  representing a spoken sentence and its Fourier transform  $X(e^{-i\omega})$ . Taking the complex conjugate of the Fourier transform, the following sequence



**Fig. 5.6** (a) Speech intelligibility for synthesized speech using magnitude spectral speech (MSS) using the original magnitude with random phase (*solid line*) or using phase spectral speech (PSS) using the original phase with random magnitude spectrum (*broken line*); (b–e) cross-correlation coefficients from Eq. 5.7 on page 79 between the original and synthesized narrow-band envelopes, taken from Fig. 4 of [3]



**Fig. 5.7** Schematic for synthesizing speech on frame-by-frame basis from Fig. 1 of [3]

$$\begin{aligned} Y(e^{-i\Omega}) &= X^*(e^{-i\Omega}) \\ &= \Re[X(e^{-i\Omega})] - i\Im[X(e^{-i\Omega})] \end{aligned} \quad (5.19)$$

is then obtained. Therefore, the waveform reconstructed from the complex conjugate spectrum becomes

$$y(n) = x(-n). \quad (5.20)$$

This sequence is called the time-reversed sequence. Here, note that time-reversed speech has the identical magnitude spectrum to the original speech, but its phase spectrum has just the opposite sign to the original phase. This phase change is due to time reversal [11, 12].

The intelligibility of the time-reversed speech is almost completely lost because of the phase effect. In addition, this phase effect can also be seen in the phase-spectral recovery of the envelope, whereas no changes occur in the magnitude spectrum of the envelope. Figure 5.8 on the facing page illustrates an example of cross-correlation coefficients for narrow-band envelopes using Eq. 5.7 on page 79 between the normal and time-reversed speech samples. The cross-correlation coefficients are significantly low, and this fact shows that the time-reversed speech might be almost perfectly unintelligible.

The decrease in the envelope correlations can be explained by phase differences between the two envelopes. Specifically, the phase difference is given by  $2\phi$  at every frequency for the two corresponding envelopes, where  $\phi$  denotes the phase of the original envelope, and thus the average for  $\cos 2\phi$  might significantly decrease from unity at every frequency band. In other words, the loss of intelligibility for a time-reversed speech sample cannot be interpreted according to the modulation index. This is because there is no difference in the magnitude spectral components of the envelopes between the original and time-reversed speech samples. Instead, it can be understood from the phase difference between the time-reversed and the original phase spectra of the envelope.

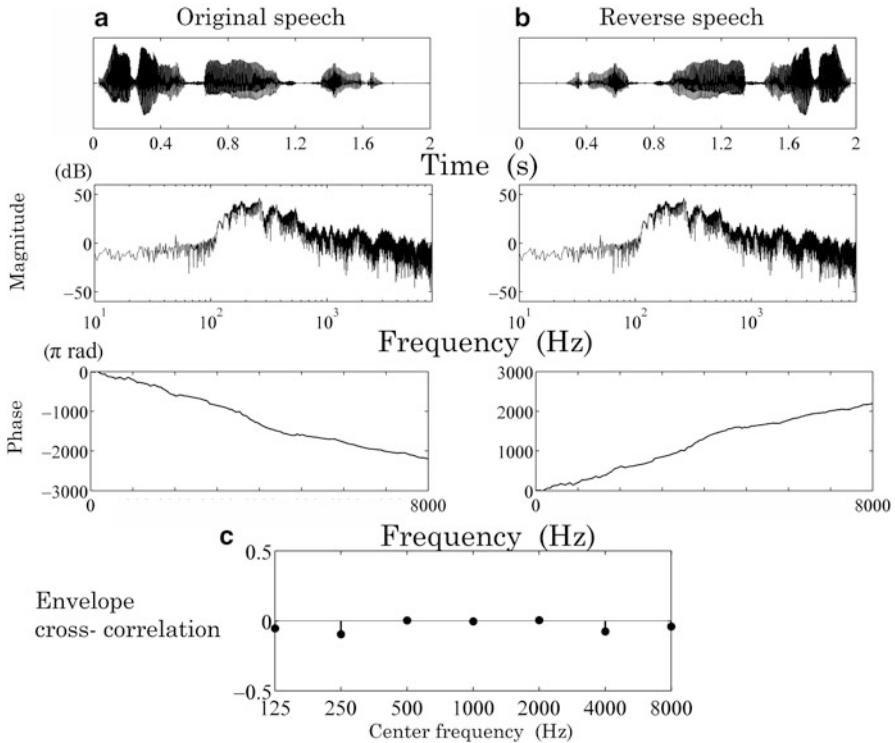
A time-reversed speech sample can be one of ideal maskers (for example, see [13]). Intuitively, this might be understood by recalling that the sum of the speech sequence  $x(n)$  and its time-reversed sequence  $x(-n)$  creates a new symmetric sequence with respect to the central position on the time axis. This kind of masking, which is produced by the masker's phase condition rather than the magnitude (or power) spectral condition, is often called informational masking [14].

## 5.3 Fundamental Frequency and Period of Envelopes

### 5.3.1 Auto-correlation of Sequence and Period Estimation

If a sequence  $x(n)$  satisfies the condition

$$x(n) = x(n + pM) \quad (5.21)$$



**Fig. 5.8** Cross-correlation coefficients for narrow-band envelopes between original (**a**) and time-reversed (**b**) speech samples where 30 pairs are used for averaging; cross-correlation coefficients (**c**) are given using Eq. 5.7 on page 79

where  $p$  is an integer, the sequence is called periodic with period  $M$ . In general, the period or the fundamental frequency, defined by the inverse of the period, is one of the representative signal signatures. However, in the real world, there are no periodic sequences observed from a strict mathematical point of view; moreover, sinusoidal waves are not seen. Therefore, the period can only be approximately estimated from a practical or perceptual point of view.

Suppose a sequence is composed of  $K$  sinusoidal components as in

$$x(n) = \sum_{k=1}^K A_k \cos(\Omega_k n + \phi_k) \quad (5.22)$$

where  $A_k > 0$  and  $\phi_k$  denote the magnitude and initial phase of the  $k$ -th sinusoidal component respectively, and  $\Omega_k$  denotes the  $k$ -th angular frequency in the series  $\Omega_1 < \Omega_2 < \dots < \Omega_K$ . A sequence expressed as above is called a compound sound or compound wave. If the angular frequency is given by

$$\Omega_k = m_k \cdot \Omega_0 \quad (5.23)$$

then the sequence  $x(n)$  is periodic and its period is given by  $T = 2\pi/\Omega_0$ . Here,  $F_0 = \Omega_0/2\pi$  is called the fundamental frequency, and  $F_k = m_k F_0$  is the  $m_k$ -th harmonic frequency and  $m_k$  is a positive integer. Similarly, the sinusoidal component of the fundamental frequency is called the fundamental, and its higher-frequency components are called harmonics. A periodic wave without harmonics is a sinusoidal wave [15].

Note that the fundamental frequency represents the greatest common divisor among the frequencies of the sinusoidal components that comprise the sequence  $x(n)$ . Therefore, the fundamental frequency can be determined even if the fundamental frequency is not contained in the sequence  $x(n)$  itself. In addition, the fundamental, in principle, is independent of the magnitude and phase spectrum. However, a sequence that is observed or recorded in a practical situation is rarely periodic, so the greatest common divisor is strictly not determined from a mathematical point of view. Therefore, in practice, another way must be taken to proceed.

The auto-correlation sequence is a candidate as intuitively understood by the definition. The phase spectral information is discarded, so that all phase angles are instead null for the auto-correlation sequence. Consequently, the power spectral information is left in the auto-correlation sequence, and thus the power spectral property must have an influence on the fundamental frequency or the period estimation.

Suppose a compound sequence  $x(n)$  is given by Eq. 5.22 on the previous page. Its auto-correlation sequence can be written as

$$r(n) = \frac{1}{2K} \sum_{k=1}^K A_k^2 \cos \Omega_k n. \quad (5.24)$$

Here, setting

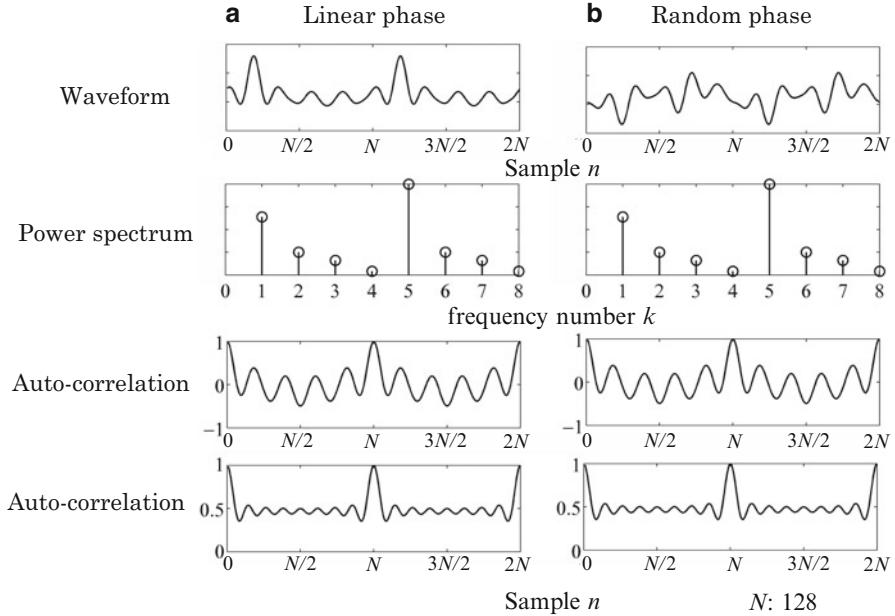
$$\frac{1}{2} A_k^2 = 1 \quad (5.25)$$

then to erase effects of the power spectral condition on the period estimation, the auto-correlation for the newly normalized sequence becomes

$$\hat{r}(n) = \frac{1}{K} \sum_{k=1}^K \cos \Omega_k n. \quad (5.26)$$

Note again that the fundamental frequency or period of the sequence  $x(n)$  can be estimated from the period of Eq. 5.26 independent of the power spectrum  $A_k^2$ .

Figure 5.9 on the facing page displays samples of periodic sequences and their auto-correlation sequences. Column (a) shows a sequence with phase  $\phi_k = -\Omega_k n_0$  where  $n_0 > 0$  is an integer, and, in column (b),  $\phi_k$  is randomly taken in the interval from 0 to  $2\pi$ . A phase that is proportional to the frequency, as in column (a), is called

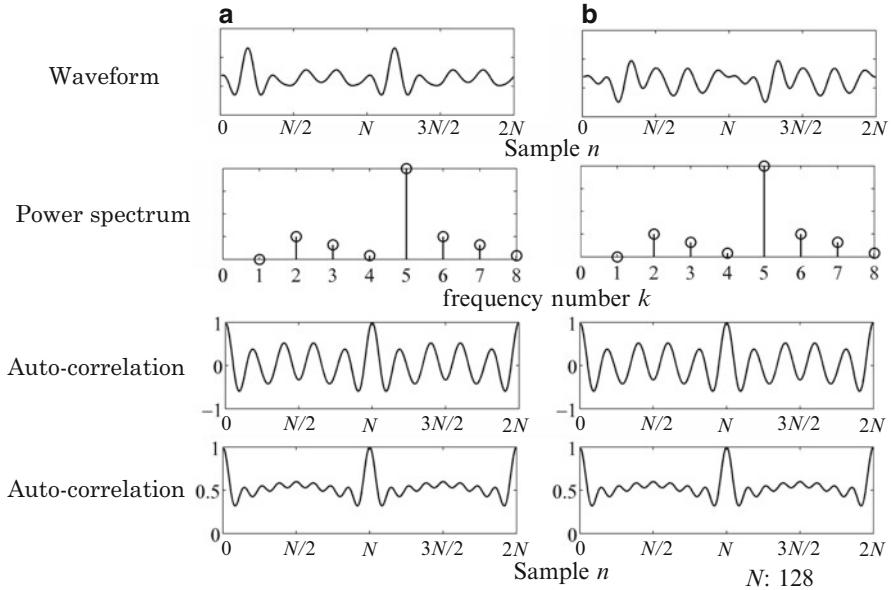


**Fig. 5.9** Samples of auto-correlation sequences for periodic compound sequences (period:  $N$ ); (a) linear phase (see text), (b) random phase (see text), top row: waveform, second row: power spectrum, third and bottom rows: auto-correlation sequences obtained using Eqs. 5.24 on the preceding page and 5.26 on the preceding page, respectively

linear. The top row shows the waveforms of the sequences, whereas the second row illustrates the power spectrum  $A_k^2$ . The third and bottom rows display the auto-correlation sequences given by Eqs. 5.24 on the preceding page and 5.26 on the preceding page, respectively. Clearly, the auto-correlation sequence can be obtained from 5.26 on the preceding page because the flat-spectral sequence gives the period of the sequence by removing spectral effects, subject to the requirement that all frequencies of the sinusoidal components of the sequence of interest can be estimated.

### 5.3.2 Missing Fundamental

The fundamental frequency can be defined even when the corresponding sinusoidal component is not contained in the sequence, that is, under missing fundamental conditions [16]. The fact that fundamental frequencies are perceived in sounds, even if missing, supports the time theory over the place (or spectral) theory [16] in modeling how ears sense the fundamental frequency. Figure 5.10 on the following page confirms that the period (or the fundamental frequency) is estimated by Eq. 5.26 on the preceding page, that is, using the auto-correlation sequence for the flat spectral sequence under the missing fundamental condition.



**Fig. 5.10** Samples of auto-correlation sequences under the missing fundamental condition; all other conditions are similar to those in Fig. 5.9 on the previous page. **(a)** Linear phase. **(b)** Random phase

### 5.3.3 Period of Envelope

The fundamental frequency or the period is an important signal signature from a perceptual point of view. The pitch perception of sound is a principle attribute of the ear; the basic structure of music is the musical scales constructed naturally from fundamental frequencies. Research-wise, there have been many studies on finding the mechanism underlying the pitch perception of the ear [16]. In general, the pitch sensed by a periodic sequence is basically understood as corresponding to the fundamental frequency of the sequence. Again, from the perceptual point of view, narrow-band envelopes are taken instead of the sequence itself [16].

Recall the sinusoidal modulation formula

$$\begin{aligned} x(n) &= \left( \frac{1 + \cos 2\Omega_e n}{2} \right) \cdot \cos \Omega_c n \\ &= \frac{1}{2} \left( \cos(\Omega_c n) + \frac{1}{2} \cos(\Omega_c + \Delta\Omega)n + \frac{1}{2} \cos(\Omega_c - \Delta\Omega)n \right) \end{aligned} \quad (5.27)$$

where  $2\Omega_e$  and  $\Omega_c$  denote the angular frequencies for the envelope and carrier, for which  $\Omega_e \ll \Omega_c$  holds, and  $\Delta\Omega = 2\Omega_e$ . The latter relation states that the envelope frequency is equal to the frequency spacing of the three sinusoidal components.

Consider a pair of adjacent harmonic components of angular frequencies  $\Omega_m$  and  $\Omega_{m+1} = \Omega_m + \Omega_0$ . Here, the frequency spacing,  $\Omega_0 = \Delta\Omega$ , can be interpreted as the envelope angular frequency of the waveform composed of the harmonics. Consequently, the fundamental frequency giving the spacing between the harmonics can be estimated from the period of the envelope.

With this in mind, the model for the pitch perception scheme using period analysis of narrow-band envelopes [16], can be intuitively understood. Again, consider the periodic sequence for which the power spectral components are represented as a series of line spectral entries. Recall here that the frequency spacing of the components determines the fundamental frequency.

A partial sum of the harmonic components can be represented essentially by a compound sequence or by envelope-modulation modeling [6]. The sinusoidal modulation formula expressed by Eq. 5.27 on the preceding page can be extended to a multi-sinusoidal modulation. Suppose the following sequence is given

$$\begin{aligned} x(n) &= \left( K + \sum_{k=1}^K A_k \cos(2(k\Omega_e n + \phi_{e_k})) \right) \cos \Omega_c n \\ &= K \cos \Omega_c n + \frac{1}{2} \sum_{k=1}^K A_k [\cos((\Omega_c + 2k\Omega_e)n + 2\phi_{e_k}) \\ &\quad + \cos((\Omega_c - 2k\Omega_e)n - 2\phi_{e_k})]. \end{aligned} \quad (5.28)$$

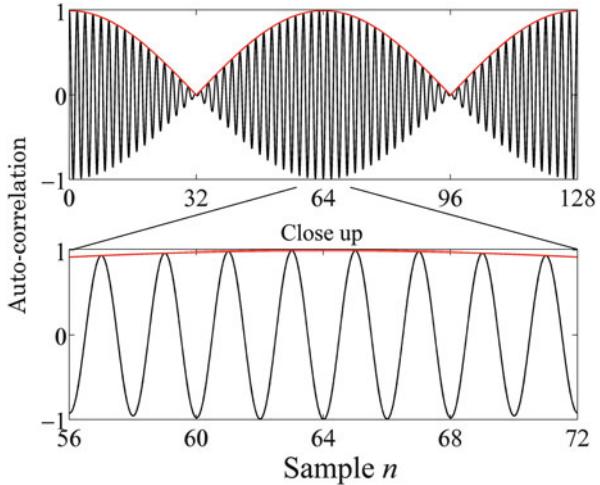
The formulation above shows a series of harmonic components, and thus the multi-sinusoidal modulation formula can also be interpreted as an expression for a partial sum of a periodic compound sequence. Here, the spectral spacing, that is,  $2\Omega_e = \Omega_0$  is estimated from the period of the envelope. This result underpins the pitch estimation model [16], in which the partial sum of the harmonics is interpreted as a narrow-band sequence and consequently, the envelope can be understood as the narrow-band envelope.

Unfortunately, the envelope scheme does not always work well in estimating the fundamental frequency [17]. A counter example occurs when the frequency interval between adjacent sinusoidal components is wider than the fundamental frequency even in a narrow frequency band [17]. Consider a pair of sinusoidal components with angular frequencies  $k_1\Omega_0$  and  $k_2\Omega_0$ . The auto-correlation sequence can be written as

$$\begin{aligned} r(n) &= \frac{1}{2} (\cos k_1 \Omega_0 n + \cos k_2 \Omega_0 n) \\ &= \cos \frac{k_1 - k_2}{2} \Omega_0 n \cos \frac{k_1 + k_2}{2} \Omega_0 n \end{aligned} \quad (5.29)$$

where  $\Omega_0$  denotes the fundamental angular frequency, and  $k_1$  and  $k_2$  are positive integers. The equation above indicates that the periodic envelope does not

**Fig. 5.11** Period estimation using the auto-correlation sequence for a pair of sinusoidal components of which the frequency interval is wider than the fundamental frequency;  $\Omega_0 = 2\pi/128$ : fundamental angular frequency,  $\Omega_1 = 63\Omega_0$  and  $\Omega_2 = 65\Omega_0$  are the angular frequencies for the sinusoidal components



have a fundamental angular frequency  $\Omega_0$ , when  $|k_1 - k_2| > 1$ , although the auto-correlation sequence itself exhibits this fundamental frequency, as shown in Fig. 5.11.

Auto-correlation analysis can be applied to estimate the frequency for a sequence composed of harmonic sequences such as a musical chord. Suppose three harmonic sequences are given:

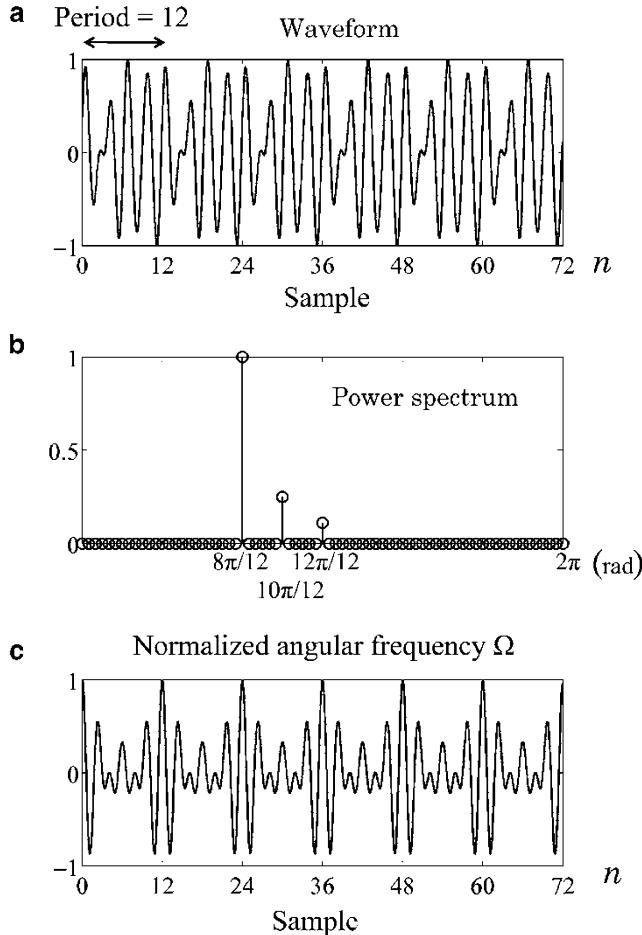
$$x_1(n) = \sum_{k=1}^I A_k \cos k\Omega_1 n \quad (5.30)$$

$$x_2(n) = \sum_{k=1}^J A_k \cos k\Omega_2 n \quad (5.31)$$

$$x_3(n) = \sum_{k=1}^K A_k \cos k\Omega_3 n \quad (5.32)$$

where  $\Omega_1 = h\Omega_0$ ,  $\Omega_2 = i\Omega_0$ ,  $\Omega_3 = j\Omega_0$ , and  $h < i < j$ . Taking the superposition of the above three harmonic sequences, the harmonic sequence  $x(n) = x_1(n) + x_2(n) + x_3(n)$  has fundamental angular frequency  $\Omega_0$ . Actually, the period of the harmonic sequence  $x(n)$  is determined by the fundamental frequency  $\Omega_0$ , that is, the greatest common divisor among  $h\Omega_0$ ,  $i\Omega_0$ , and  $j\Omega_0$ .

The fundamental angular frequency  $\Omega_0$  can be estimated using the auto-correlation sequence in which all the power spectral components are normalized (flattened) to unity. Figure 5.12 on the facing page shows an example of an estimation of the fundamental frequency for a compound harmonic sequence; panel (a) illustrates the waveform, panel (b) is the power spectral components, and



**Fig. 5.12** Period estimation for harmonic sequences given by Eqs. 5.30–5.32 where  $\Omega_0 = 2\pi/12$ ,  $h = 4$ ,  $i = 5$ ,  $j = 6$  and  $K = 1$ ; (a) waveform, (b) power spectral sequence, and (c) auto-correlation sequence after normalizing power spectrum to be unity

panel (c) shows the auto-correlation sequence after normalizing the power spectral components. Clearly, the period, or the inverse of the fundamental frequency, can be easily estimated by the auto-correlation sequence in panel (c).

However, the following query might be posed from a perceptual point of view. Figure 5.12 shows a case when the frequency ratio of the fundamentals for the three harmonic components is given as  $4 : 5 : 6(h : i : j)$  as for a major tonic chord. Therefore, the fundamental frequency is just two octaves lower than that of the lowest fundamental of the three harmonic sequences. Probably, the estimated fundamental frequency might not be acceptable as an estimation of pitch perception from a perceptual point of view.

The methods for signal signature analysis must be chosen depending on the purpose. Again, suppose three sinusoidal frequency components such as  $f_c$ ,  $f_c \pm \Delta f$ , are given. If the goal of the signal analysis is estimating the period (fundamental frequency) or beat, then analyzing  $\Delta f$  is important. Beating is central in tuning musical instruments, such as a piano. The beat period is given as  $1/\Delta f$ . Alternatively, the center frequency might be a significant component needing to be definitively estimated, from a perceptual point of view, when estimated periods appear not to be good measures of perceptual pitch perception.

However, the central component never represents the envelope that yields the beats. Therefore, if the entire waveform is the main interest, independent of the observation intervals of the waveform, the three sinusoidal components must be detected separately, even if the three components exist very closely on the frequency scale. It seems that the ear is good at detecting the signature appropriately.

## References

1. R. Drullman, Temporal envelope and fine structure cues for speech intelligibility. *J. Acoust. Soc. Am.* **97**(1), 585–592 (1995)
2. T. Houtgast, H.J.M. Steeneken, R. Plomp, A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* **77**(3), 1069–1077 (1985)
3. M. Kazama, S. Gotoh, M. Tohyama, T. Houtgast, On the significance of phase in the short term Fourier spectrum for speech intelligibility. *J. Acoust. Soc. Am.* **127**(3), 1432–1439 (2010)
4. K. Yoshida, M. Kazama, M. Tohyama, Pitch and speech-rate conversion using envelope modulation modeling. *ICASSP I SP-P04.04* 425–428 (2003)
5. K. Terada, M. Tohyama, T. Houtgast, The effect of envelope or carrier delays on the precedence effect. *Acustica* **91**(6) 1016–1019 (2005)
6. M. Tohyama, *Sound and Signals* (Springer, 2011)
7. M.R. Schroeder, Modulation transfer functions: definition and measurement. *Acustica* **49**(3), 179–182 (1981)
8. M.R. Schroeder, *Fractals, Chaos, Power Laws* (W.H. Freeman and Company, 1991)
9. A.V. Oppenheim, J.S. Lim, The importance of phase in signals. *Proc. IEEE* **69**(5), 529–541 (1981)
10. L. Liu, J. He, G. Palm, Effects of phase on the perception of intervocalic stop consonants. *Speech Commun.* **22**, 403–417 (1997)
11. K. Saberi, D.R. Perrot, Cognitive restoration of reversed speech. *Nature* **398**, 760 (29 April 1999)
12. L. Longworth-Reed, E. Brandewie, P. Zahorik, Time-forward speech intelligibility in time-reversed rooms. *J. Acoust. Soc. Am. Express Lett.* **125**(1), EL13–EL19 (2009, published online 22 December 2008)
13. Y. Hara, M. Tohyama, K. Miyoshi, Effects of temporal and spectral factors of maskers on speech intelligibility. *Appl. Acoust.* **73**, 893–899 (2012)
14. R. Lutfi, How much masking in informational masking? *J. Acoust. Soc. Am.* **88**(6), 2607–2610 (1990)
15. H.v. Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage fuer die Theorie der Musik, On the Sensation of Tone* (translated by A.J. Ellis) (Dover, 1954)
16. R. Meddis, L. O'Mard, A unitary model of pitch perception. *J. Acoust. Soc. Am.* **102**(3), 1811–1820 (1997)
17. M.R. Schroeder, *Number Theory in Science and Communication* (Springer, 1997)

# Chapter 6

## Transfer Function of Linear Systems

The transfer function is fundamental in describing linear systems. This chapter describes introductory but basic aspects of transfer functions such as minimum-phase and non-minimum-phase decomposition including the introduction of cepstral sequences from a point of view of waveform analysis. The logarithm of the transfer function is the starting point for the analytical formulation of the power spectral density function as well as for the magnitude and accumulate-phase spectral responses. In particular, the geometrical interpretation of the phase is helpful in understanding the discontinuous phase function in the limit. Filtering is a central issue of the linear system theory. This chapter describes ideal low-pass filtering in terms of the Dirichlet kernel.

### 6.1 Zeros of the Transfer Function

#### 6.1.1 *z*-Transform of Impulse Response

Let  $h(n)$  denote the impulse response of a linear system. The  $z$ -transform of the impulse response  $H(z^{-1})$  is called the transfer function of the linear system. Assuming the impulse response has a finite record length  $N$ , the transfer function  $H(z^{-1})$  is factorable

$$H(z^{-1}) = K \prod_{l=1}^{N-1} (1 - a_l z^{-1}) \quad (6.1)$$

---

The original version of this chapter was revised.  
An erratum to this chapter can be found at DOI [10.1007/978-4-431-54424-1\\_10](https://doi.org/10.1007/978-4-431-54424-1_10)

where  $z = a_l$  is called a zero for the transfer function as

$$H(z^{-1})|_{z=a_l} = 0 \quad (6.2)$$

holds. Moreover, the transfer function for a finite-length impulse response can be specified by its zeros up to a constant  $K$ .

In contrast, poles or singularities where the transfer function cannot be defined are also found in general. However, the transfer function for a finite-length impulse response has only trivial poles at  $z = 0$  as seen in the factorized form above.

### 6.1.2 Magnitude Frequency Response for Zero

Consider an impulse response composed of a single echo that is expressed by

$$h(n) = 1 + a\delta(n - m). \quad (6.3)$$

The  $z$ -transform for the impulse response becomes

$$H(z^{-1}) = 1 + az^{-m} \quad (6.4)$$

where  $a$  is assumed to be a real number, and thus the zeros are located at

$$z = z_0 = \begin{cases} |a|^{1/m} e^{j \frac{(2l+1)\pi}{m}}, & a > 0, \\ |a|^{1/m} e^{j \frac{2l\pi}{m}}, & a < 0 \end{cases} \quad (6.5)$$

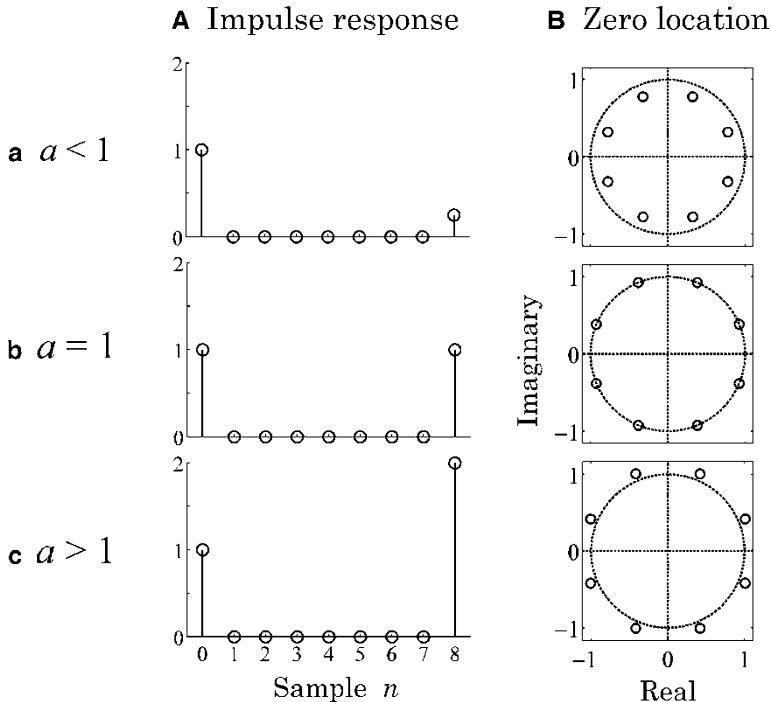
and  $0 \leq l \leq m - 1$ . Factorization and the identification of all zeros are closely related.

Figure 6.1 on the facing page shows the location and arrangement of the zeros of  $H(z^{-1})$  for  $m = 8$  and in (a)  $a < 1$ , in (b)  $a = 1$ , and in (c)  $a > 1$ . That is, when the magnitude of the single echo is smaller (greater) than unity, the zeros are located on a circle inside (outside) the unit circle, and are on the unit circle when the magnitude of the delayed component is equal to the direct component. The  $m$  phase angles of the zeros are equally spaced over the interval  $[0, 2\pi]$ .

### 6.1.3 Power Spectral Density Function for Zeros

The frequency response for the transfer function using Eq. 6.4 becomes

$$\begin{aligned} H(z^{-1})|_{z=e^{j\Omega}} &= 1 + ae^{-im\Omega} \\ &= (1 + a \cos m\Omega) - ia \sin m\Omega. \end{aligned} \quad (6.6)$$



**Fig. 6.1** Sample of zero locations for single echo systems where the magnitude of echo is  $a$  and delay time is  $m = 8$  samples; (a)  $a < 1$ , (b)  $a = 1$ , (c)  $a > 1$ , and magnitude of direct component is normalized to unity. (A) Impulse response. (B) Zero location

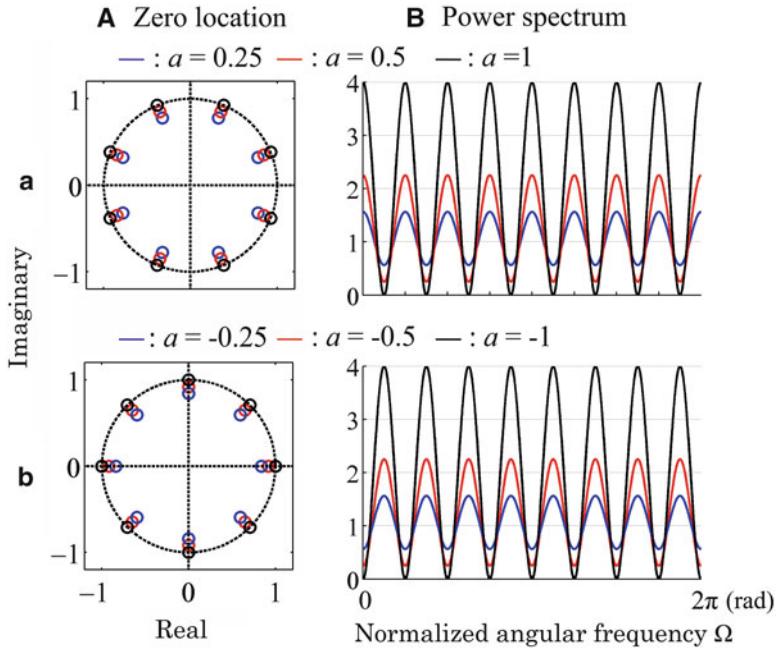
Thus, the power spectral density function is written as

$$|H(e^{-i\Omega})|^2 = (1 + a^2) + 2a \cos m\Omega. \quad (6.7)$$

If  $a = 1$ , the power spectral density function becomes

$$|H(e^{-i\Omega})|^2 = 4 \cos^2 \frac{m}{2} \Omega. \quad (6.8)$$

Figure 6.2 on the following page illustrates examples of the power spectral density functions. It can be seen that the zeros produce a periodic array of troughs in the power spectral density function with the periodicity resulting from the equal spacing of zeros on the circle of radius  $|a|^{1/m}$  as given by Eq. 6.5 on the preceding page. In particular, the troughs are not only minima but null when  $|a| = 1$ . Specifically, the troughs are shallower when the zeros are farther away from the unit circle; conversely, the troughs are deeper when the zeros approach the unit circle.



**Fig. 6.2** Examples of power spectral density functions for single echo system; (A) location of zeros, (B) power spectral density functions where  $m = 8$  and  $a = 0.25, 0.5, 1.0$  for blue, red, and black lines in (a), similarly,  $a = -0.25, -0.5, -1.0$  in (b)

The frequency characteristics given by Eq. 6.6 on page 96 can be rewritten as

$$\begin{aligned} H(e^{-i\Omega}) &= 1 + ae^{-im\Omega} \\ &= ae^{-im\Omega}(1 + a^{-1}e^{im\Omega}). \end{aligned} \quad (6.9)$$

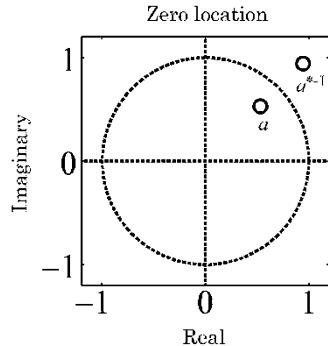
Therefore, the power spectral density function can also be rewritten as

$$\begin{aligned} |H(e^{-i\Omega})|^2 &= |a|^2 |1 + a^{-1}e^{im\Omega}|^2 \\ &= |a|^2 \left| 1 + \frac{1}{a^*} e^{-im\Omega} \right|^2 \end{aligned} \quad (6.10)$$

if  $a$  is extended to include complex numbers. The result above indicates that the power spectral density function is identical between a pair of zeros such as  $z = z_0$  and  $z = (1/z_0)^*$  up to a constant  $|z_0|^2$ . Therefore a pair of zeros, such as those shown in Fig. 6.3 on the facing page, is called a reciprocal complex-conjugate pair of zeros with respect to the unit circle.

From Fig. 6.2, the spacing between adjacent troughs resulting from zeros becomes narrow as the delay time of the echo lengthens ( $m$  becomes large).

**Fig. 6.3** Example of a reciprocal complex-conjugate pair for which the power spectral density function might be the same as that for each zero



This is because the Fourier transform of  $\delta(n-m)$  is given by  $e^{-im\Omega}$ . That is, the zeros determine the frequencies at which the sinusoidal responses become null through the interference between the direct sound and the echo:

$$y(n) = \cos \Omega_0 n + \cos \Omega_0(n-m) = 0. \quad (6.11)$$

This explains why the troughs become deeper or shallower depending on the magnitude of the echo  $|a|$ . That is, cancellation produces in general the minima for

$$y(n) = \cos \Omega_0 n + a \cos \Omega_0(n-m) \quad (6.12)$$

instead of nulls.

The zeros change the power spectral density function from the original of an input sequence to that for the output. Consider input sequence  $x(n)$  and its Fourier transform  $X(e^{-i\Omega})$ . The power spectral density function of the output response  $y(n)$  becomes

$$|Y(e^{-i\Omega})|^2 = |X(e^{-i\Omega})|^2 \cdot |H(e^{-i\Omega})|^2. \quad (6.13)$$

Assuming  $a$  is a real variable, the equation above can be rewritten as

$$|Y(e^{-i\Omega})|^2 = |X(e^{-i\Omega})|^2 [(1 + a^2) + 2a \cos m\Omega]. \quad (6.14)$$

Thus, periodically located minima may appear in the power spectral density function for the output response.

### 6.1.4 Auto-correlation Sequence for Output Response Through a Single Echo System

The minima in the power spectral density function exhibit path information. However, path effects on the perception of sound depend also on the conditions of the source. Consider the auto-correlation sequence for the input sequence  $\rho_{xx}(n)$ . The auto-correlation for the output response is given by

$$\rho_{yy}(n) = \rho_{xx} * \rho_{hh}(n). \quad (6.15)$$

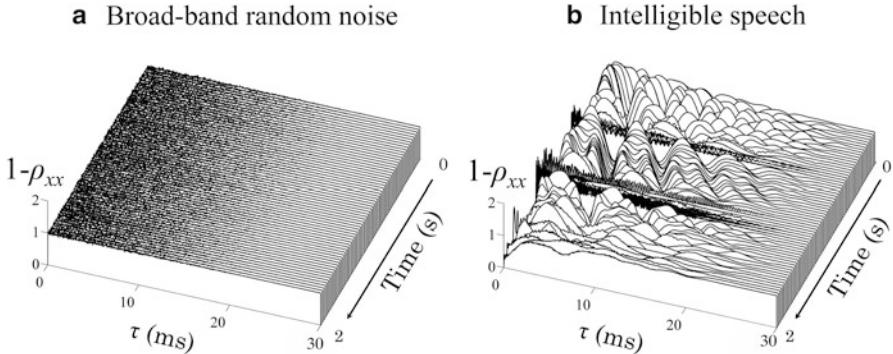
The auto-correlation sequence in the temporal domain might be helpful in understanding the effect of a single echo on the perception of sound rather than the periodically located zeros on the spectral domain [1].

A single echo might be noticeable because the auto-correlation sequence of the input source is separately observed between the direct sound and the single echo. However, a broad-band random noise source might be an exception. The auto-correlation sequence for a broad-band noise source resembles an impulsive sequence. Therefore, the auto-correlation sequences of the output response can be separated into components for the direct and echo portion, even if the delay time of the single echo becomes very short. A separated echo though is not noticeable, and sound can in general be fused into one. This is probably because it is quite unlikely to perceive the difference between the direct and delayed noise samples when a subject listens simultaneously to both sounds. In contrast, a speech sample, for example, is quite likely to be distinguished between the direct and delayed samples if the delay time becomes somewhat long. Such differences between random noise and speech samples might be interpreted as a difference in the temporal envelopes. That is, the envelope might be assumed to be almost stationary for a broad-band random noise, whereas it seems non-stationary for an intelligible speech sample.

Figure 6.4 on the facing page displays examples of frame-wise auto-correlation sequences for (a) broad-band random-noise and (b) intelligible-speech samples. A triangular window with negative slope is applied to every short frame at intervals of 30 ms. Intuitively, the figures exhibit stationarity for (a) random noise and non-stationarity for (b) intelligible speech.

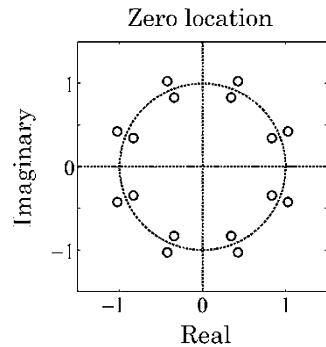
### 6.1.5 Logarithmic Expression for Power Spectral Density Function

The power spectral density function is customarily expressed using the logarithm. Recall that the square of the magnitude for the transfer function can be written as



**Fig. 6.4** Samples of frame-wise auto-correlation sequences for (a) broad-band random noise and (b) intelligible speech samples where a triangular window sequence with negative slope is applied to every short frame of interval 30 ms

**Fig. 6.5** Samples of reciprocal complex-conjugate pairs of zeros



$$\begin{aligned} |H(z^{-1})|^2 &= H(z^{-1}) \cdot H^*(z^{-1}) \\ &= H(z^{-1}) \cdot H(z) \end{aligned} \quad (6.16)$$

for a real impulse response with a finite record length  $N$ . The transfer function  $H(z^{-1})$  can be factorized into a product of factors each of which shows a single zero  $a_l$  such as

$$H(z^{-1}) = K \prod_{l=1}^{N-1} (1 - a_l z^{-1}). \quad (6.17)$$

Thus, the zeros for  $|H(z^{-1})|^2$  are composed of reciprocal complex-conjugate pairs with respect to the unit circle, as shown in Fig. 6.5 where  $N - 1 = 8$ . In addition, because the impulse response is a real sequence, both zeros of the pair lying inside and outside the unit circle are determined by complex conjugate pairs as illustrated in this figure.

Following the symmetric distribution of zeros, the power spectral density function can be rewritten as a product of the form

$$\left|H(z^{-1})\right|^2 = R \cdot F(z^{-1}) \cdot G(z) \quad (6.18)$$

where  $R$  denotes a constant,  $F(z^{-1})$  is characterized by zeros inside the unit circle, whereas for  $G(z)$  the zeros reside outside the unit circle. Factor out a complex conjugate pair of zeros for  $F(z^{-1})$  from its factorized form

$$F(z^{-1}) = \prod_{p=1}^{N-1} (1 - b_p z^{-1}) \quad (6.19)$$

where  $b_p$  denotes the zeros inside the unit circle. That is

$$F_p(z^{-1}) = (1 - b_p z^{-1})(1 - b_p^* z^{-1}). \quad (6.20)$$

The logarithmic function of  $F_p(z^{-1})$  is

$$\log_e F_p(z^{-1}) = \log_e(1 - b_p z^{-1}) + \log_e(1 - b_p^* z^{-1}). \quad (6.21)$$

Recalling that

$$\log_e(1 - x) = - \sum_{n=1}^{\infty} \frac{x^n}{n} \quad (6.22)$$

for  $|x| < 1$ , then

$$\begin{aligned} \log_e(1 - b_p z^{-1}) &= - \sum_{n=1}^{\infty} \frac{b_p^n}{n} z^{-n} \\ \log_e(1 - b_p^* z^{-1}) &= - \sum_{n=1}^{\infty} \frac{(b_p^*)^n}{n} z^{-n}. \end{aligned} \quad (6.23)$$

for  $|b_p z^{-1}| < 1$ [2]. This condition holds also on the unit circle because  $|b_p| < 1$ . Therefore,

$$\begin{aligned} \log_e F_p(z^{-1}) &= - \sum_{n=1}^{\infty} \frac{b_p^n + (b_p^*)^n}{n} z^{-n} \\ &= -2 \sum_{n=1}^{\infty} \frac{\Re[b_p^n]}{n} z^{-n} \end{aligned} \quad (6.24)$$

indicating that the logarithm  $\log_e F_p(z^{-1})$  can be interpreted as the  $z$ -transform of the time sequence for  $n \geq 1$  for

$$f_p(n) = -2 \frac{\Re[b_p^n]}{n}. \quad (6.25)$$

Note here

$$f_p(n) = -\frac{\Re[b_p^n]}{n}. \quad (6.26)$$

if  $b_p$  is real. Consequently, the logarithmic function  $\log_e F(z^{-1})$  can be expressed as

$$\begin{aligned} \log_e F(z^{-1}) &= \sum_p \log_e F_p(z^{-1}) \\ &= \sum_{n=1}^{\infty} f(n) z^{-n} \end{aligned} \quad (6.27)$$

where

$$f(n) = \sum_p f_p(n). \quad (6.28)$$

In contrast, a non-causal time sequence is obtained for the outside zeros. Suppose that  $G(z)$  can be factorized as

$$G(z) = \prod_{q=1}^{N-1} (1 - c_q^{-1} z). \quad (6.29)$$

By analogy with the above, factoring a complex conjugate pair of zeros, the logarithmic function can be written as

$$\begin{aligned} \log_e G_q(z) &= \log_e(1 - \frac{1}{c_q} z) + \log_e(1 - \frac{1}{c_q} z^*) \\ &= 2 \sum_{n=-1}^{-\infty} \frac{\Re[c_q^n]}{n} z^{-n} \end{aligned} \quad (6.30)$$

if  $|c_q^{-1} z| < 1$ ; the unit circle is also included because  $|c_q^{-1}| < 1$ . Consequently,

$$\log_e G(z) = -2 \sum_{n=1}^{\infty} \left[ \sum_p \frac{\Re[b_p^n]}{n} \right] z^n$$

$$= \sum_{n=1}^{\infty} f(n)z^n \quad (6.31)$$

is derived with  $c_q = (1/b_p)^*$ .

Hence the logarithm of the power spectral density can be represented as a  $z$ -transform:

$$\begin{aligned} \log_e F(z^{-1})G(z)|_{z=e^{j\Omega}} &= \sum_{n=1}^{\infty} f(n)(z^n + z^{-n}) \Big|_{z=e^{j\Omega}} \\ &= 2 \sum_{n=1}^{\infty} f(n) \cos n\Omega. \end{aligned} \quad (6.32)$$

Hence, the logarithmic power spectral function is an even function of  $\Omega$ ; its non-causal time sequence is also even, and has an infinite record length. This is because the logarithmic function has singularities owing to its zeros, and thus, the zeros outside the unit circle make the time sequence non-causal.

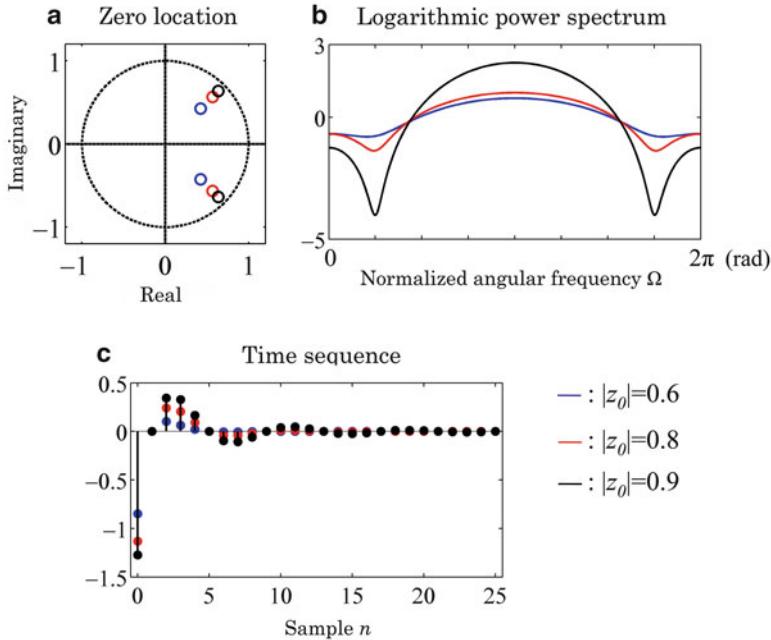
Figure 6.6 on the facing page illustrates samples of the logarithmic function and their time sequences. Panel (a) shows the locations of the zeros, whereas (b) depicts the logarithmic power-spectral density function with its time sequence illustrated in (c). The troughs in the power-spectral density function become shallower as the zeros move farther away from the unit circle, and consequently, the effective record length of the time sequence becomes shorter.

## 6.2 Phase and Accumulated Phase

### 6.2.1 Phase Characteristics for Zeros

The frequency response of the phase can be defined as well as that for the magnitude response. Consider the transfer function for a single-echo system such as given by Eq. 6.4 on page 96. The transfer function can be factorized using the zeros:

$$\begin{aligned} H(z^{-1}) &= 1 + az^{-m} \\ &= K \prod_{l=1}^m (1 - a_l z^{-1}) \end{aligned} \quad (6.33)$$



**Fig. 6.6** Example of the logarithm of the power spectral density and its time sequence; (a) locations of complex conjugate pair of zeros  $z_0$  and  $z_0^*$  where  $z_0 = |z_0|e^{i\pi/4}$ ; (b) logarithmic power spectral density function; (c) time sequence for the logarithm of the power spectral density

where

$$z_0 = |a|^{1/m} e^{i \frac{(2l+1)\pi}{m}} \quad (6.34)$$

and  $a > 0$ . The phase frequency response can be expressed as

$$\begin{aligned} \angle H(e^{-i\Omega}) &= \sum_{l=1}^m \Phi_l(e^{-i\Omega}) \\ &= \Im \left[ \sum_{l=1}^m \log_e(1 - a_l e^{-i\Omega}) \right] \end{aligned} \quad (6.35)$$

where

$$\Phi_l(e^{-i\Omega}) = \angle(1 - a_l e^{-i\Omega}). \quad (6.36)$$

### 6.2.2 Analytic Expression for Phase Frequency Response

The logarithmic function in Eq. 6.35 on the previous page can be expanded into power series [2]. Take a complex conjugate pair of zeros such as  $a_l$  and  $a_l^*$  where  $|a_l| < 1$ . The transfer function constructed by the pair of zeros is defined as

$$H_l(z^{-1}) = (1 - a_l z^{-1})(1 - a_l^* z^{-1}). \quad (6.37)$$

Note that the transfer function above has double poles at the origin, specifically the center of the unit circle, in addition to a complex conjugate pair of zeros. Hence, with the exception of poles and zeros, taking the logarithm of the transfer function yields

$$\log_e H_l(z^{-1}) = \log_e(1 - a_l z^{-1}) + \log_e(1 - a_l^* z^{-1}) \quad (6.38)$$

where  $|a_l z^{-1}| < 1$ . Recalling the power series expansion in Eq. 6.23 on page 102,

$$\log_e H_l(z^{-1}) = -2 \sum_{n=1}^{\infty} \frac{\Re[a_l^n]}{n} z^{-n}. \quad (6.39)$$

Therefore the phase frequency characteristics is expressed as

$$\begin{aligned} \Phi_l(e^{-i\Omega}) &= \angle H_l(e^{-i\Omega}) \\ &= \Im[\log_e H_l(e^{-i\Omega})] \\ &= 2 \sum_{n=1}^{\infty} \frac{\Re[a_l^n]}{n} \sin(n\Omega). \end{aligned} \quad (6.40)$$

The time sequence from which the phase response can be derived is a real causal sequence, and the phase response curve is an odd function of the normalized angular frequency  $\Omega$ .

Suppose that a pair of complex conjugate zeros  $|a_l|e^{\pm i\Omega_l}$  are given. The equation above indicates that the phase response starts at 0 when  $\Omega = 0$  and returns to 0, when  $\Omega$  reaches  $\pi$ . Now looking at the group delay for the phase response and assuming that the phase curve is a differentiable function, the group delay can be estimated as a limit when  $|a_l|$  approaches unity from inside the unit circle:

$$\begin{aligned} g_{l_d}(\Omega) &= -\frac{\partial \Phi_l(e^{-i\Omega})}{\partial \Omega} \\ &= -2 \sum_{n=1}^{\infty} \Re[a_l^n] \cos n\Omega. \end{aligned} \quad (6.41)$$

Thus, the group delay at  $\Omega = 0$  is given by

$$g_{l_d}(\Omega)|_{\Omega=0} = -2 \sum_{n=1}^{\infty} \Re[a_l^n]. \quad (6.42)$$

Consequently, taking the limit  $\epsilon \rightarrow 0$  subject to condition  $|a_l| \cong 1 - \epsilon$ , the group delay becomes

$$\begin{aligned} g_{l_d}(\Omega)|_{\Omega=0} &= -2 \sum_{n=1}^{\infty} \cos n\Omega_l \\ &= -2 \sum_{n=1}^{\infty} \Re[e^{in\Omega_l}] \\ &= 1. \end{aligned} \quad (6.43)$$

Specifically, the first derivative of the phase response becomes  $-1$  at  $\Omega = 0$  in the limit when the zeros approach the unit circle from inside the unit circle; the same occurs for  $\Omega = \pi$ .

Estimating the phase derivative gives an estimate of the phase jump at  $\Omega = \pm\Omega_l$ , which determine the angles at the zeros, in the limit when the zeros approach the unit circle. In accordance with a phase derivative of  $-1$  at  $\Omega = 0$  and  $\Omega = \pi$ , the phase approaches  $\Omega_l$  when the angular frequency  $\Omega$  tends to  $\Omega_l$  from the left-hand side; in contrast, the phase is  $\pi - \Omega_l$  from the right-hand side. These results indicate that the phase response becomes discontinuous in the limit. The phase jump, which represents the discontinuity of the phase, can be obtained from

$$\Delta\Phi_l(\Omega_l) = (\pi - \Omega_l) - (-\Omega_l) = \pi. \quad (6.44)$$

A phase jump of  $+\pi$  can be expected for an interior zero in the limit when the zero approaches the unit circle.

The example above describes the phase response for a complex conjugate pair of zeros inside the unit circle. If the zeros are located outside the unit circle, then take a pair of complex conjugate zeros at  $b_l = |b_l|(e^{\pm i\Omega_l})$  where  $b_l = (a_l^*)^{-1}$ . The transfer function constructed by this pair of zeros is defined as

$$\hat{H}_l(z) = |b_l|^2 z^{-2} (1 - (b_l)^{-1}z)(1 - (b_l^*)^{-1}z). \quad (6.45)$$

Note here this transfer function has double poles at the center of the unit circle in addition to the complex conjugate pair of zeros. Thus, with the exception of the poles and zeros, taking the logarithm of the transfer function gives

$$\log_e \hat{H}_l(z) = \log_e(1 - (b_l)^{-1}z) + \log_e(1 - (b_l^*)^{-1}z) + \log_e z^{-2} \quad (6.46)$$

to within a constant term. Recalling the power series expansion in Eq. 6.23 on page 102, the equation above is rewritten as

$$\log_e \hat{H}_l(z) = -2 \sum_{n=1}^{\infty} \frac{\Re[(a_l)^n]}{n} z^n + \log_e z^{-2}. \quad (6.47)$$

In contrast to Eq. 6.39 on page 106, this result indicates the time series from which the phase response can be derived is a real but non-causal sequence. Therefore, the phase frequency characteristics is given by substituting  $z = e^{j\Omega}$  for  $z$ :

$$\begin{aligned} \hat{\phi}_l(e^{j\Omega}) &= \Im \left[ \log_e \hat{H}_l(e^{j\Omega}) \right] \\ &= -2 \sum_{n=1}^{\infty} \frac{\Re[(a_l)^n]}{n} \sin(n\Omega) - 2\Omega. \end{aligned} \quad (6.48)$$

There is a significant difference between Eqs. 6.40 on page 106 and 6.48, i.e., between zeros interior and exterior to the unit circle. There are double poles at the origin for both cases; however, contributions of the poles to the phase cancel for zeros inside the circle, whereas they do not for zeros outside. Consequently, the phase returns to zero at  $\Omega = \pi$  and  $\Omega = 2\pi$  for complex conjugate pairs inside the unit circle, whereas the phase becomes  $-2\pi$  and  $-4\pi$  for the two outside zeros of the pair.

In addition, looking at the group delay, the phase jump expected by the outside zero is just opposite to that for the inside zero. The group delay can be estimated as the limit when  $|b_l|$  approaches unity from outside the unit circle:

$$\hat{g}_{l_d}(\Omega) = 2 \sum_{n=1}^{\infty} \Re[(a_l)^n] \cos n\Omega + 2. \quad (6.49)$$

The group delay at  $\Omega = 0$  is given by

$$\hat{g}_{l_d}(\Omega)|_{\Omega=0} = 2 \sum_{n=1}^{\infty} \Re[(a_l)^n]. \quad (6.50)$$

Consequently, taking the limit  $\epsilon \rightarrow 0$  subject to  $|b_l| \cong 1 + \epsilon$ , the group delay becomes

$$\begin{aligned} \hat{g}_{l_d}(\Omega)|_{\Omega=0} &= 2 \sum_{n=1}^{\infty} \cos n\Omega_l + 2 \\ &= 1. \end{aligned} \quad (6.51)$$

Hence, similar to inside zeros, the group delay becomes 1 at  $\Omega = 0$  in the limit, and is also the same at  $\Omega = \pi$ .

However, the difference in the phase response at  $\Omega = \pi$  yields the opposite phase jump at  $\Omega_l$ . For inside zeros, according to which the phase derivatives are  $-1$  at  $\Omega = 0$  and  $\Omega = \pi$ , the phase approaches  $-\Omega_l$  when the angular frequency  $\Omega$  approaches  $\Omega_l$  from the left-hand side. This is also true for the outside zeros. In contrast, the phase tends to  $-\pi - \Omega_l$  from the right-hand side. Hence, the phase jump is estimated to be

$$\hat{\Delta}\Phi_l(\Omega_l) = (-\pi - \Omega_l) - (-\Omega_l) = -\pi. \quad (6.52)$$

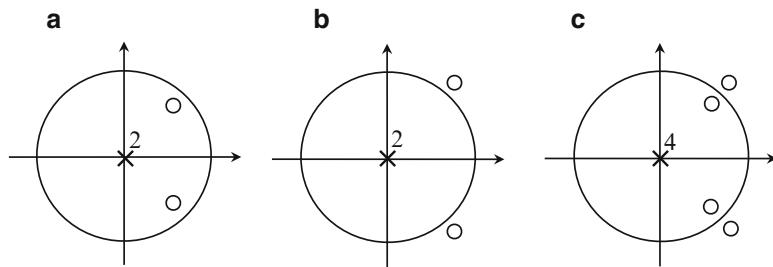
Therefore a phase jump of  $-\pi$  can be expected for an outside zero in the limit as the zeros approach the unit circle.

The differences in the phase responses indicate that the local behavior of the phase response might cancel if the inside and outside pairs are superposed. Recalling Eqs. 6.40 on page 106 and 6.48 on the preceding page, the phase response for both zero pairs is expressed as

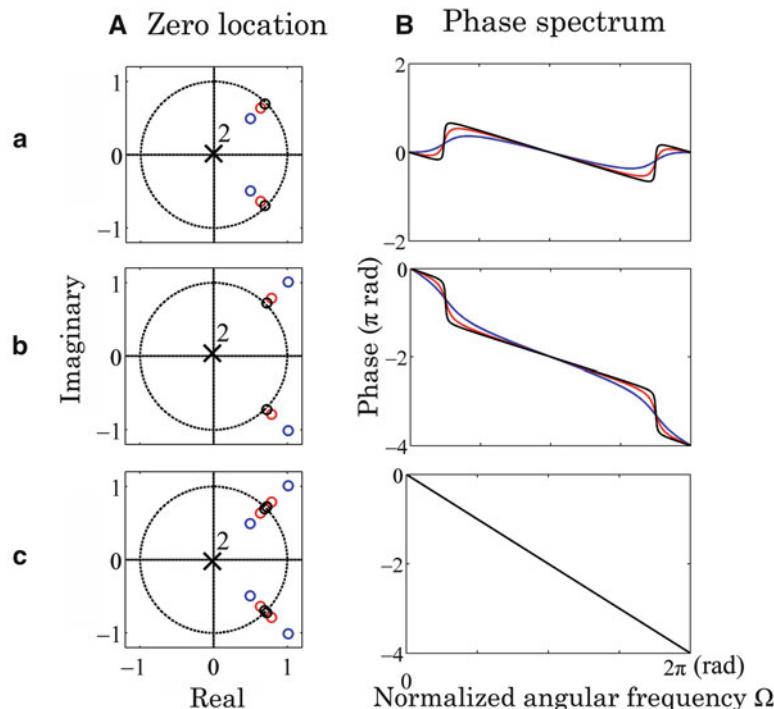
$$\begin{aligned} \Phi_{LP}(e^{-i\Omega}) &= \Phi_l(e^{-i\Omega}) + \hat{\Phi}_l(e^{i\Omega}) \\ &= 2 \sum_{n=1}^{\infty} \frac{\Re[a_l^n]}{n} \sin(n\Omega) - 2 \sum_{n=1}^{\infty} \frac{\Re[(a_l^*)^n]}{n} \sin(n\Omega) - 2\Omega \\ &= -2\Omega. \end{aligned} \quad (6.53)$$

This result shows that a linear phase system can be made if the response is described by a combination of reciprocal complex-conjugate pairs of zeros with respect to the unit circle.

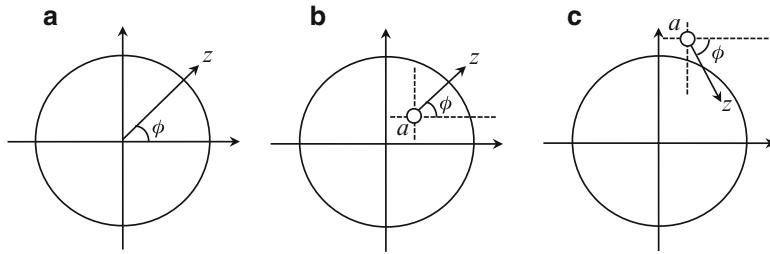
Figure 6.7 on the following page illustrates examples of complex-conjugate pairs of zeros (a) inside, (b) outside, and (c) on both sides of the unit circle. In particular the pairs are reciprocally located with respect to the unit circle in (c). Double poles are located at the origin in (a) and (b), whereas quadratic poles are set at the origin in (c). Figure 6.8 on the following page presents the phase responses corresponding to the pole/zero locations (a), (b), and (c) in Fig. 6.7 on the following page, respectively. The curves (blue, red, and black lines corresponding to the locations of zeros) are drawn using Eqs. 6.40 on page 106, 6.48 on the preceding page, and 6.53, respectively. The zero phase in case (a), the negative phase in (b), and the linear (negative) phase which is proportional to the frequency in (c) are all confirmed. In addition, the phase jump around the zeros rises steeply as the zeros approach the unit circle.



**Fig. 6.7** Examples of poles and reciprocal locations of zeros; (a) complex conjugate pair of zeros inside the unit circle, (b) complex conjugate pair of outside zeros, (c) reciprocal pairs of zeros with respect to unit circle



**Fig. 6.8** Phase responses corresponding to the pole/zero locations (a), (b), and (c) in Fig. 6.7 where blue, red, and black lines, corresponding to the locations of zeros of Eqs. 6.40 on page 106, 6.48 on page 108, and 6.53 on the previous page, respectively, are drawn. (A) Zero location. (B) Phase spectrum



**Fig. 6.9** Geometric interpretation of phase angle of a complex function; (a)  $f(z) = z$ , (b)  $f(z) = z - a$  for  $|a| < 1$ , and (c)  $|a| > 1$

### 6.2.3 Geometric Interpretation of Phase Response

Analytic expressions for the phase responses are derived using the power series expansion formula. However, the limit as the zeros approach the unit circle can be intuitively understood following the geometry of complex functions on the complex frequency plane. Figure 6.9 shows the phase angles of the complex functions  $f(z) = z$  (panel (a)) and  $f(z) = z - a$  for  $|a| < 1$  (panel (b)) and  $|a| > 1$  (panel (c)). In the complex frequency plane, the zero of  $f(z) = z$  is at the origin whereas for  $f(z) = z - a$  the zero is at  $z = a$ . Hence, the differences in the phase angles for the functions arise from the location of the zero.

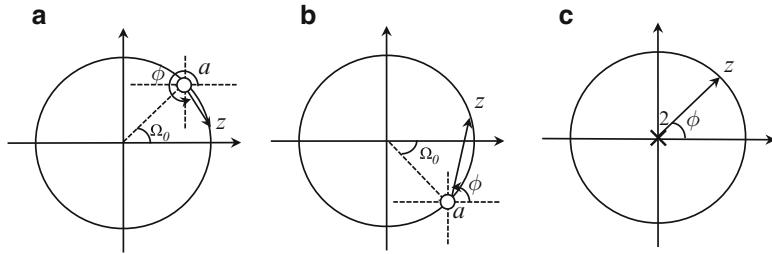
Figure 6.10 on the following page illustrates examples of the phase angle for conjugate pairs of zeros (panels (a) and (b)), and for double poles at the origin in the  $z$ -plane (panel (c)). The phase angle  $\Phi(\Omega)$  changes when the angular frequency  $\Omega$  moves around the origin on the unit circle from  $\Omega = 0$  to  $\Omega = 2\pi$ . In panel (a), the angle  $\Phi_a(\Omega)$  can be expressed algebraically as

$$\Phi_a(\Omega) = \begin{cases} \frac{1}{2}(\Omega + |\Omega_0| - \pi), & 0 \leq \Omega < |\Omega_0|, \\ \frac{1}{2}(\Omega + |\Omega_0| + \pi), & |\Omega_0| < \Omega \leq 2\pi \end{cases} \quad (6.54)$$

where  $|\Omega_0|$  denotes the phase angle of the zero. Similarly, in panel (b), the phase angle which corresponds to the complex conjugate of that for the zero in (a), is written as

$$\Phi_b(\Omega) = \Phi_{a^*}(\Omega) = \begin{cases} \frac{1}{2}(\Omega + 2\pi - |\Omega_0| - \pi), & 0 \leq \Omega < 2\pi - |\Omega_0|, \\ \frac{1}{2}(\Omega + 2\pi - |\Omega_0| + \pi). & 2\pi - |\Omega_0| < \Omega \leq 2\pi \end{cases} \quad (6.55)$$

where  $2\pi - |\Omega_0|$  gives the angle of the zero location. For panel (c) corresponding to double poles located at the origin, the phase angle  $\Phi_c(\Omega)$  changes  $-4\pi$  as the angular frequency moves anti-clockwise around the unit circle. Note here that the



**Fig. 6.10** Schematic of the phase angles  $\Phi(\Omega)$  for zeros when (a)  $\Omega_0 = \pi/4$ , (b)  $\Omega_0 = -\pi/4$ , and (c) double poles occur at the origin

phase change because a pole is just opposite to that for a single zero located at the same position instead of the pole. Mathematically,

$$\angle[(z-a)|_{z=e^{i\Omega}}] = -\angle[(z-a)^{-1}|_{z=e^{i\Omega}}]. \quad (6.56)$$

Consequently, the total phase response from the poles and zeros in Fig. 6.10 is given by

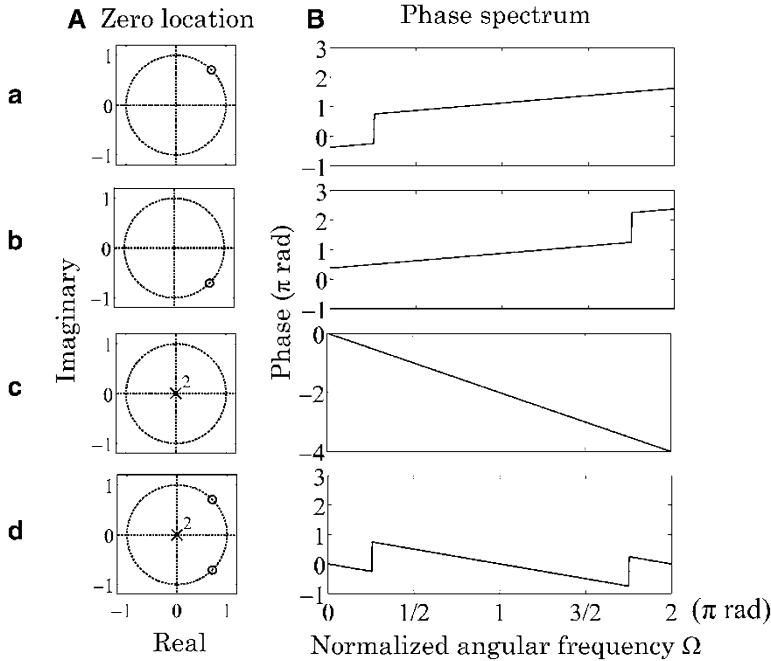
$$\begin{aligned} \Phi_{abc}(\Omega) &= \Phi_a(\Omega) + \Phi_b(\Omega) + \Phi_c(\Omega) \\ &= \begin{cases} -\Omega, & 0 \leq \Omega < |\Omega_0|, \\ -\Omega + \pi, & |\Omega_0| < \Omega < 2\pi - |\Omega_0|, \\ -\Omega + 2\pi, & 2\pi - |\Omega_0| < \Omega \leq 2\pi. \end{cases} \end{aligned} \quad (6.57)$$

Figure 6.11 on the facing page shows the phase response given by Eq. 6.57. The phase jump of  $\pi$  owing to the zeros, which can be intuitively understood as shown in Figs. 6.12 on the facing page and 6.13 on page 114, is verified.

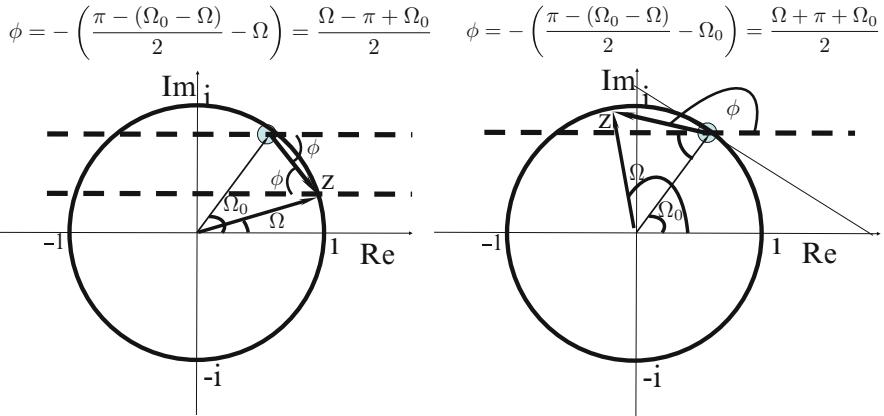
This geometric phase expression corresponds to that for the limit case described in the previous sub-section. Thus the time series used for the analytic representation of the phase response can be interpreted as Fourier coefficients for the Fourier series expansion of the phase response. Therefore the analytic phase expression converges to the average of the left- and right-side limits at the discontinuity corresponding to the zero location in the limit [3].

### 6.2.4 Poles, Zeros, and Accumulated Phase

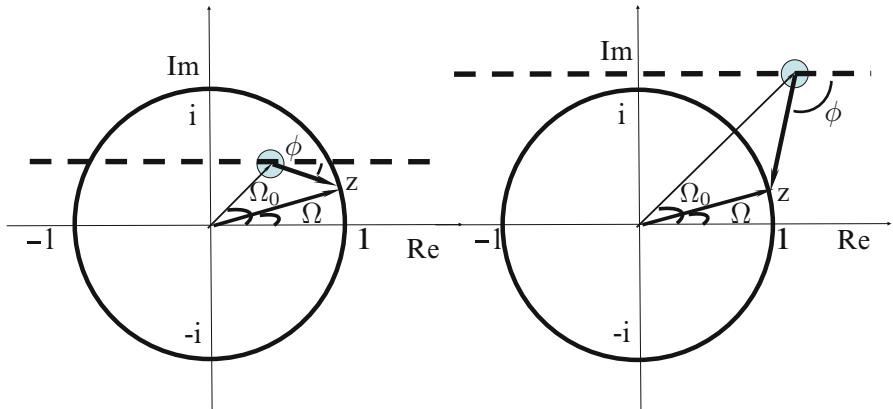
Like the transfer function, frequency characteristics can also be identified and represented using poles and zeros. The phase difference at  $\Omega$  from  $\Omega = 0$  is called the accumulated phase of the system of interest. Specifically, the phase angle  $\Phi(e^{-i\Omega})$  gives the accumulated phase when  $\Phi(e^{-i\Omega})|_{\Omega=0}$  is set to 0.



**Fig. 6.11** Phase responses occurring for poles and zeros (a), (b), and (c) corresponding to Fig. 6.10 on the preceding page, and their total responses (d); phase responses are given by Eqs. 6.54 on page 111 for (a), 6.55 on page 111 for (b), and 6.57 on the preceding page for (d). (A) Zero location. (B) Phase spectrum



**Fig. 6.12** Schematic of the phase jump owing to zero



**Fig. 6.13** Schematic of the phase change owing to inside or outside zero

The accumulated phase for a single pole inside the unit circle is  $-2\pi$  when  $\Omega$  moves from 0 to  $2\pi$ . In contrast, the phase is  $2\pi$  for a single zero inside the unit circle, whereas the phase for a single zero outside the unit circle does not accumulate but returns to 0. Suppose that there are  $N_p$  poles inside the unit circle, and  $N_z^-$  and  $N_z^+$  zeros inside and outside the unit circle, respectively where  $N_p = N_z^- + N_z^+$ . The accumulated phase is

$$\Phi(e^{-i\Omega})|_{\Omega=2\pi} = -2\pi(N_p - N_z^-) \quad (6.58)$$

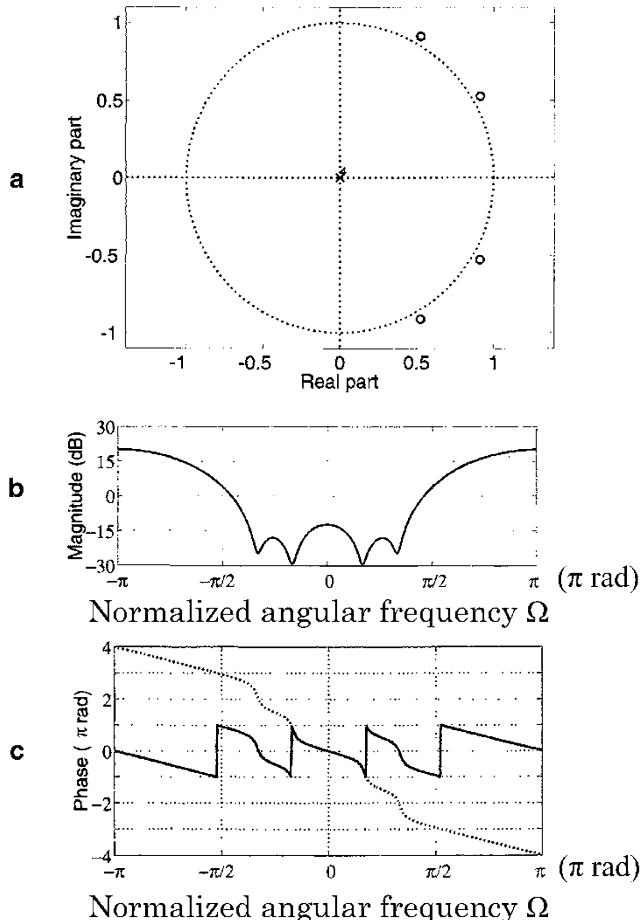
which becomes 0 when  $N_z^- = N_p$ , that is,  $N_z^+ = 0$ . In contrast, it yields  $-2\pi N_p$  when  $N_z^- = 0$ , that is,  $N_z^+ = N_p$ .

The phase response represents the continuous accumulated-phase function. However, frequently the principal value of the phase, which is wrapped into the interval between 0 and  $2\pi$  as shown in Fig. 6.14 on the facing page, is taken as indicating the phase characteristics. Such a wrapped discontinuous phase cannot be used to represent linear systems. Therefore the accumulated continuous phase is called the unwrapped phase in contrast to the wrapped phase. The unwrapped phase can be obtained by integrating the group delay:

$$\Phi(e^{-i\Omega}) = - \int_0^\Omega \tau(\Omega) d\Omega \quad (6.59)$$

where  $\tau(\Omega)$  denotes the group delay.

The accumulated phase can be generalized using the residue theorem for complex functions [4]. Suppose that  $f(z)$  is analytic, except at the poles within a closed curve  $c$ , and there are no poles and zeroes on  $c$ ; then



**Fig. 6.14** Example of magnitude, wrapped, and unwrapped phase frequency responses: (a) pole and zero locations, (b) magnitude response, (c) wrapped (solid) and unwrapped (dotted) phase responses from Fig. 4.1.4 in [2]

$$\Im \left[ \int_c \frac{f'(z)}{f(z)} dz \right] = -2\pi(N_p - N_z^-). \quad (6.60)$$

The formula is better understood by rewriting the integrand as the first derivative of the logarithm of  $f(z)$ .

## 6.3 Minimum Phase and Cepstral Sequences

### 6.3.1 *Magnitude and Phase Characteristics for Minimum-Phase Transfer Function*

Phase characteristics are important in categorizing linear systems. In particular, the minimum-phase property is fundamental. A linear causal system is called a minimum phase, when its transfer function has no zero outside the unit circle. Following the definition of minimum phase, the total accumulated phase of a minimum-phase system must be 0 with  $N_p = N_z^-$ . Recall the factorization of the transfer function written by Eq. 6.1 on page 95; in the factorized form,  $|a_l| < 1$  for causal and stable minimum-phase systems.

The real (imaginary) part of the Fourier transform of the impulse response sequence, can be derived from the imaginary (real) part, subject to the impulse response being a real causal sequence. However, there is no definite relationship between the magnitude and phase parts for the Fourier transform in general. Interestingly, the same type of relationship holds between the magnitude and phase components for minimum phase transfer functions.

Recall Sects. 6.1.5 on page 100 and 6.2.1 on page 104 where logarithmic expressions were introduced into the transfer functions. The logarithmic power spectral response and phase characteristics can be analytically expressed using the time sequences in which the logarithms are represented in Fourier series expansions. This kind of time sequence is called a cepstral sequence [2,5,6]. Consider a transfer function  $H(z^{-1})$  for a real, causal, and stable sequence  $h(n)$ . Its logarithm

$$C(z^{-1}) = \log_e H(z^{-1}) \quad (6.61)$$

has singularities inherited from the poles and zeros of the transfer function  $H(z^{-1})$ . That is,  $C(z^{-1})$  has poles inside the unit circle in the  $z$ -plane corresponding to the poles of the transfer function; in addition, it has poles corresponding to the zeros of  $H(z^{-1})$  both inside and outside the unit circle. As  $h(n)$  is a causal and stable sequence, its zeros could be distributed inside and/or outside the unit circle.

Here, suppose that  $H(z^{-1})$  has no zeros outside the unit circle; that is, the linear system of interest is minimum phase. Taking the inverse  $z$ -transform of  $C(z^{-1})$ , the time sequence  $c(n)$ , which is called a complex cepstral sequence (or cepstrum), is derived as a real, stable, and causal sequence. Specifically, following Eq. 6.23 on page 102,  $C(z^{-1})$  can be expressed as

$$C(z^{-1}) = \sum_{n=1}^{\infty} c(n)z^{-n} \quad (6.62)$$

using the real causal time sequence  $c(n)$ . Consequently,  $C(e^{-i\Omega})$ , the Fourier transform of  $c(n)$ , is

$$\begin{aligned} C(e^{-i\Omega}) &= \log_e H(e^{-i\Omega}) \\ &= \log_e |H(e^{-i\Omega})| + i\Phi(e^{-i\Omega}) \end{aligned} \quad (6.63)$$

where

$$H(e^{-i\Omega}) = |H(e^{-i\Omega})| e^{i\Phi(e^{-i\Omega})} \quad (6.64)$$

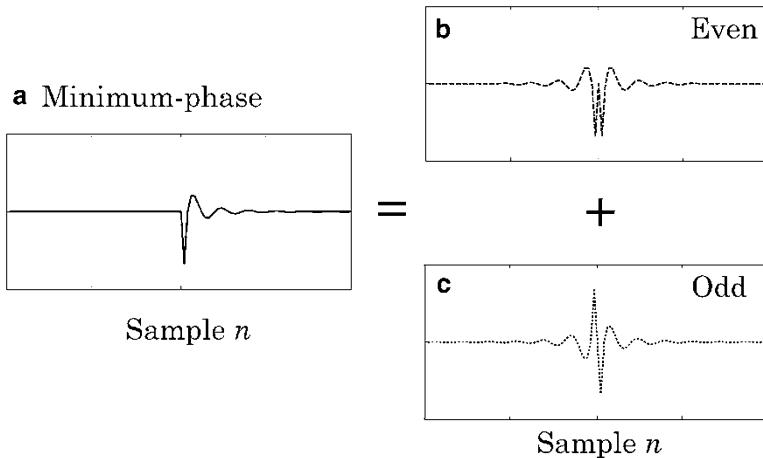
and has the same type of relationship between the logarithmic magnitude and phase as that for the real and imaginary parts of  $H(e^{-i\Omega})$ . This result indicates that the relationship between the real and imaginary parts of the Fourier transform for a real, causal, and stable sequence is not expected for the magnitude and phase; however, if the real, causal, and stable time sequence is a minimum phase, then the logarithmic magnitude (phase) can be derived from the phase (logarithmic magnitude) instead. That is, the real causal cepstral sequence  $c(n)$  decomposes into even and odd sequences.

### 6.3.2 Decomposition of Minimum-Phase Cepstrum into Magnitude and Phase Terms

Given that a real causal sequence can be decomposed into even and odd ones, consider then a real causal cepstral sequence  $c_{min}(n)$  that is derived from a minimum-phase and real impulse response  $h_{min}(n)$ . Customarily called a minimum-phase cepstrum, it is a real causal sequence that can be divided into even and odd sequences (see schematic in Fig. 6.15 on the following page). Here, recalling once more the Fourier transform of a real sequence, the even sequence (dashed line) yields the real part of its Fourier transform. The real part can be written as

$$\begin{aligned} \Re [C_{min}(e^{-i\Omega})] &= \log_e |H(e^{-i\Omega})| \\ &= \frac{1}{2} \log_e |H(e^{-i\Omega})|^2 \end{aligned} \quad (6.65)$$

where  $C_{min}(e^{-i\Omega})$  denotes the Fourier transform of the minimum-phase cepstrum, and thus the real part indicates the logarithmic magnitude or power spectral density that is described in Sect. 6.1 on page 95. Therefore the even time sequence (dashed line) is customarily called the magnitude cepstrum.



**Fig. 6.15** Schematic illustrating the decomposition of minimum-phase cepstrum into magnitude and phase cepstra where the causal sequence (solid line) represents the minimum-phase (**a**) cepstrum and the even (**b**)/odd (**c**) sequence (dashed/dotted) line represents the magnitude/phase cepstrum from Fig. 14.11 in [6]

Similarly, the imaginary part of  $C_{min}(e^{-i\Omega})$

$$\Im[C_{min}(e^{-i\Omega})] = \angle[H(e^{-i\Omega})] = \Phi(e^{-i\Omega}) \quad (6.66)$$

obtained from the odd sequence (dotted line), is termed the minimum-phase phase cepstrum. Consequently, the relationship between the logarithmic magnitude and phase spectral functions can be established for a real and minimum-phase sequence by decomposing the minimum-phase cepstrum:

$$\begin{aligned} c_{min}(n) &= c_{min-even} + c_{min-odd} \\ &= c_{min-mag} + c_{min-phase} \end{aligned} \quad (6.67)$$

where

$$\begin{aligned} c_{min-even} &= c_{min-mag} \\ c_{min-odd} &= c_{min-phase}. \end{aligned} \quad (6.68)$$

Here,  $c_{min-mag}$  is the magnitude cepstrum drawn as the even sequence (dashed line), whereas  $c_{min-phase}$  is the minimum-phase phase cepstrum drawn as the odd sequence (dotted line).

## 6.4 Decomposition of Transfer Function into Minimum-Phase and All-Pass Transfer Functions

### 6.4.1 Decomposition of Transfer Function by Pole-Zero Diagram

The logarithmic expression of the transfer function has singularities inside and/or outside the unit circle in consequence of the zeros for the transfer function, as described in Sect. 6.3.1 on page 116. This fact indicates that the transfer function can be decomposed into a minimum-phase component and its complement. Suppose that a transfer function can be factorized as

$$H(z^{-1}) = K(1 - az^{-1})(1 - a^*z^{-1})(1 - bz^{-1})(1 - b^*z^{-1}) \quad (6.69)$$

where  $|a| < 1$ ,  $|b| > 1$ , and  $K$  is a constant. Recall Fig. 6.3 on page 99, which displays the locations of reciprocal complex-conjugate zeros with respect to the unit circle. There are no differences in the power spectral density functions except for the constant between the zeros making up the reciprocal complex-conjugate pair. Take the single zero located at  $z = b$  outside the unit disc. If the zero moves into its reciprocal location given by  $b^{-1}$ , the power spectral density function corresponding to the zero becomes

$$\begin{aligned} |H_{(b^*)^{-1}}|^2|_{z=e^{i\Omega}} &= |(1 - (b^*)^{-1}z^{-1})|^2|_{z=e^{i\Omega}} \\ &= (1 - (b^*)^{-1}z^{-1})(1 - b^{-1}z)|_{z=e^{i\Omega}} \\ &= |b^{-1}|^2 (1 - b^*z)(1 - bz^{-1})|_{z=e^{i\Omega}} \\ &= |b^{-1}|^2 |H_b(z^{-1})|^2|_{z=e^{i\Omega}}. \end{aligned} \quad (6.70)$$

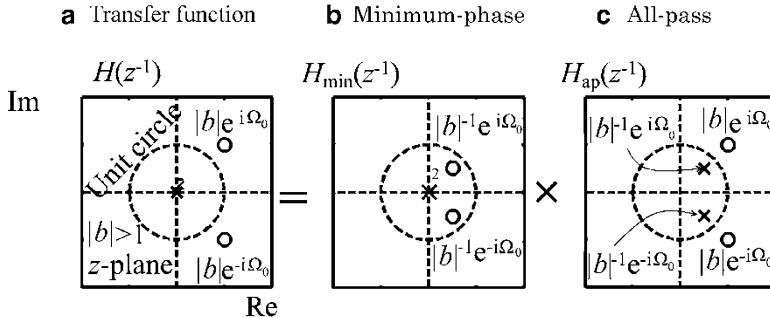
Consequently, the decomposition of the pole-zero plot holds for the transfer function with respect to the unit circle as shown in Fig. 6.16 on the following page.

In Fig. 6.16 on the following page, panel (a) shows the pole-zero plot for the original transfer function; similarly, panel (b) presents the minimum-phase component that contains all the zeros inside the unit circle. Panel (c) gives the all-pass component left after extracting the minimum-phase part from the original transfer function. That is, the figure indicates that the transfer function can be decomposed into

$$H(z^{-1})|_{z=e^{i\Omega}} = H_{min}(z^{-1}) \cdot H_{ap}(z^{-1})|_{z=e^{i\Omega}} \quad (6.71)$$

or

$$h(n) = h_{min} * h_{ap}(n). \quad (6.72)$$



**Fig. 6.16** Decomposition of pole-zero plot into minimum-phase and all-pass component for transfer function where left panel (a) shows for the original transfer function, central one (b) illustrates for minimum-phase, and plot (c) gives all-pass components, respectively from Fig. 13.42 in [6].

Here  $H_{ap}(z^{-1})$  and  $h_{ap}(n)$  denote the all-pass components for the original transfer function  $H(z^{-1})$  and impulse response  $h(n)$ .

Take the single zero component for  $z = b$  once more. The corresponding all-pass component for the transfer function  $H_b(z^{-1})$  can be written as

$$H_{b-ap}(z^{-1}) = \frac{1 - bz^{-1}}{1 - (b^*)^{-1}z^{-1}} \quad (6.73)$$

which on the unit circle can be rewritten as

$$\begin{aligned} H_{b-ap}(e^{-i\Omega}) &= \frac{1 - be^{-i\Omega}}{1 - (b^*)^{-1}e^{-i\Omega}} \\ &= -be^{-i\Omega} \frac{1 - b^{-1}e^{i\Omega}}{1 - (b^*)^{-1}e^{-i\Omega}}. \end{aligned} \quad (6.74)$$

Therefore, taking the magnitude,

$$|H_{b-ap}(e^{-i\Omega})| = |b|. \quad (6.75)$$

The result above indicates that the magnitude response is a constant independent of frequency, thus clarifying the term “all-pass” given to this component. It is also consistent with the minimum-phase component maintaining the magnitude of the frequency response for the original transfer function.

The flat-magnitude frequency response of the all-pass component is because of the reciprocal complex-conjugate locations of the poles (within the unit disc) and outside zeros with respect to the unit circle. In contrast to the reciprocal pair of zeros inside and outside the unit disc, the effects to the magnitude frequency response from inside poles disappear, being cancelled by the outside zeros. Consequently,

only the phase carries frequency-dependent information. Following these results, the frequency characteristics for the transfer function can be decomposed into

$$\begin{aligned} H(e^{-i\Omega}) &= |H(e^{-i\Omega})| e^{i\Phi(e^{-i\Omega})} \\ &= H_{min}(e^{-i\Omega}) \cdot H_{ap}(e^{-i\Omega}) \\ &= |H_{min}(e^{-i\Omega})| e^{i(\Phi_{min}(e^{-i\Omega}) + \Phi_{ap}(e^{-i\Omega}))} \end{aligned} \quad (6.76)$$

where

$$\Phi(e^{-i\Omega}) = \Phi_{min}(e^{-i\Omega}) + \Phi_{ap}(e^{-i\Omega}) \quad (6.77)$$

and

$$|H_{ap}(e^{-i\Omega})| = 1. \quad (6.78)$$

Note that the magnitude and phase response can be obtained from each other for the minimum-phase component, in contrast to that for the all-pass component which are independent of each other.

The all-pass system is historically important in audio engineering. Artificial reverberators are basically designed based on all-pass systems [7], because an all-pass system does not change the magnitude (or power) of the spectral characteristics of sound that is fed into the system. This can be understood by recalling that the auto-correlation sequence for an all-pass system is represented by a  $\delta$ -function. That is, the power spectral density function of the all-pass system is a constant independent of the frequency. Therefore the auto-correlation sequence of the output sequence through the all-pass system can be formulated as

$$\begin{aligned} r_{ap-out}(n) &= r_{xx} * r_{ap}(n) \\ &= r_{xx} * K\delta(n) \\ &= Kr_{xx}(n) \end{aligned} \quad (6.79)$$

where  $r_{xx}(n)$  and  $r_{ap}(n)$  denote the auto-correlation sequences for the input sound and the auto-correlation sequence of the impulse response of the all-pass system. Specifically, the auto-correlation sequence is retained even after passing through the all-pass system.

In contrast to the constant magnitude spectral characteristics, the output sound changes quite perceptively in passing through the all-pass system. This change in sound arises from changes in the fine structure of the phase spectral characteristics [7] brought about by the all-pass system. In particular, the group delay causes significant changes in the envelopes of the input sound, and thus the modified envelopes are perceived as reverberation. This reverberation effect can be intuitively

understood by observing over a long duration the impulse response for the all-pass system that is produced by the poles. That is, phase effects determined by the poles cannot be cancelled by the outside zeros.

This perceptual or reverberation effect by the phase spectrum explains why speech intelligibility is lost, even if the magnitude spectral property of speech can be preserved over long frame lengths. Conversely, even if the magnitude spectral information is lost, intelligible speech is preserved, as long as the phase information of the intelligible speech is maintained over these long frame lengths. The phase spectral effects on perception of sound might be significant in the time domain rather than the frequency plane.

#### 6.4.2 Cepstral Decomposition of Transfer Function into Minimum-Phase and All-Pass Components

The decomposition of transfer functions are normally performed according to the decomposition of cepstrum instead of using the pole-zero plots. The real causal cepstrum can be divided into the magnitude and phase cepstral sequences as described in Sect. 6.3.2 on page 117. However, a cepstral sequence is not always a causal sequence because of outside zeros of the transfer function.

Cepstral sequences for transfer functions are expressed in general as

$$\begin{aligned} c(n) &= c_{mag}(n) + c_{phase}(n) \\ &= c_{min}(n) + c_{ap}(n) \end{aligned} \quad (6.80)$$

following Eq. 6.76 on the previous page, where  $c_{min}(n)$  and  $c_{ap}(n)$  denote the cepstral sequences for the minimum-phase and all-pass components, respectively. In addition, the minimum-phase part can be separated into the magnitude and phase cepstral sequences, whereas the all-pass component is composed of the corresponding phase cepstrum only. Therefore the cepstrum given by the equation above can be rewritten as

$$\begin{aligned} c(n) &= c_{mag}(n) + c_{phase}(n) \\ &= c_{min}(n) + c_{ap}(n) \\ &= c_{mag}(n) + c_{min-phase}(n) + c_{ap-phase}(n) \\ &= c^-(n) + c^+(n), \end{aligned} \quad (6.81)$$

where  $c^\pm(n)$  denote the non-causal and causal parts of the cepstrum, respectively.

Recalling the minimum-phase cepstrum is a causal sequence, the non-causal part of the cepstrum must represent the non-causal part of the all-pass phase cepstrum.

Decomposing the odd sequence for the all-pass (phase) cepstrum into non-causal and causal parts  $c_{ap}^{\pm}(n)$ , the minimum-phase cepstrum can be extracted for  $n > 0$  as

$$c_{min}(n) = c(n) - [-c_{ap}^+(n)] \quad (6.82)$$

where  $c_{ap}^-(n) = -c_{ap}^+(-n)$  for  $n > 0$ . This follows from the fact that the non-causal part of the cepstrum  $c(n)$  with  $n < 0$  is equal to the non-causal part of the all-pass phase cepstrum. Consequently, the minimum-phase and all-pass parts of the transfer function can be extracted following the decomposition scheme of the cepstrum without needing to use the pole-zero plots.

### 6.4.3 Samples of Cepstral Sequences and Decomposition

In Sects. 6.1.5 on page 100 and 6.2.1 on page 104, analytic expressions of the magnitude and phase responses are derived using power series of their logarithms. Cepstral sequences are also obtained according to the power series expansions [2]. Consider once again the transfer function

$$H_a(z^{-1}) = (1 - az^{-1})(1 - (a^*)^{-1}z^{-1}) \quad (6.83)$$

and its logarithmic function

$$\log_e H_a(z^{-1}) = \log_e(1 - az^{-1}) + \log_e(1 - (a^*)^{-1}z^{-1}). \quad (6.84)$$

Recall the power series expansion expressed in Eqs. 6.24 on page 102, 6.30 on page 103, and 6.47 on page 108. If  $|az^{-1}| < 1$ , following Eq. 6.24 on page 102, the power series expansion of the logarithm above can be written as

$$\log_e H_a(z^{-1}) = -2 \sum_{n=1}^{\infty} \frac{\Re[a^n]}{n} z^{-n}. \quad (6.85)$$

In contrast, if  $|az^{-1}| > 1$ ,

$$\log_e H_a(z^{-1}) = \log_e |a|^2 + \log_e z^{-2} + 2 \sum_{n=-1}^{-\infty} \frac{\Re[a^n]}{n} z^{-n} \quad (6.86)$$

is obtained from Eqs. 6.30 on page 103 and 6.47 on page 108.

The cepstral sequence or cepstrum, which can be defined by the inverse Fourier transform of the logarithm of the spectral function, is derived from the inverse  $z$ -transform. Specifically, if  $|a| < 1$ , the cepstral sequence becomes

$$c^-(n) = \begin{cases} -2\frac{\Re[a^n]}{n}, & n > 0 \\ 0, & n < 0. \end{cases} \quad (6.87)$$

In contrast, if  $|a| > 1$ , taking the inverse Fourier transform by substituting  $z = e^{i\Omega}$  for  $\log_e z^{-2}$  and  $\log_e |a|^{-2}$ , then

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log_e(e^{-2i\Omega}) e^{i\Omega n} d\Omega = \begin{cases} -2\frac{\cos n\pi}{n}, & n \neq 0 \\ 0, & n = 0 \end{cases} \quad (6.88)$$

and

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log_e |a|^2 e^{i\Omega n} d\Omega = \log_e |a|^2 \delta(n) \quad (6.89)$$

are derived, where the interval of integration is taken between  $-\pi$  and  $\pi$  so that if  $\log_e(e^{-2i\Omega})$  is an odd function over the interval, the integral is well defined. Consequently, the cepstral sequence is given by

$$c^+(n) = \begin{cases} -2\frac{\cos n\pi}{n}, & n > 0 \\ \log_e |a|^2, & n = 0 \\ 2\frac{\Re[a^n]}{n} - 2\frac{\cos n\pi}{n}, & n < 0 \end{cases} \quad (6.90)$$

subject to condition  $|a| > 1$ .

The minimum-phase cepstrum  $c^-(n)$  can now be obtained from  $c^+(n)$  for the non-minimum-phase cepstrum as follows. Recall that the non-causal part for the non-minimum phase part gives the all-pass cepstrum of an odd sequence as

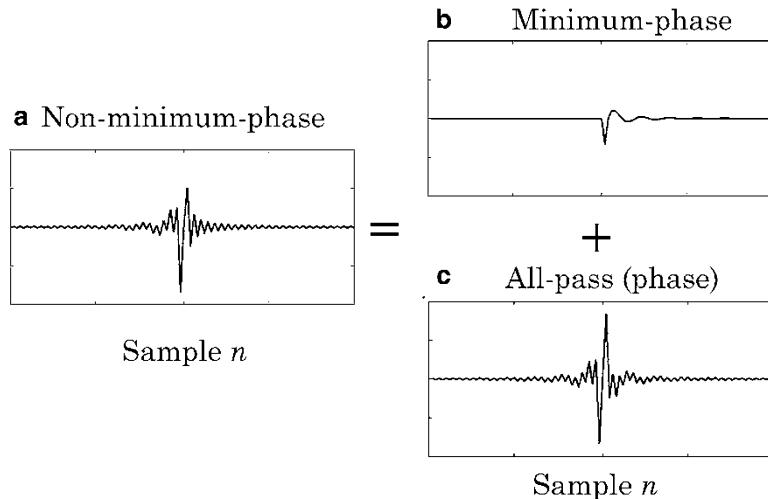
$$c_{ap}^+(n) = 2\frac{\Re[a^n]}{n} - 2\frac{\cos n\pi}{n} \quad (6.91)$$

for  $n < 0$  and is an odd sequence for  $n > 0$ ; it takes value 0 at  $n = 0$  by setting the all-pass magnitude spectrum to unity. Removing the all-pass cepstrum from  $c^+(n)$ ,

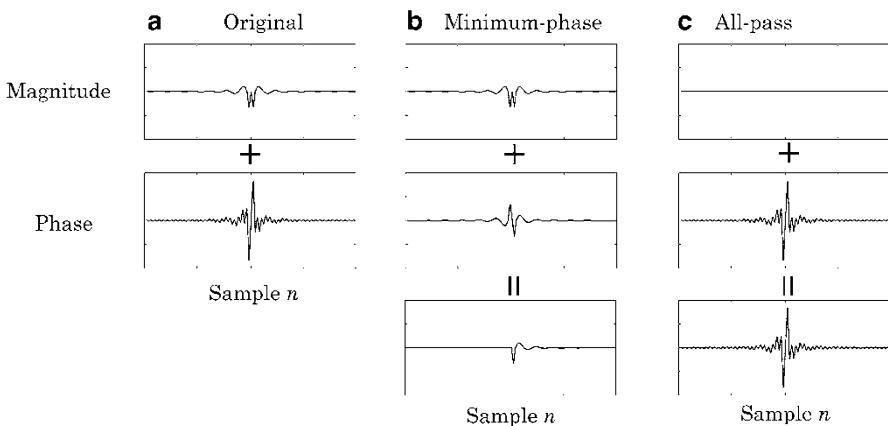
$$c^+(n) - c_{ap}^+(n) = \begin{cases} -2\frac{\Re[a^n]}{n}, & (n > 0) \\ 0, & (n = 0). \end{cases} \quad (6.92)$$

This result gives just the minimum-phase cepstrum. Similarly, the decomposition scheme into the magnitude and phase cepstral components holds.

Figure 6.17 on the facing page illustrates the samples of cepstral sequences expressed by Eqs. 6.87 and 6.90 for minimum-phase and non-minimum-phase cepstral sequences, respectively. Moreover, Fig. 6.18 on the facing page is a schematic display for the cepstral decomposition into magnitude and phase, and minimum-phase and all-pass cepstral sequences, respectively.



**Fig. 6.17** Examples of cepstral sequences; (a) non-minimum-phase cepstrum  $c^+(n)$  given by Eq. 6.90 on the preceding page where  $|a| = 1/0.95$  and  $\Omega_0 = \pm\pi/4$ ; (b) minimum-phase cepstral sequence  $c^-(n)$  given by Eq. 6.87 on the preceding page where  $|a| = 0.95$ ; (c) all-pass (phase) cepstrum  $c_{ap}^+(n)$  given by Eq. 6.91 on the preceding page



**Fig. 6.18** Schematic of cepstral decomposition; magnitude (a) and phase (b) cepstral sequences, and minimum-phase (B) and all-pass (C) cepstral sequences from Fig. 14.11 in [6]

## 6.5 Linear Phase and Ideal Low-Pass Filter

### 6.5.1 Linear Phase System

A reciprocal complex-conjugate pair of zeros with respect to the unit circle yields a linear phase response, in which the phase is linearly proportional to the frequency, as in Eq. 6.53 on page 109. Taking pairs of reciprocal complex-conjugate zeros. The transfer function corresponding to the zeros is given by

$$\begin{aligned} H_{sym-z}(z^{-1}) &= (1 - az^{-1})(1 - (a^*)^{-1}z^{-1})(1 - a^*z^{-1})(1 - a^{-1}z^{-1}) \\ &= |a^{-1}|^2 z^{-2} \left| (1 - az^{-1})(1 - a^*z^{-1}) \right|^2 \\ &= |a^{-1}|^2 z^{-2} [|a|^2 z^2 - 2\Re[a](1 + |a|^2)z^1 + 1 + |a|^4 \\ &\quad + (2\Re[a])^2 - 2\Re[a](1 + |a|^2)z^{-1} + |a|^2 z^{-2}] \text{ for } Z = e^{i\Omega} \end{aligned} \quad (6.93)$$

Interestingly, the impulse response is a symmetric (an even) sequence. The phase then becomes  $-2\Omega$  indicating a linear phase, and thus the group delay 2 gives the time delay of the envelopes. The results, indicating an even impulse-response sequence obtains from the linear phase characteristics, can be extended to general cases not only for two pairs of zeros but  $N$  pairs in general.

However, an even impulse-response sequence is not always linear phase. The frequency response for an even sequence can be formally expressed as

$$H_{even}(e^{-i\Omega}) = K \cdot R(e^{-i\Omega}) \cdot e^{-iN\Omega} \quad (6.94)$$

where  $K$  is a real constant and  $R(e^{-i\Omega})$  denotes a real function. If there is no sign change in the real function of  $\Omega$ , then the phase becomes  $-N\Omega$  indicating phase linearity. However, the real function possibly changes sign depending on  $\Omega$ , and the phase response cannot be linear in phase. In that case, the phase characteristics follow this linear trend in phase.

### 6.5.2 Ideal Low-Pass Filter and Fourier Transform

Low-pass filtering is a fundamental tool for signal analysis and processing. In particular, an ideal low-pass filter is fundamental to discrete signal processing and Fourier transformation. Suppose that the following Fourier transform of a function is given,

$$H_{LF}(e^{-i\Omega}) = \begin{cases} \frac{1}{2\Omega_c}, & |\Omega| \leq \Omega_c, \\ 0 & elsewhere \end{cases} \quad (6.95)$$

over the interval  $-\pi \leq \Omega < \pi$ . Taking the inverse Fourier transform, its impulse response is then expressed as

$$h_{LF}(n) = \frac{1}{2\pi} \frac{1}{2\Omega_c} \int_{-\Omega_c}^{\Omega_c} e^{i\Omega n} d\Omega = \frac{1}{2\pi} \frac{\sin \Omega_c n}{\Omega_c n} \quad (6.96)$$

for  $-\infty < n < \infty$ . The impulse response is an even non-causal sequence.

Therefore the continuous spectral function can be formally represented by the Fourier transform as

$$H_{LF}(e^{-i\Omega}) = \sum_{n=-\infty}^{+\infty} \frac{1}{2\pi} \frac{\sin \Omega_c n}{\Omega_c n} e^{-i\Omega n}. \quad (6.97)$$

Here note that the series above gives the average at a point of discontinuity, such as the cut-off frequency, as

$$H_{LF}(e^{-i\Omega})|_{\Omega=\Omega_c} = \frac{1}{2} \left[ H_{LF}(e^{-i\Omega})|_{\Omega=\Omega_c-0} + H_{LF}(e^{-i\Omega})|_{\Omega=\Omega_c+0} \right] \quad (6.98)$$

where  $\pm 0$  denote the limits taken from the right and left of  $\Omega_c$ , respectively. For the ideal low-pass filter defined by Eq. 6.95 on the preceding page, the infinite series above converges to  $1/4\Omega_c$  at the cut-off frequency [3]. Figure 6.19 on the following page presents an example of Fourier series expansion. Here, the convergence of the Fourier series at the discontinuous point can be graphically understood. The series converges to the average for both sides at the discontinuous point as the number of terms (length of the series) becomes infinite. However, an overshoot and undershoot remain for the left- and right-hand limits evaluated at the point of discontinuity. The length of the over (under)-shoot is known in general to be

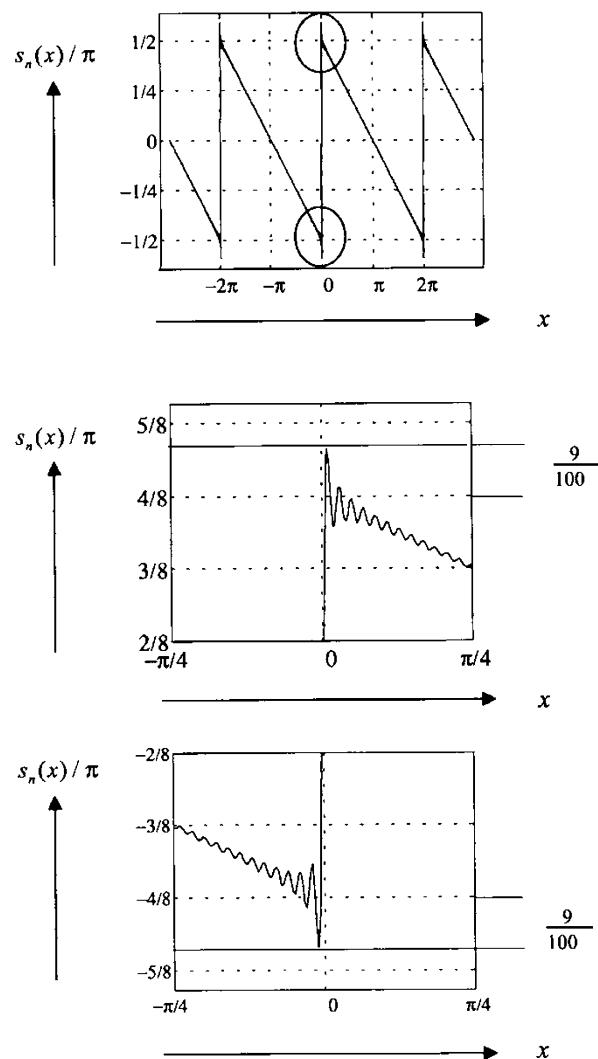
$$\Delta_{sh} \cong 0.089 \cdot |f(x_{dc} + 0) - f(x_{dc} - 0)| \quad (6.99)$$

where  $\Delta_{sh}$  denotes the length of the over (under)-shoot in terms of  $f(x_{dc} \pm 0)$ ,  $x_{dc}$  denotes the point of discontinuity, and  $|f(x_{dc} + 0) - f(x_{dc} - 0)|$  gives the length of the interval between the limits on the left and right of  $x_{dc}$  [3, 8].

### 6.5.3 Resonator and Cut-Off Frequency of Low-Pass Filter

The spectral function for the ideal low-pass filter is interesting from a point of view of poles and zeros. Intuitively, the impulse response can be interpreted as a sinusoidal sequence of the angular frequency  $\Omega_c$ , denoting the cut-off frequency, with decreasing amplitude as time progresses. Although the sinusoidal sequence does not show a naturally decaying sequence following an exponential function, it

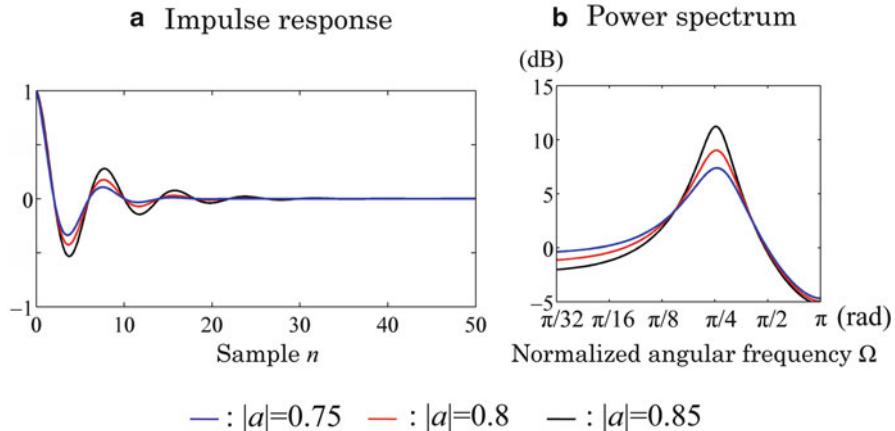
**Fig. 6.19** Sample of convergence of Fourier series expansion



can be understood as the free oscillation of a single resonator defined by a conjugate pair of poles instead. Consider the transfer function

$$H_p(z^{-1}) = \frac{1}{(1 - az^{-1})(1 - a^*z^{-1})} \quad (6.100)$$

where  $a = |a|e^{j\Omega_c}$  and  $a^*$  denotes the complex conjugate of  $a$ . In principle, a resonance system has frequency characteristics similar to a low-pass filter in which the frequency of the free oscillation corresponds to the cut-off frequency. In addition, the impulse response, which can be defined by the inverse  $z$ -transform



**Fig. 6.20** Impulse responses (a) and power spectral function (b) for a single resonator where the pole is located as  $z_0 = |a|e^{i\Omega_0}$ ,  $\Omega_0 = \pm\pi/4$ , and  $|a| = 0.75, 0.8, 0.85$

of the Eq. 6.100 on the preceding page is obtained by rewriting the transfer function into

$$H_p(z^{-1}) = \frac{A}{1 - az^{-1}} + \frac{B}{1 - a^*z^{-1}} \quad (6.101)$$

where

$$A = \frac{-a}{a^* - a} = B^*. \quad (6.102)$$

Thus, an infinite length of the impulse response, indicating an exponential decay in the sinusoidal sequence, is produced. The frequency of the sinusoidal sequence corresponds to the resonant frequency of the resonating system.

Figure 6.20 displays examples of the impulse response and power spectral characteristics of the resonator. The cut-off characteristics can be controlled by changing  $|a|$ . Indeed, as the poles move inside the unit circle, the slope at the cut-off frequency decreases.

#### 6.5.4 Zeros for a Truncated Ideal Low-Pass Filter

If the impulse response record is truncated to finite length, the transfer function can be represented by a polynomial instead of an infinite series. Recalling Eq. 6.97 on page 127; the partial sum for  $-N \leq n \leq N$  is taken for the truncation. The Fourier transform of the truncated impulse response can be written as

$$\begin{aligned}
H_{N-LF}(e^{-i\Omega}) &= \sum_{n=-N}^{+N} h(n)e^{-i\Omega n} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} H_{LF}(e^{-i\Omega'}) \sum_{n=-N}^{+N} e^{-i(\Omega-\Omega')n} d\Omega' \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} H_{LF}(e^{-i\Omega'}) D_N(\Omega - \Omega') d\Omega'
\end{aligned} \tag{6.103}$$

where

$$D_N(u) = \sum_{n=-N}^{+N} e^{inu} = \frac{\sin(N + \frac{1}{2})u}{\sin \frac{u}{2}} \tag{6.104}$$

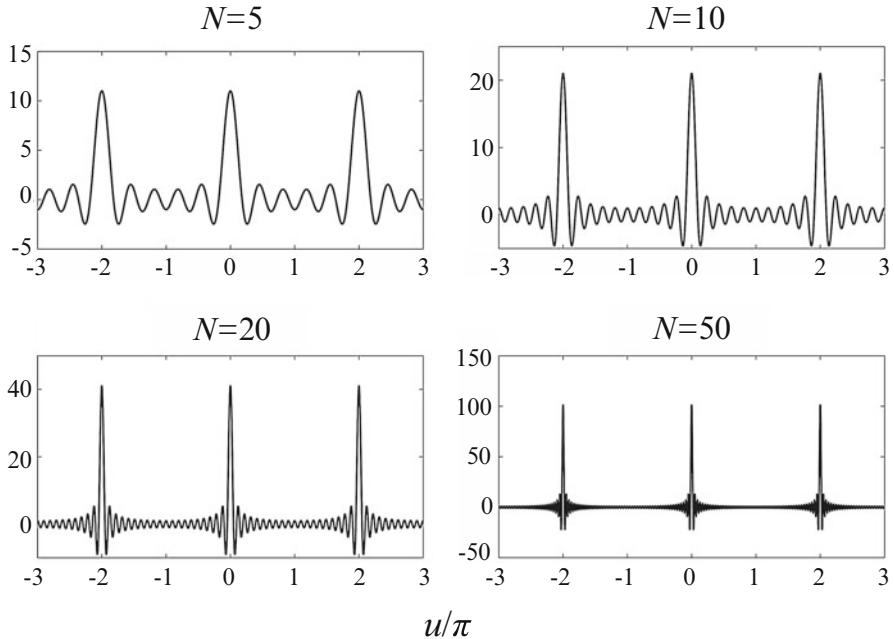
is the Dirichlet kernel [3]. Hence, the spectral function can be written as a convolution between the original, that might be obtained if the length of the series was infinite, and the Dirichlet kernel, which yields the spectral function of the truncating window function.

Figure 6.21 on the facing page depicts four Dirichlet kernels of differing lengths. The kernel approaches the  $\delta$ -function in the interval between  $-\pi$  and  $\pi$ , as the truncated sequence is lengthened; however, in general, it takes positive and negative signs for finite  $N$ . Therefore, the spectral response may also take positive and negative signs under convolution with the kernel. Indeed, Fig. 6.22 on the facing page shows the spectral function after convolution clearly deviating from the ideal defined by Eq. 6.95 on page 126, even in the region far from the cut-off frequency point (the point of discontinuity). Consequently, there are zeros observable under the truncation in the frequency response.

A similar truncation also applies for the impulse response of the resonator. The complex conjugate pair of poles that identifies the resonator disappears, but the zeros, except for the original pole, are densely and regularly placed on the circle with radius given by the distance of the pole locations from the origin of the  $z$ -plane. Figure 6.23 on page 132 shows an example of distributions of zeros for the polynomial  $H(z^{-1}) = \sum_{n=0}^{N-1} a_n z^{-n}$  in the complex frequency plane [2]. The polynomial  $H(z^{-1})$  is periodic in the unit circle. Suppose that  $a_n$  is a real random variable that follows a normal distribution. A set of  $a_n$  represents random amplitudes of reflections. To simulate the decay process of impulse responses, an exponential windowing function is applied to the random polynomials. In terms of acoustics, the variance of  $a_n$  decreases exponentially as the order of the coefficients increases. Figure 6.23 on page 132 illustrates samples of the distribution of zeros.

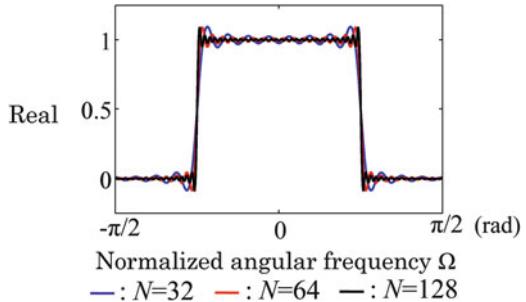
The density is concentrated on the line that is estimated by the reverberation time  $T_R$ . The high density line of the zeros is estimated by

$$\delta = \delta_0 \cong 6.9/T_R \tag{6.105}$$



**Fig. 6.21** Samples of Dirichlet kernels with  $N = 5, 10, 20, 50$

**Fig. 6.22** Samples of spectral functions for an ideal low-pass filter after convolution with Dirichlet kernels

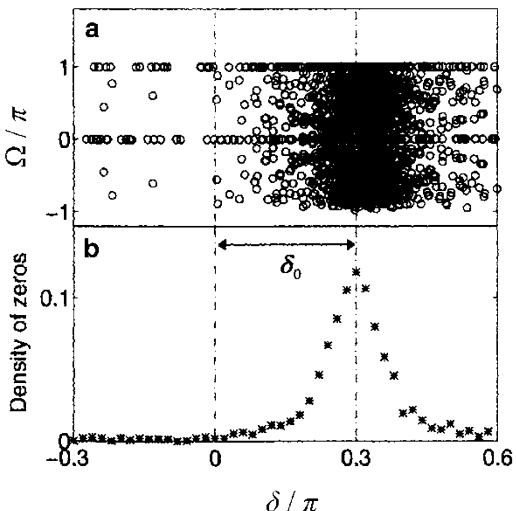


The reverberation time  $T_R$  is defined as the 60 dB decay time in the sound energy decay process. If the energy decay process follows  $E(t) = e^{-2\delta_0 t}$ , then the reverberation time is obtained by solving

$$10 \log e^{-2\delta_0 T_R} = -60. \quad (6.106)$$

The impulse response that is formulated by an exponentially decaying function such as that for a resonator can be interpreted as the limit case of the order of the polynomial approaches infinity. The concentrated distribution of the zeros can be replaced by a single pole instead. In other words the hidden resonance appears the order comes close to the limit [9].

**Fig. 6.23** Distribution of zeros of random sequence after exponential windowing:  
**(a)** distribution of zeros for 20 random polynomials of order 10,  
**(b)** density of zeros for distribution of **(a)** from Fig. 7.2.4 in [2]

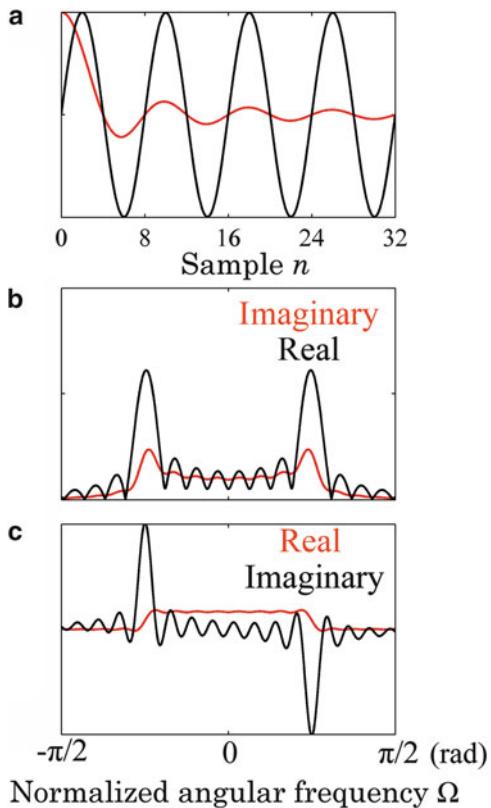


Interestingly, the pole-like characteristics (hidden resonance) can be seen in the spectral function even for the truncated ideal low-pass filter. After applying triangular window function with positive slope, the Fourier transform of the causal part of the impulse response given by Eq. 6.96 on page 127 appears as in Fig. 6.24 where the impulse response is truncated at  $N = \pm 32$ . The cut-off frequency is interpreted as corresponding to a virtual pole (or hidden resonance frequency).

### 6.5.5 Linear Phase and Ideal Low-Pass Filter

Linear phase filters are important tools in signal processing. This is because the narrow-band envelopes of an input sequence are simply delayed without deformation. The delay can be controlled by adding a time delay to the zero-phase sequence. An ideal low-pass filter is a candidate for linear-phase low-pass filtering, because it has a symmetric impulse response sequence. However, the ideal low-pass filter has no phase linearity from a theoretical point of view. As denoted by Eq. 6.99 on page 127 in Sect. 6.5.2 on page 126, even if an infinite length of series were taken, the spectrum, which is obtained by the Fourier transform of the impulse response, is able to take positive and negative signs by the so-called Gibbs phenomenon [3, 8]. Taking the auto-correlation sequence after applying a symmetric triangular windowing with negative slope is a possible way to obtain the zero-phase spectrum. A symmetric triangular window such as  $w_e(n) = N - 1 \pm n$  for  $n = 0, \pm 1, \dots, \pm(N - 1)$ , which itself is an auto-correlation sequence for a rectangular sequence, is a candidate for a window sequence that always yields the

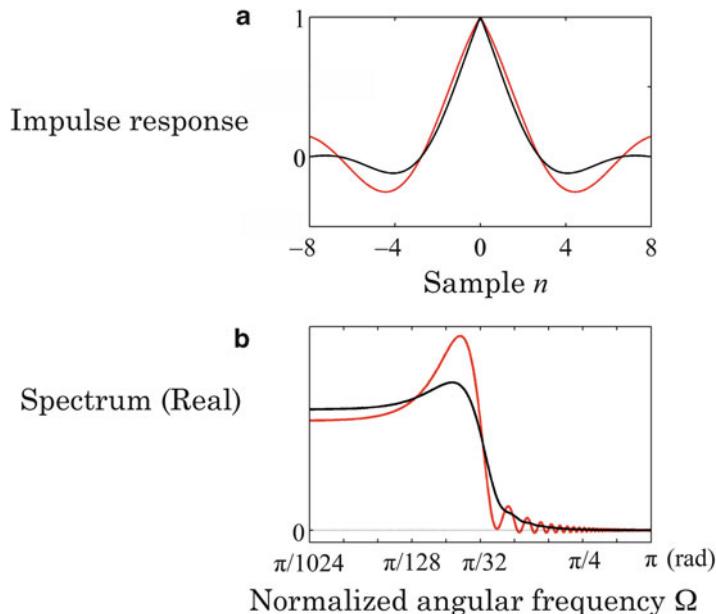
**Fig. 6.24** Spectral samples of causal part for truncated impulse response of ideal low-pass filter by triangular window with positive slope (red without windowing, black with windowing) where maximum of samples is normalized; (a) impulse response, (b and c) spectral components



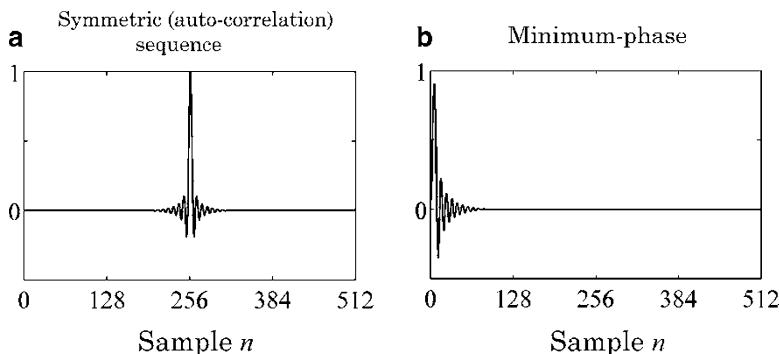
non-negative spectrum even after windowing the auto-correlation sequences. By regarding the auto-correlation sequence (after windowing) of the truncated ideal low-pass filter as the impulse response of a low-pass filter, it has a zero-phase spectrum. Thus, a linear-phase low-pass filter can be made by adding a finite time delay. Figure 6.25 on the following page illustrates the auto-correlation sequence and its zero-phase spectrum.

The linear phase system, which has a symmetric impulse response, might be good from the view point of preserving narrow-band envelopes. However, slow transient responses are inevitable because of the symmetric waveform of the impulse response. This fact might be problematic from an audio engineering type of application. A minimum-phase filter is another candidate of the low-pass filter. Figure 6.26 on the following page presents a minimum-phase filter constructed from the impulse response shown in Fig. 6.25 on the following page. The phase response is no longer the zero-phase, thus phase linearity is not achieved even if a time delay is added.

However, the transient response is not so slow, because the impulse response is causal and non-symmetric. The difference between the linear and



**Fig. 6.25** Impulse response equivalent to an auto-correlation sequence of a truncated ideal low-pass filter (a) and its spectral (real) function (b) where red: without windowing (rectangular windowing), black: triangular windowing with negative slope



**Fig. 6.26** Impulse response of minimum-phase low-pass filter made from auto-correlation sequence of truncated ideal low-pass filter where (a) shows the auto-correlation sequence used as impulse response and (b) gives minimum-phase derived from (a)

minimum-phase filters might be an interesting aspect for audio engineering application from a perceptual point of view, in particular, under binaural listening conditions [10]. Details of design methods for digital filters including low-pass filters are intensively described from theoretical and practical perspectives in [5,11].

## References

1. Y. Ando, *Auditory and Visual Sensation* (Springer, 2009)
2. M. Tohyama, T. Koike, *Fundamentals of Acoustic Signal Processing* (Academic Press, 1998)
3. A. Zygmund, *Trigonometric Series* (The University Press, Cambridge, 1988)
4. S. Lang, *Complex Analysis* (Springer, 1999)
5. A.V. Oppenheim, R.W. Shafer, *Digital Signal Processing* (Prentice-Hall, 1975)
6. M. Tohyama, *Sound and Signals* (Springer, 2011)
7. M.R. Schroeder, Improved quasi-stereophony and colorless artificial reverberation. *J. Acoust. Soc. Am.* **33**, 1061–1064 (1961)
8. P.J. Nahin, *Dr. Euler's Fabulous Formula: Cures Many Mathematical Ills* (Princeton University Press, 2006)
9. Y. Takahashi, M. Tohyama, Y. Yamasaki, Cumulative spectral analysis for transient decaying signals in a transmission system including a feedback loop. *J. Audio Eng. Soc.* **54**(7/8), 620–629 (2006)
10. Y. Hirata, Linear phase problem. Private communication, unpublished (2012)
11. G. Strang, T. Naguyen, *Wavelets and Filter Banks* (Wellesley Cambridge Press, 1997)

# Chapter 7

## Sampling Theorem and Discrete Fourier Transform

The Fourier transform of a discrete sequence yields a continuous complex function, the so-called continuous spectrum. This chapter describes how to express a periodic sequence using sampled spectral sequences in accordance with the sampling theorem and discrete Fourier transformation. The pair of time and spectral sequences forms a discrete Fourier transform (DFT) pair. The sampling theorem gives conditions and formulation for sampling a continuous function as a discrete sequence from which the original continuous function can be reconstructed without any deformation. The interpretation of the theorem is given in terms of the DFT and the partial sum of the Fourier series expansion for a continuous periodic function. In understanding the DFT schemes, it is informative seeing what happens if the sampling theorem is violated. Modifications of sequences through interpolation and decimation are displayed using numerical examples.

### 7.1 Sampling of Spectral Function

#### 7.1.1 *Simultaneous Equations for Representation of Periodic Sequence*

Suppose that a periodic sequence is expressed as a superposition of harmonic frequency components. Introducing complex sinusoidal functions, a periodic sequence can be written as

$$x(n) = \sum_{k=0}^{N-1} X(k) e^{i \frac{2\pi}{N} k n} \quad (7.1)$$

---

The original version of this chapter was revised.

An erratum to this chapter can be found at DOI [10.1007/978-4-431-54424-1\\_10](https://doi.org/10.1007/978-4-431-54424-1_10)

which can be interpreted as an expression using the spectral sequence  $X(k)$  sampled from the continuous spectral function  $X(e^{-i\Omega})$  at phase points  $\Omega = \Omega_k = (2\pi/N)k$ . In addition, an  $N$ -point sequence of  $x(n)$  is necessary for the complete determination of the  $X(k)$  with  $k = 0, 1, 2, \dots, N - 1$ . If the sequence  $x(n)$  is observed at  $n = 0, 1, 2, \dots, N - 1$ , then a set of  $N$  linearly independent simultaneous equations can be constructed to obtain the  $N$  unknown variables  $X(k)$ .

In general, matrix manipulations are necessary to solve the set of simultaneous linear equations [1]. However, recalling the orthogonality relation for the complex sinusoidal functions

$$\begin{aligned} \sum_{n=0}^{N-1} e^{i\frac{2\pi}{N}kn} \cdot e^{-i\frac{2\pi}{N}k'n} &= \sum_{n=0}^{N-1} e^{i\frac{2\pi}{N}(k-k')n} \\ &= N\delta(k - k' - pN) \end{aligned} \quad (7.2)$$

where  $p$  is an integer, the unknown variables  $X(k)$  can be obtained explicitly as

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-i\frac{2\pi}{N}kn}. \quad (7.3)$$

This expression is analogous to describing the inverse of a unitary matrix  $A$  as the transpose of the complex conjugate matrix of  $A$  [1].

### 7.1.2 Periodic Property in Time and Frequency Planes

Continuous periodic functions can be expanded as a Fourier series [2]. Spectral properties of the continuous periodic function are then expressed as a discrete sequence in the time domain corresponding to the coefficients in the Fourier series expansion. This expansion implies that, if a continuous periodic function is given in the frequency domain, its Fourier series representation yields a discrete time sequence. Specifically, a sequence in the time (frequency) domain yields a continuous periodic function in the frequency (time) domain. In terms of signal processing, the procedure whereby a continuous periodic function is obtained from a sequence is called the Fourier transform. The reverse procedure, to obtain the sequence from a continuous periodic function, is called the inverse Fourier transform, which replaces the Fourier series expansion.

The equations in the previous sub-section, in contrast, describe the relationship between a pair of two periodic sequences, a discrete time (spectral) sequence and a discrete spectral (time) sequence, one being the discrete Fourier transform (DFT) of the other.

## 7.2 Discrete Fourier Transform and Periodic Property

### 7.2.1 Discrete Fourier Transform Pair

Consider an  $N$ -point sequence  $x(n)$  with  $0 \leq n \leq N - 1$ . Assuming that the sequence is a single period taken from a periodic sequence with period  $N$ , then the sequence can be expressed as

$$\begin{aligned}\hat{x}(n) &= \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kn} \\ &= \begin{cases} x(n), & 0 \leq n \leq N - 1, \\ x(n + pN) & \text{otherwise,} \end{cases}\end{aligned}\quad (7.4)$$

where  $p$  is an integer, and

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} \hat{x}(n) e^{-j\frac{2\pi}{N}kn}. \quad (7.5)$$

With a finite record length of  $N$ , Eq. 7.5 defines the DFT of sequence  $x(n)$ , whereas Eq. 7.4 defines the inverse DFT. The discrete sequence of a spectrum or a line spectrum is a periodic sequence with period  $N$ , because the Fourier coefficients are periodic;

$$\begin{aligned}X(k + pN) &= \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi}{N}kn} \cdot e^{-j2\pi p} \\ &= X(k),\end{aligned}\quad (7.6)$$

where  $p$  is an integer. Similarly,  $\hat{x}(n)$  is periodic with period  $N$ . The proof uses the orthogonality of the sinusoidal functions, Eq. 7.2 on the preceding page:

$$\begin{aligned}\hat{x}(n) &= \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kn} \\ &= \frac{1}{N} \sum_{m=0}^{N-1} x(m) \sum_{k=0}^{N-1} e^{j\frac{2\pi}{N}k(n-m)} \\ &= \sum_{m=0}^{N-1} x(m) \delta(n - m - pN)\end{aligned}$$

$$= \begin{cases} x(n), & 0 \leq n \leq N - 1, \\ x(n + pN) & \text{otherwise.} \end{cases} \quad (7.7)$$

A DFT pair can be interpreted in terms of a filter bank. Using its DFT, a periodic sequence  $x(n)$  with period  $N$  can be represented in the form

$$\begin{aligned} x(n) &= \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kn} \\ &= \sum_{k=0}^{N-1} \frac{1}{N} \sum_{m=0}^{N-1} x(m) e^{j\frac{2\pi}{N}k(n-m)} \\ &= \sum_{k=0}^{N-1} x_k(n) \end{aligned} \quad (7.8)$$

where

$$\begin{aligned} x_k(n) &= \frac{1}{N} \sum_{m=0}^{N-1} x(m) e^{j\frac{2\pi}{N}k(n-m)} \\ &= X(k) e^{j\frac{2\pi}{N}kn}. \end{aligned} \quad (7.9)$$

This expression shows that the periodic sequence is a superposition of sub-band sequences for the filter bank. Here each sub-band filter has a complex sinusoidal impulse response

$$h_k(n) = e^{j\frac{2\pi}{N}kn} \quad (7.10)$$

for the  $k$ -th sub-band. Therefore every sub-band sequence  $x_k(n)$  is periodic with period  $N$ ; then

$$X(k) = x_k(pN) \quad (7.11)$$

holds.

### 7.2.2 Interpolation of Time Sequence

A periodic sequence is represented by its discrete spectral sequence. Sampling is a process whereby a discrete sequence is obtained from a continuous function. In contrast, reconstructing the original continuous function from a discrete sequence of a sampled signal of interest is called interpolation. Interestingly, the continuous

function can be reconstructed from a sampled discrete sequence, subject to the condition that the sequence can be represented by a DFT.

From its periodicity, the representation of a periodic sequence can be rewritten as

$$\begin{aligned} \sum_{k=-N}^N X(k) e^{i \frac{2\pi}{N} k n} &= \sum_{k=-(N-1)}^0 X(k) e^{i \frac{2\pi}{N} k n} + \sum_{k=0}^{N-1} X(k) e^{i \frac{2\pi}{N} k n} + X(0) \\ &= 2x(n) + X(0). \end{aligned} \quad (7.12)$$

Thus, assuming formally that

$$e^{i \frac{2\pi}{N} k n} = e^{i \frac{2\pi}{N} k t} \Big|_{t=n}, \quad (7.13)$$

then

$$\begin{aligned} x(t) &= \frac{1}{2} \sum_{k=-N}^N X(k) e^{i \frac{2\pi}{N} k t} - \frac{X(0)}{2} \\ &= \frac{1}{2} \sum_{k=-N}^N \left[ \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-i \frac{2\pi}{N} k n} \right] e^{i \frac{2\pi}{N} k t} - \frac{X(0)}{2} \end{aligned} \quad (7.14)$$

obtains. That is, the continuous function  $x(t)$  can be determined from the discrete spectral sequence  $X(k)$  by interpolation.

Swapping the order of the two summations, the interpolation is rewritten as

$$x(t) = \frac{1}{2N} \sum_{n=0}^{N-1} x(n) D_N(s) - \frac{X(0)}{2} \quad (7.15)$$

where  $D_N(s)$  denotes the Dirichlet kernel and  $s = (2\pi/N)(t - n)$ .

The interpolation formula can be established by verifying that

$$x(m) = x(t)|_{t=m}. \quad (7.16)$$

Recalling the Dirichlet kernel

$$D_N \left( \frac{2\pi}{N} (m - n) \right) = \begin{cases} 1 & m \neq n \\ 2N + 1 & m = n, \end{cases} \quad (7.17)$$

the interpolation formula written as 7.15 gives

$$\begin{aligned} x(t)|_{t=m} &= \frac{1}{2N} \left[ (2N+1)x(m) + \sum_{n=0, n \neq m}^{N-1} x(n) - \sum_{n=0}^{N-1} x(n) \right] \\ &= x(m). \end{aligned} \quad (7.18)$$

The result above determines that the continuous function can be interpolated by the convolution of the discrete sequence and the Dirichlet kernel.

### 7.2.3 Interpolation of Discrete Spectral Sequence

A similar discussion holds in spectral interpolation from a discrete spectral sequence, subject to the condition that the spectral sequence  $X(k)$  is obtained from a periodic sequence  $x(n)$ . Suppose that a spectral sequence  $X(k)$  is assumed to be the DFT of a periodic sequence  $x(n)$  with period  $N$ . Following the same argument given in the previous sub-section, the continuous spectral function  $X(\xi)$  from which the spectral discrete sequence  $X(k)$  is sampled can be interpolated as

$$X(\xi) = \frac{1}{2N} \sum_{k=0}^{N-1} X(k) D_N(r) - \frac{x(0)}{2} \quad (7.19)$$

where  $r = (2\pi/N)(\xi - k)$ .

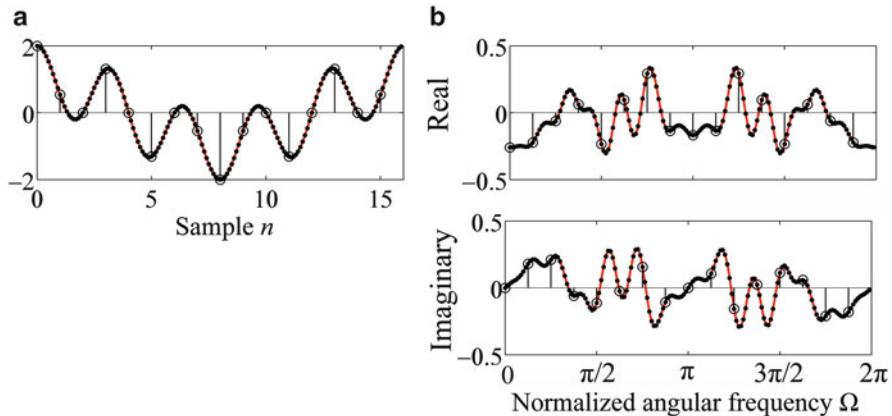
Figure 7.1 on the facing page shows an example of a time (a) and spectral (b) interpolation of a periodic sequence using Eqs. 7.15 on the previous page and 7.19 for (a) and (b), respectively. Numeric examples evidence this graphically; however, the time and spectral sequences are periodic, and such sequences can be represented by a finite record length that is equal to the period. The interpolation scheme holds in accordance with this periodicity.

The continuous function can be reconstructed from a periodic sampled sequence using the DFT between the periodic sequences. However, the spectral sequence is not always periodic, even if the original continuous function is accessible. A sampling theorem is given that enables conditions to be specified for when the original continuous function can be reconstructed, even if only a sampled version of the function is available.

## 7.3 Sampling Theorem

### 7.3.1 Partial Sum of Fourier Series

The sampling theorem is very fundamental in discrete waveform analysis; the literature and many textbooks fully describe this theorem [3,4]. However, understanding its mathematical implications is difficult because mathematical expressions of



**Fig. 7.1** Examples of time (a) and spectral (b) interpolation from discrete periodic sequences; open circles: sampled discrete sequences from original function (red), dotted: interpolated samples

Fourier-transformed non-periodic functions are necessary to formulate the theorem. Here, instead of a rigorous discussion, a different path is taken so that the theorem can be understood more intuitively. For that purpose, consider the sampling of a continuous periodic function  $x(t)$  whose spectral sequence is not always periodic. In addition, assume that  $x(t)$  has a finite-length spectral sequence.

Written as a real Fourier series,  $x(t)$  has the form

$$x(t) = \frac{1}{2}A_0 + \sum_{k=1}^{\infty} (A_k \cos kt + B_k \sin kt) \quad (7.20)$$

where the period is  $2\pi$ . Here a partial sum is taken by truncating terms with  $k > N$ . The sum can be written as

$$\begin{aligned} x_N(t) &= \frac{1}{2}A_0 + \sum_{k=1}^N (A_k \cos kt + B_k \sin kt) \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} x(\tau) \left[ \frac{1}{2} + \sum_{k=1}^N \cos k(t - \tau) \right] d\tau \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} x(\tau) D_N(t - \tau) d\tau, \end{aligned} \quad (7.21)$$

where

$$A_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} x(t) dt \quad (7.22)$$

$$A_k = \frac{1}{\pi} \int_{-\pi}^{\pi} x(t) \cos kt dt \quad (7.23)$$

$$B_k = \frac{1}{\pi} \int_{-\pi}^{\pi} x(t) \sin kt dt. \quad (7.24)$$

With

$$2 \sin \frac{u}{2} \left( \frac{1}{2} + \sum_{k=1}^N \cos ku \right) = \sin \left( N + \frac{1}{2} \right) u \quad (7.25)$$

then

$$\begin{aligned} \frac{D_N(u)}{2} &= \frac{1}{2} + \sum_{k=1}^N \cos ku \\ &= \frac{\sin(N + \frac{1}{2})u}{2 \sin \frac{u}{2}}. \end{aligned} \quad (7.26)$$

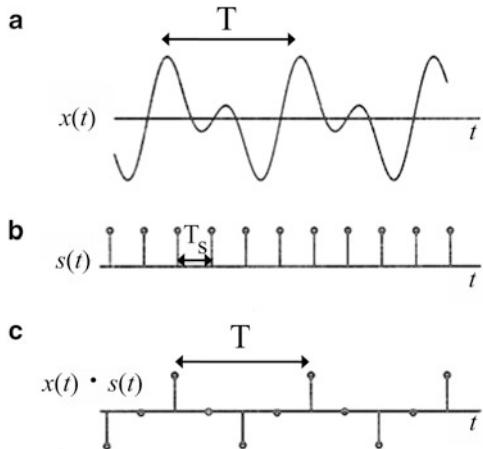
Here  $D_N(u)$  again denotes the Dirichlet kernel. The partial sum stated above can also be derived in complex function form,

$$\begin{aligned} x_N(t) &= \sum_{k=-N}^N C_k e^{ikt} \\ &= \frac{1}{2\pi} \sum_{k=-N}^N \int_{-\pi}^{\pi} x(\tau) e^{-ik\tau} d\tau e^{ikt} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} x(\tau) D_N(t - \tau) d\tau \end{aligned} \quad (7.27)$$

where  $D_N(u)$  is the Dirichlet kernel. The partial sum is given by the convolution of the original function defined before the truncation with the Dirichlet kernel. This result can be intuitively understood from a signal processing point of view, recalling that the Dirichlet kernel corresponds to the impulse response of an ideal low-pass filter corresponding to taking the partial sum.

Recall the interpolation from a discrete sequence to obtain the continuous function. The interpolation formula was expressed as a convolution (see Eq. 7.18 on page 142). The reconstructed original function is represented by a partial sum of the sampled periodic sequence. This suggests that if the partial sum of a discrete sequence, obtained after sampling the original continuous function, satisfies the necessary and sufficient condition required in reconstructing the original function, then the original continuous function can be reconstructed by convolution with the Dirichlet kernel even from the sampled sequence.

**Fig. 7.2** Example of sampling a continuous periodic function with period  $T$  (a), sampling function with period  $T_s$  (b), and periodic sampled sequence with period  $T$  (c) from Fig. 2.4.1 in [4]



### 7.3.2 Periodic Property of Spectrum for Sampled Sequence

Recall the spectral property of a periodic function. The spectral function is described as a discrete sequence whose interval represents the fundamental frequency (i.e., the inverse of the period). Figure 7.2 gives an example of a sampling of a continuous function with period  $T$  (a) yielding periodic sequence (c); (b) shows a periodic pulse sequence representing a sampling obtained with period  $T_s$ . The sampling sequence can be formally expressed in continuous function form using continuous variable  $t$  in a Fourier series expansion

$$s(t) = \sum_{l=-\infty}^{\infty} \delta(t - lT_s) = \sum_{k=-\infty}^{\infty} C_k e^{j \frac{2\pi k}{T_s} t} \quad (7.28)$$

where  $l$  is an integer and

$$C_k = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} T_s \delta(t) e^{-j \frac{2\pi k}{T_s} t} dt = T_s. \quad (7.29)$$

Therefore, the sampled sequence can also be represented using a continuous variable

$$x_s(t) = x(t)s(t) = x(t)T_s \sum_{k=-\infty}^{\infty} e^{j \frac{2\pi k}{T_s} t} \quad (7.30)$$

where the sampled sequence is assumed periodic by choosing  $T = NT_s$ . Assuming periodicity, the sampled spectral sequence is written as

$$\begin{aligned}
X_s(k) &= \frac{1}{T} \int_{-T/2}^{T/2} x_s(t) e^{-i \frac{2\pi k}{T} t} dt \\
&= \frac{T_s}{T} \int_{-T/2}^{T/2} \left[ \sum_{k'=-\infty}^{\infty} e^{i \frac{2\pi k'}{T_s} t} \right] x(t) e^{-i \frac{2\pi k}{T} t} dt \\
&= \frac{T_s}{T} \sum_{k'=-\infty}^{\infty} \int_{-T/2}^{T/2} x(t) e^{-i \left( \frac{2\pi k}{T} - \frac{2\pi k'}{T_s} \right) t} dt \\
&= T_s \cdot X * S(2\pi k/T)
\end{aligned} \tag{7.31}$$

where

$$X \left( \frac{2\pi}{T} k \right) = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-i \frac{2\pi k}{T} t} dt \tag{7.32}$$

$$S \left( \frac{2\pi}{T} k \right) = \sum_{k'=-\infty}^{\infty} \delta \left( \frac{2\pi}{T} (k - Nk') \right). \tag{7.33}$$

This result shows that the Fourier transform of the sampled sequence is periodic with period  $N$ , which denotes the number of samples obtained in a single cycle of the original function  $x(t)$ , i.e.,  $T = NT_s$ . Here, the Fourier transform of the sampled function is equivalent to the DFT of the sampled sequence to within a normalization factor [4];

$$\begin{aligned}
X_s(k) &= \frac{1}{T} \int_{-T/2}^{T/2} x_s(t) e^{-i \frac{2\pi k}{T} t} dt \\
&= \frac{T_s}{T} \int_{-T/2}^{T/2} x(t) \sum_{n=-(N/2-1)}^{N/2} \delta(t - nT_s) e^{-i \frac{2\pi k}{T} t} dt \\
&= \frac{1}{N} \sum_{n=-(N/2-1)}^{N/2} x(n) e^{-i \frac{2\pi kn}{N}}
\end{aligned} \tag{7.34}$$

where  $x(n) = T_s x(nT_s)$ .

The spectral sequence of a sampled sequence derived by sampling a continuous periodic function is represented as a periodic expansion of the original spectral sequence before sampling. The period is given by  $1/T_s$  where  $T_s$  denotes the sampling interval. Nevertheless, note again that the original spectral sequence is not periodic in general.

### 7.3.3 Sampling Partial Sum

Suppose that the original continuous function with period  $T$  is composed of a finite number of sinusoidal components in a Fourier series expansion. Such a function can be expressed as a partial sum. Assuming the function has  $K$  sinusoidal (or spectral) components

$$x(t) = \sum_{k=-(K/2-1)}^{K/2} X(k) e^{ik\omega_0 t}. \quad (7.35)$$

The spectral sequence of a sampled time sequence is then given by

$$X_s(k) = X * S(k) \quad (7.36)$$

where  $N$  denotes the number of samples contained in a single period and

$$S(k) = \sum_{k'=-\infty}^{\infty} \delta(k - k'N). \quad (7.37)$$

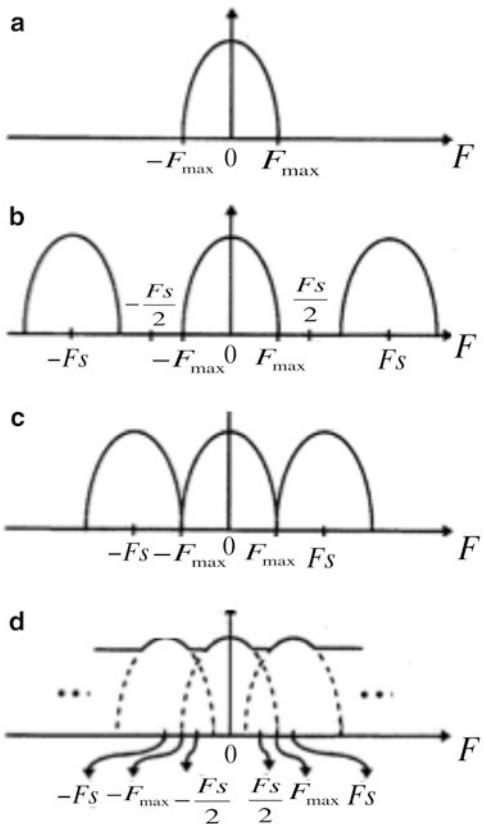
Figure 7.3 on the following page gives various examples of the spectral sequence after sampling [4]. Here, the relationship between  $K(2F_{\max})$  and  $N(F_s)$  is crucial in understanding the sampling theorem. The top panel (a) shows the original spectral sequence composing the partial sum before sampling. The length of the spectral sequence is  $K(2F_{\max})$ . Panels (b), (c), and (d) illustrate the spectral sequences observed after sampling. For panel (b),  $N > K(F_s > 2F_{\max})$ . The spectral sequence is periodic with period  $N(F_s)$  and the original sequence is repeated without any changes to the periodic sequence. The original spectral sequence is the partial sum of the spectral sequence after sampling.

Panels (c) and (d) give results for  $N = K(F_s = 2F_{\max})$  and  $N < K(F_s < 2F_{\max})$ , respectively. In contrast to panel (b), in both (c) and (d), the original spectral sequence can no longer be preserved after sampling. The examples (i.e., Eq. 7.36) determine that the sampling period  $T/N(1/F_s = T_s)$  must be shorter than  $T/K(1/2F_{\max})$  to reconstruct the original spectral sequence even after sampling.

### 7.3.4 Sampling Theorem

To reconstruct the original function from the sampled sequences, sampling frequencies  $F_s = 1/T_s$  must be higher than  $2F_{\max}$  where  $F_{\max}$  denotes the highest frequency component of the partial sum that is present in the original continuous periodic function. That is, one needs only the partial sum of the periodic sequence obtained by the DFT of the time sequence sampled from the original function

**Fig. 7.3** Examples of spectral sequences and sampling conditions; (a) original spectral sequence with length  $K(2F_{\max})$ ; (b)–(d) spectral sequences after sampling where (b)  $T/N < T/K(F_s > 2F_{\max})$ , (c)  $N = K(F_s = 2F_{\max})$ , and (d)  $N < K(F_s < 2F_{\max})$  from Fig. 2.4.2 in [4]



to reconstruct the original function. This reconstruction formula, which holds if  $F_s > 2F_{\max}$  is called the sampling theorem.

Formally, the reconstruction formula is the same as that for interpolation defined by Eq. 7.15 on page 141. That is,

$$x_o(t) = \frac{1}{2N} \sum_{n=0}^{N-1} x_s(n) D_N \left( \frac{2\pi}{N}(t-n) \right) - \frac{X_s(0)}{2} \quad (7.38)$$

where  $x_o(t)$  is the original function that is represented by the partial sum,  $x_s(n)$  the sequence sampled from the original function, and  $D_N(s)$  denotes the Dirichlet kernel. The sampling theorem says that the original continuous function can be perfectly reconstructed by interpolation with the Dirichlet kernel or by taking only the partial sum of the spectral sequence. The latter can be interpreted as low-pass filtering. Actually, the sampling theorem [3–5] holds not only for periodic functions but for non-periodic functions that are represented by continuous spectral functions

instead of line spectral sequences. The Dirichlet kernel can then be replaced by the sync function representing the impulse response of an ideal low-pass filter.

## 7.4 Discrete Fourier Transform and Sampling Theorem

### 7.4.1 Sampling of Sinusoidal Function

Consider a sampled sequence of a complex sinusoidal function

$$x(n) = A e^{i(2\pi/N)k_o n}. \quad (7.39)$$

Taking the DFT yields

$$\begin{aligned} X(k) &= \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-i(2\pi/N)kn} \\ &= \frac{1}{N} A \sum_{n=0}^{N-1} e^{-i(2\pi/N)(k-k_o)n} \\ &= A \delta(k - k_o). \end{aligned} \quad (7.40)$$

That is, the spectral sequence of a complex sinusoid is a single pulse on  $k = k_o$ . If the real sinusoidal sequences

$$A \cos \frac{2\pi}{N} k_o n = x(n) = \frac{A}{2} (e^{i(2\pi/N)k_o n} + e^{-i(2\pi/N)k_o n}) \quad (7.41)$$

$$B \sin \frac{2\pi}{N} k_o n = y(n) = \frac{B}{2i} (e^{i(2\pi/N)k_o n} - e^{-i(2\pi/N)k_o n}) \quad (7.42)$$

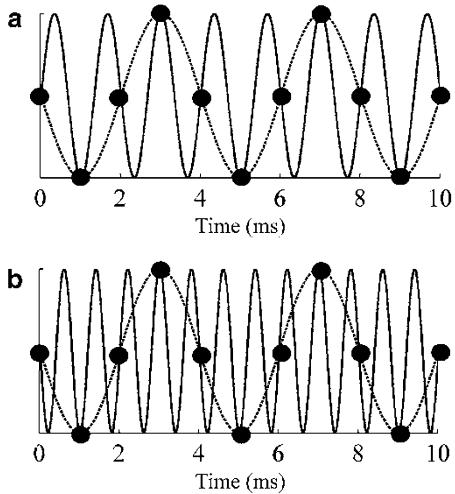
are given, the DFTs are then

$$X(k) = \frac{A}{2} [\delta(k - k_o) + \delta(k + k_o)] \quad (7.43)$$

$$Y(k) = \frac{B}{2i} [\delta(k - k_o) - \delta(k + k_o)] \quad (7.44)$$

respectively. Here, the frequency for a negative  $k$  is called the negative frequency. Note that the real sinusoidal functions are decomposed into a pair of complex sinusoidal functions with positive and negative frequency components. In addition, the spectral sequence for the cosine (sine) function is a real (imaginary) even (odd) sequence. Therefore the DFT of a real sequence is a complex sequence composed of a real even part and an imaginary odd component.

**Fig. 7.4** Examples of sinusoidal sequences sampled every 1 ms ( $F_s = 1,000$  Hz); the sinusoidal functions have frequency (a) 750 Hz, (b) 1,250 Hz from Fig. 21.6 in [6]



The DFT generates complex conjugate pairs such that

$$X^*(k) = X(N - k) \quad (7.45)$$

for  $0 < k < N/2 - 1$  where

$$X(k) = X(-(N - k)) = X(k + pN) \quad (7.46)$$

for  $0 \leq k \leq N - 1$  and  $p$  is an integer. Here, recall the definition of the normalized angular frequency,  $\Omega = \omega T_s = 2\pi/F_s$ . The index of the spectral sequence  $k$  determines the frequency of  $F_s/2$  when  $k = N/2$ , and similarly it determines the sampling frequency  $F_s$  by  $k = N$ . Therefore, in upholding the sampling theorem, any sinusoidal component must be located at  $k < N/2$ , and consequently, the spectral sequence generates a complex conjugate pair of regions where  $k < N/2$  for one and  $k > N/2$  for the other of the pair. Although the spectral sequence seems to have higher frequency components that violate the upper limit regulated by the sampling theorem, those higher components are complex conjugates to those sinusoidal components that meet the sampling theorem requirements.

Figure 7.4 presents examples of sampled sinusoidal functions [6]. Panel (a) shows a sinusoidal function with frequency 750 Hz and a sampled sequence sampled at every 1 ms (or  $F_s = 1,000$  Hz). Interestingly, the sampled sequence looks like a sinusoidal sequence with a frequency of 250 Hz. This is a result that appears to violate the sampling theorem. Suppose that the  $N$ -point gives the length of the DFT. The spectral sequence produced by sampling the sinusoidal function formally corresponds to a pair of samples at  $k = 3N/4$  and  $k = N/4$ . Consequently, the result is interpreted as the original sinusoidal function having a frequency of 250 Hz, because  $k = N/4$  corresponds to the positive frequency component for a sinusoidal

function with frequency 250 Hz. Similarly,  $k = 3N/4$  is associated with negative frequency. The apparent contradiction is explained as a violation of the sampling theorem, which produces an effect called aliasing, where a ‘stranger’ comes into the negative (positive) frequency region from the positive (negative) region.

Panel (b) again shows the same spectral sequence, but the sampling sequence looks like a sinusoid with frequency 250 Hz, despite the original sinusoidal sequence having a frequency of 1,250 Hz. This is because the sampled sequence is obtained with  $F_s = 1,000$  Hz, which does not meet the sampling condition. These examples should help in obtaining an intuitive understanding of the sampling theorem.

## 7.5 Interpolation and Decimation of Sequences

### 7.5.1 *Interpolation of Sequences*

Another expression is given for the interpolation of sequences that was described in Sect. 7.2.3 on page 142. Suppose that a time sequence  $x(n)$  is given in time interval  $0 \leq n \leq N - 1$ . The spectral interpolation can be formulated in the time domain as well as the spectral domain in which the Dirichlet kernel is defined. Defining  $\hat{x}(n)$  as the sequence obtained by appending zeros to  $x(n)$  after  $n > N - 1$

$$\hat{x}(n) = \begin{cases} x(n), & 0 \leq n \leq N - 1 \\ 0, & N - 1 < n \leq M - 1 \end{cases} \quad (7.47)$$

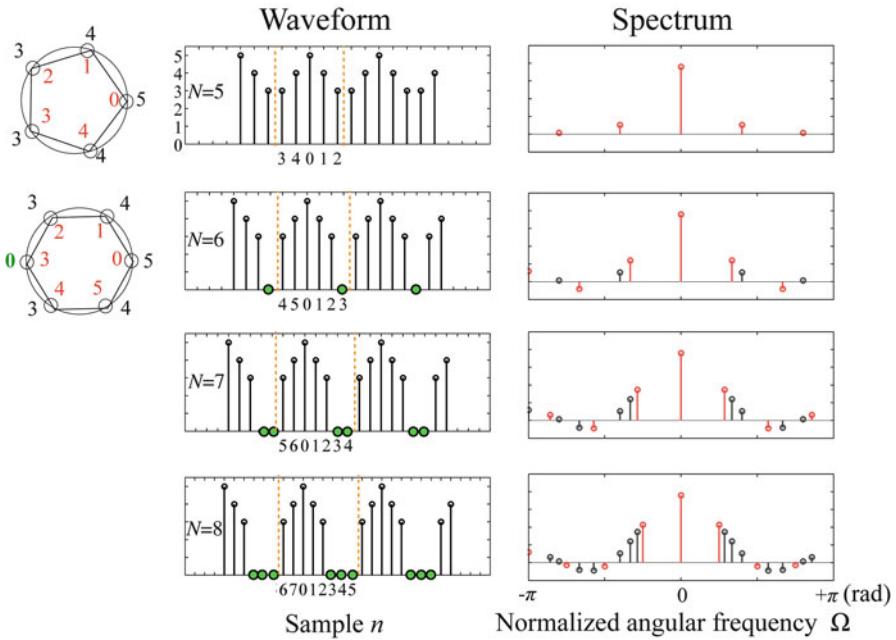
its DFT is then given by

$$\hat{X}(k) = \frac{1}{N} \sum_{n=0}^{N-1} \hat{x}(n) e^{-i(2\pi k/M)n}. \quad (7.48)$$

This equation can be rewritten as

$$\begin{aligned} \hat{X}(l) &= \frac{1}{N} \sum_{n=0}^{N-1} \left[ \sum_{k=0}^{N-1} X(k) e^{i(2\pi k/N)n} \right] e^{-i(2\pi l/M)n} \\ &= \frac{1}{2N} \sum_{k=0}^{N-1} X(k) \cdot D_N(r) - \frac{x(0)}{2} \end{aligned} \quad (7.49)$$

where  $r = 2\pi(l/M - k/N)$ . The result is the same as Eq. 7.19 on page 142. The original spectral sequence with length  $N$  is interpolated as one with length  $M$ . Thus, a spectral interpolation can be performed by appending trailing zeros to the time sequence as well as by convolution in the spectral domain using the Dirichlet kernel. Similarly, interpolation of a time sequence can be realized by adding zeros to a spectral sequence.



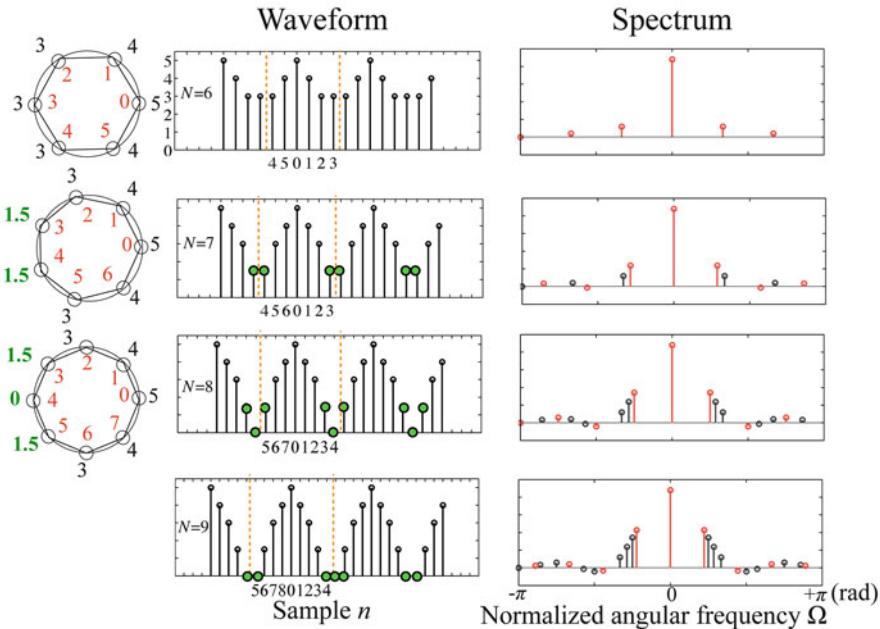
**Fig. 7.5** Interpolation of time sequence from symmetric spectral sequence composed of odd number of entries

However, appending zeros to spectral sequences or symmetric time sequences such as the auto-correlation sequences may be puzzling because of the periodic nature of the sequences. Figures 7.5 and 7.6 on the facing page help in giving an intuitive understanding of this procedure from a practical point of view.

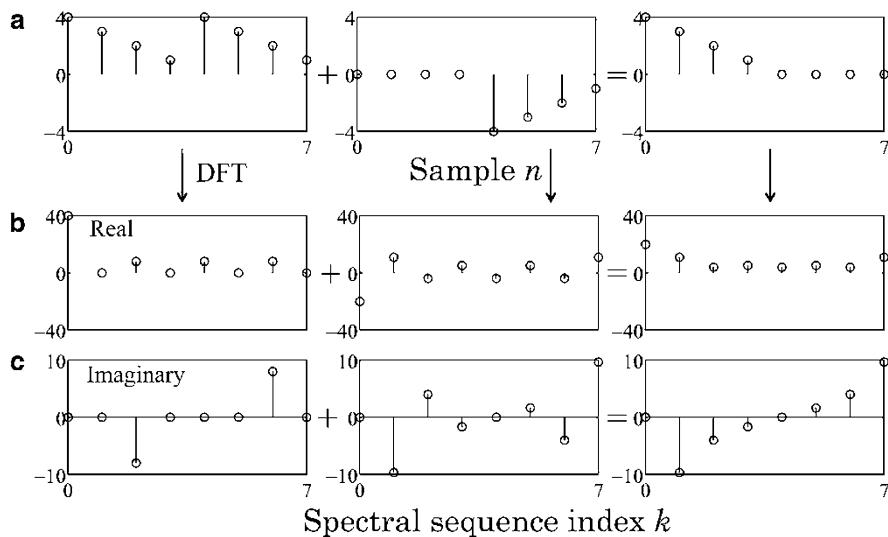
### 7.5.2 Interpolation and Decimation Samples

Figure 7.7 on the facing page gives a different view to the relationship between interpolation and decimation. A sequence in which zeros are appended can be decomposed into two sequences, one composed of two cycles of the original sequence and the other a shifted opposite-sign sequence. The original sequence is just repeated, if the interpolated samples are removed and replaced by the zeros.

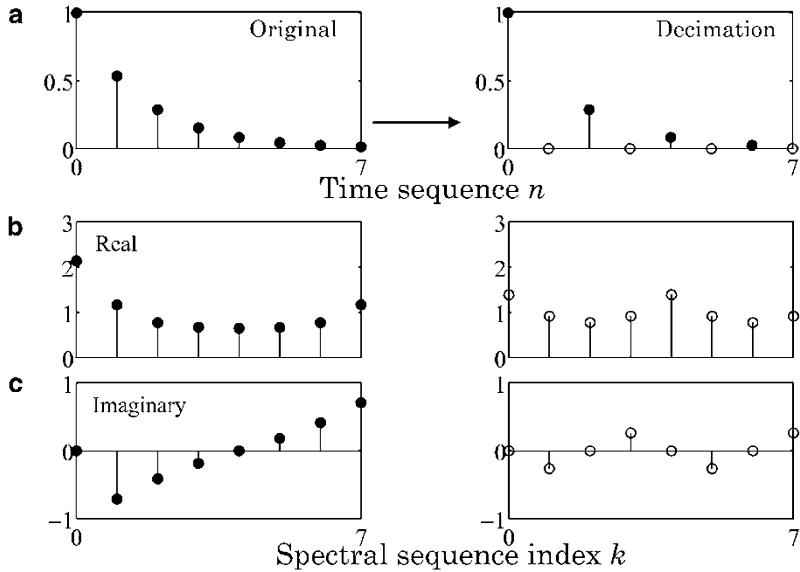
Suppose the original time (spectral) sequence of  $N$ -samples is one with modified  $M$ -samples after decimation. The spectral (time) sequence is also periodic with period  $M$ . If the original sequence is preserved even in the modified sequence of period  $M$ , decimation can be safely carried out. Take once more a time sequence composed of  $N$  samples. A reconstructed time sequence using only  $M = N/p$  spectral components can be written as



**Fig. 7.6** Similar to Fig. 7.5 on the preceding page but with even number of entries



**Fig. 7.7** Relationship between interpolation and decimation



**Fig. 7.8** Sample of decimation of time sequence

$$\begin{aligned}
 \hat{x}(n) &= \sum_{k=0}^{\frac{N}{p}-1} X(p \cdot k) e^{i(2\pi p \cdot k / N)n} \\
 &= \sum_{k=0}^{M-1} \left[ \frac{1}{N} \sum_{l=0}^{N-1} x(l) e^{-i(2\pi p \cdot k / N)l} \right] e^{i(2\pi p \cdot k / N)n} \\
 &= \frac{1}{N} \sum_{k=0}^{M-1} \left[ \frac{1}{N} \sum_{l=0}^{N-1} x(l) e^{-i(2\pi k / M)l} \right] e^{i(2\pi k / M)n} \\
 &= \frac{1}{N} \sum_{l=0}^{N-1} x(l) \sum_{k=0}^{M-1} e^{-i(2\pi k / M)(l-n)} \\
 &= \frac{M}{N} \sum_{l=0}^{N-1} x(l) \delta(l - n \pm Mq)
 \end{aligned} \tag{7.50}$$

where  $p$  and  $q$  are positive integers. The reconstructed sequence is periodic with period  $M$  and different from the original sequence  $x(n)$ .

There are examples of modifications of sequences in Figs. [7.8, 7.9 on the facing page](#), [7.10 on the facing page](#), and [7.11 on page 156](#); these help in intuitively understanding DFT scheme. Following Eq. [7.48 on page 151](#) the inverse DFT can be generalized as

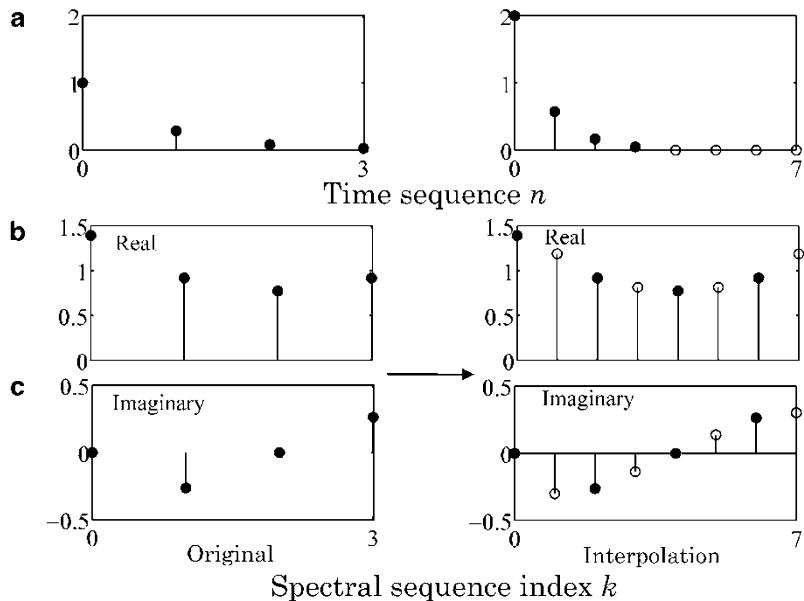


Fig. 7.9 Spectral interpolation

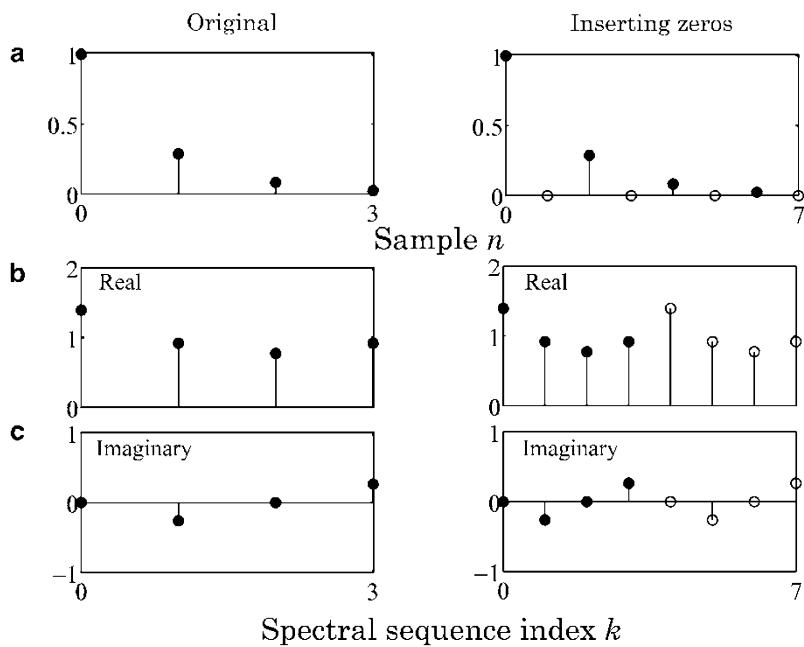
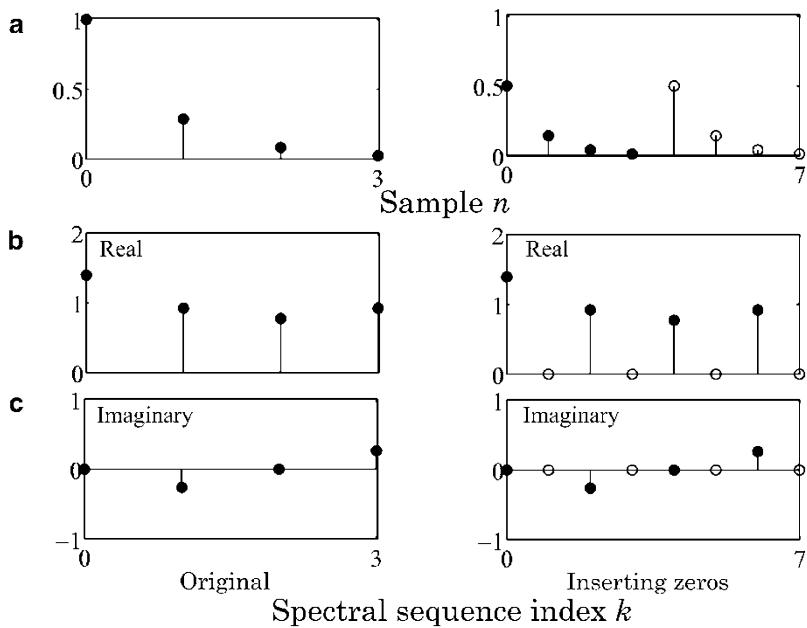


Fig. 7.10 Inserting zeros in time sequence



**Fig. 7.11** Inserting zeros in spectral sequence

$$\hat{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \left[ \sum_{n=0}^{N-1} x(n) e^{-i(2\pi k/N)n} \right] e^{i(2\pi k/M)n}. \quad (7.51)$$

If  $M > N$  ( $M < N$ ) is taken, then interpolation (decimation) is performed on the time sequence  $x(n)$ . Similarly, the modifications are performed on the spectral sequence following the next equation

$$\hat{X}(k) = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \sum_{k=0}^{N-1} X(k) e^{i(2\pi k/N)n} \right] e^{-i(2\pi k/M)n}. \quad (7.52)$$

### 7.5.3 Sampling Frequency Conversion

The interpolation formula can also be interpreted as a conversion of a sampling frequency. To simplify the formulation, a low-pass filtered analytic sequence is introduced here;

$$\begin{aligned} z_L(n) &= x(n) + i y(n) \\ &= \sum_{k=0}^{L-1} \hat{X}(k) e^{i(2\pi kn/N)} \end{aligned} \quad (7.53)$$

where

$$\hat{X}(k) = \begin{cases} 2X(k) & (0 < k < N/2) \\ 0 & (N/2 + 1 \leq k \leq N - 1) \end{cases} \quad (7.54)$$

and

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-i(2\pi kn/N)} \quad (7.55)$$

$$X(0) = X(N/2) = 0 \quad (7.56)$$

under the condition that  $L < N/2$ . Here, the analytic sequence can be obtained from a real sequence  $x(n)$  by the convolution formula for the time domain:

$$z_L(n) = \frac{2}{N} \sum_{q=0}^{L-1} x(q) D_L(n-q) \quad (7.57)$$

where

$$D_L(n) = \sum_{k=0}^{L-1} e^{i(2\pi kn/N)} \quad (7.58)$$

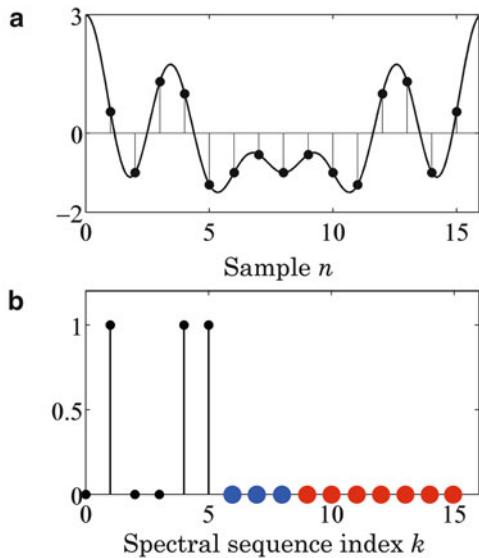
$$= \begin{cases} L, & (n = pN) \\ \frac{1-e^{i(2\pi L n/N)}}{1+e^{i(2\pi n/N)}}, & (n \neq pN) \end{cases} \quad (7.59)$$

and  $p$  an integer. Therefore, if integer  $M$ , which can be greater (smaller) than  $N$ , is substituted for  $N$  in Eq. 7.53, an up-sampled (down-sampled) sequence is obtained. Here  $D_L(n)$  can be interpreted as an analytic analog of the Dirichlet kernel. In addition, when  $M = N$  the analytic formula for the original sequence is derived.

An up-sampled sequence  $z_{LU}(n)$  can be written as

$$z_{LU}(n) = \frac{2}{N} \sum_{q=0}^{N-1} x_L(q) D(n-q) \quad (7.60)$$

**Fig. 7.12** Sample of a time sequence and its power spectral sequence; (a) original waveform (*solid line*) and samples of a sequence composed of 16 entries (real part of analytic sequence), (b) power spectral sequence including zeros for the negative frequency (*red plots*) components



where

$$D(n - q) = \sum_{k=0}^{L-1} e^{j2\pi k(n/M - q/N)} \quad (7.61)$$

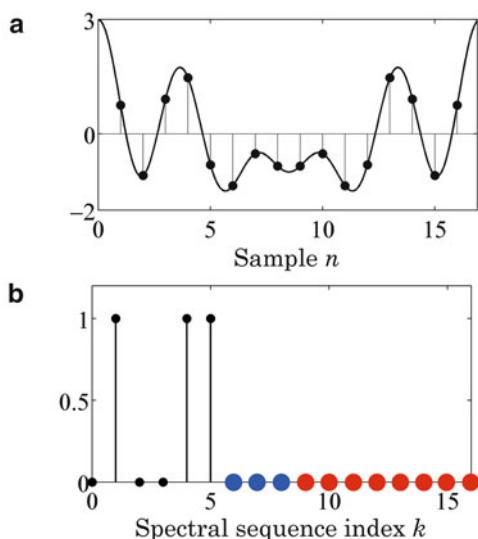
and  $M > N$ . By taking the real part of  $z_{LU}(n)$ , the up-sampled version of  $x(n)$  is derived.

Figure 7.12a shows the original sequence  $x(n)$ , which is composed of 16 samples; its power spectral sequence is illustrated in Fig. 7.12b. The solid line in Fig. 7.12a shows the interpolated curve, from which the sequence in Fig. 7.12a is sampled. It can be up-sampled to the sequence made of 17 samples; see Fig. 7.13 on the facing page (a) (time sequence) and Fig. 7.13 on the facing page (b) (power spectral sequence). Here, the solid line is identical to that in Fig. 7.12a. Thus, the up-sampled sequence is understood as being a re-sampling of the original function (the solid curve in panel (a)). In addition, it can be seen that a single zero is added into the spectral sequence in the negative frequency region (red plots in Fig. 7.13 on the facing page (b)).

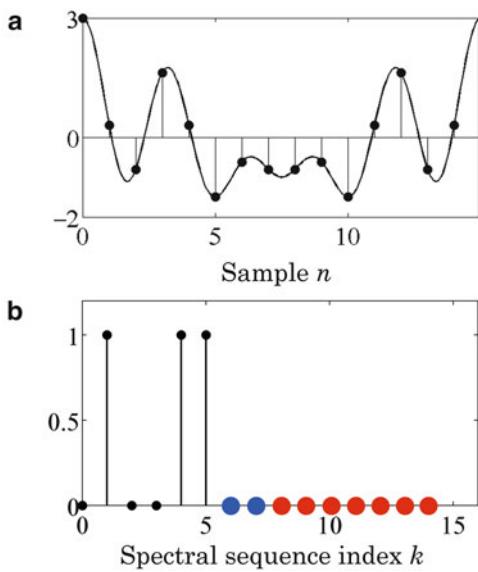
Similarly, by taking  $M$  smaller than  $N$  but greater than  $2L$ , a down-sampled version of the original sequence  $x(n)$  can be produced. The original sequence (Fig. 7.12) is down-sampled to a new sequence composed of 15 samples (Fig. 7.14 on the facing page (a)); the solid curve gives the same original function. In contrast to up-sampling, a single zero is removed from the spectral sequence in the positive frequency components (blue plots in Fig. 7.14 on the facing page (b)).

Up- and down-sampling can be interpreted as interpolation and decimation, respectively. A formulation has been given in analytical-sequence form, but applying convolution over the time domain instead of the frequency plane. The kernel

**Fig. 7.13** Up-sampled sequences from the sequence shown in Fig. 7.12a on the preceding page to 17 samples; (a) and (b) show similar sequences to Fig. 7.12a and b on the preceding page, respectively where the solid line is identical to that for Fig. 7.12a on the preceding page



**Fig. 7.14** Down-sampled sequences from sequences in Fig. 7.12 on the preceding page to 15 samples, where panels are similarly interpreted as in Figs. 7.12 on the preceding page and 7.13. (a) Sampled sequence. (b) Power spectral sequence



of the convolution can be interpreted as a representation of the Dirichlet kernel in the complex frequency plane. The representation of an analytic sequence using this kernel can be used for sampling frequency conversion. Sampling frequency conversion can also be performed in the complex frequency plane. Zeros are added to (removed from) the spectral zeros in the frequency region for an analytic time-

sequence so that the number of samples might be adjusted to a different up-sampled (down-sampled) frequency.

## References

1. D.C. Lay, *Linear Algebra and Its Applications* (Addison-Wesley, 1994)
2. A. Zygmund, *Trigonometric Series* (The University Press, Cambridge, 1988)
3. R. Adams, T. Kwan, Theory and VLSI architectures for asynchronous sample-rate converters. *J. Audio Eng. Soc.* **41**(7/8), 539–555 (1993)
4. M. Tohyama, T. Koike, *Fundamentals of Acoustic Signal Processing* (Academic Press, 1998)
5. R.E. Crochiere, L. Rabiner, Interpolation and decimation of digital signals—a tutorial review. *Proc. IEEE* **69**(3), 300–331 (1981)
6. W.M. Hartmann, *Signals, Sound, and Sensation* (Springer, 1997)

# Chapter 8

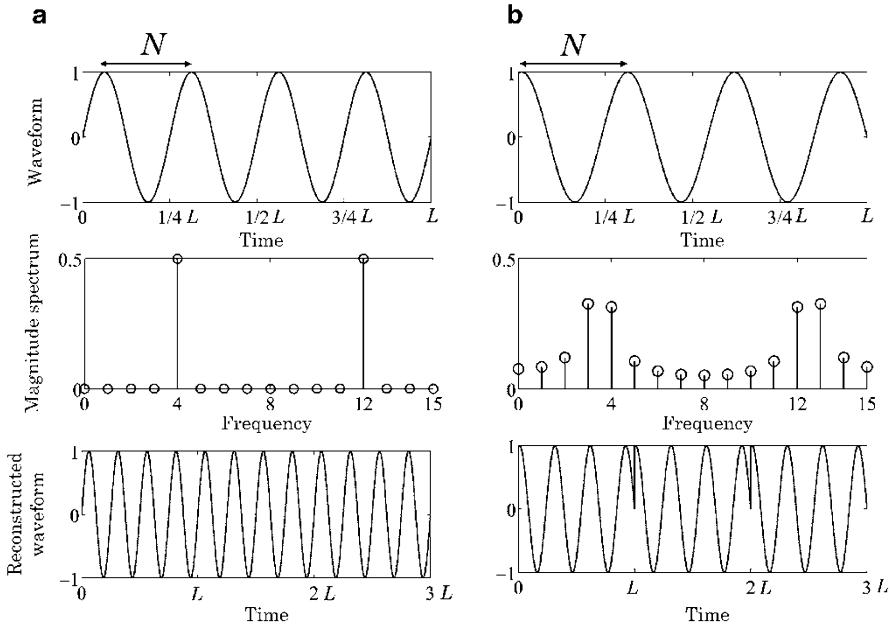
## Sinusoidal Representation of Sequence

This chapter describes spectral peak selection used in estimating the true spectrum of the target sequence, which is assumed to be composed of sinusoidal sequences. A compound sinusoidal sequence can be identified by repeating spectral peak selection from the interpolated spectral sequence iteratively independent of the observation length, provided the sinusoidal components are time independent. The frame-wise approach of spectral peak selection can also be applied to the estimation of fundamental frequencies of music tones. This chapter also describes clustered line-spectral modeling (CLSM) formulated to estimate the frequency components of a sequence composed of densely clustered sinusoidal components. CLSM is one approach to represent a sequence the envelope of which carries the signal signature. A finite-approximation to the Weierstrass function, which is an example of a harmonic sinusoidal sequence on the logarithmic frequency scale, is an interesting example to apply spectral peak selection. Spectral peak selection is also one approach to perform extrapolation (or prediction) of the target sequence outside its observation range.

### 8.1 Spectral Peak Selection

#### 8.1.1 *Representation of Sinusoidal Sequence by Discrete Fourier Transform*

The discrete Fourier transform (DFT) is a fundamental tool in representing periodic signals. A sinusoidal sequence is the fundamental periodic signal having just one single frequency. What happens if the resulting DFT sequence of length  $L$  is considered a sinusoidal sequence? Naturally, the DFT uniquely estimates the fundamental frequency; however, the result is not so straightforward. Figure 8.1 on the following page gives the spectral sequences of two periodic waves with



**Fig. 8.1** Examples of spectral analysis of sinusoidal sequences by DFT; **(a, left column)** DFT length  $L$  equals  $mN$  where  $N$  denotes the period of the sinusoidal sequence and  $m$  is a positive integer; **(b, right column)**  $L \neq mN$ ; *top row*: periodic sequences, *middle row*: magnitude spectral sequences with positive and negative frequency components obtained using the DFT, *bottom row*: periodic sequences reconstructed using the DFT with interpolated curve (*solid line*)

period  $N$ . The top two rows depict the periodic sequences, and their magnitude spectral sequences with positive and negative frequency components. On the left, the DFT length  $L$  is equal to  $mN$ , whereas for the right  $L \neq mN$ . Here  $m$  is a positive integer. In the bottom row, note that the periodic sequences, which were reconstructed from the DFT, are different.

Understanding the results on the left is straightforward. Specifically, the pair of DFT sequences (from the DFT and inverse DFT) yields the sinusoidal sequence. In contrast, the right gives a sinusoidal sequence that can no longer be reconstructed from the DFT sequence. This is, nevertheless, a quite natural consequence from a mathematical point of view. The DFT pair generates a periodic sequence; however, its period  $L$  is not always equal to  $mN$  of the sinusoidal sequence of interest. Therefore, the reconstructed sequence of period  $L$  does not look like it is sampled from the sinusoidal curve when the sequence is periodically repeated. Actually, the solid curve at the bottom right depicts the interpolation of the reconstructed sequence. The reconstructed curve is different from the original sinusoidal curve.

This simple example suggests that frequency analysis is not an easy task. These results are intuitively understood by recalling examples of Fourier series expansions for sinusoidal functions

$$\cos \mu x = \frac{2\mu \sin \mu \pi}{\pi} \left( \frac{1}{2\mu^2} - \frac{\cos x}{\mu^2 - 1} + \frac{\cos 2x}{\mu^2 - 2^2} - \dots \right) \quad (8.1)$$

$$\sin \mu x = -\frac{2\mu \sin \mu \pi}{\pi} \left( -\frac{\sin x}{\mu^2 - 1} - \frac{2 \sin 2x}{\mu^2 - 2^2} + \frac{3 \sin 3x}{\mu^2 - 3^2} - \dots \right) \quad (8.2)$$

### 8.1.2 Spectral Estimation of Sinusoidal Sequences by Spectral Interpolation

The DFT is a mathematical tool to determine a signal's fundamental and harmonic frequencies. However, note that the fundamental determined by the DFT, while independent of the signal signature, is actually dependent on the length of the window. This means that if the fundamental is to be observed it cannot be estimated using the DFT, except for when the length of the sequence  $L$  taken for observation (i.e., the window length) is  $L = mN$ ; here  $N$  denotes the fundamental period of the sequence to be estimated and  $m$  an integer. Therefore, from a practical point of view, a spectral analysis of a sequence, if indeed valid, seems almost impossible using the DFT.

Symbolically, the relationship between the original and observed spectra is

$$X_0 * W(k) = X(k) \quad (8.3)$$

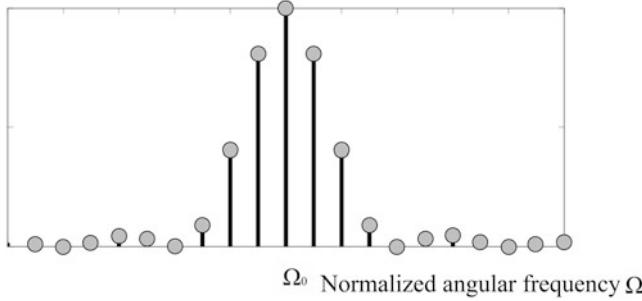
where  $X_0(k)$  denotes the original spectrum for the original sequence  $x_0(n)$ ,  $W(k)$  the spectrum of the window function used for the DFT,  $X(k)$  the observed spectrum, and  $x(n) = x_0(n) \cdot w(n)$ .

Suppose the sinusoidal sequence has the analytic form

$$x_0(n) = e^{j\Omega_0 n}. \quad (8.4)$$

By applying a rectangular window function of length  $N$  to this sequence and taking the Fourier transform

$$\begin{aligned} X(e^{-j\Omega}) &= \frac{1}{N} \sum_{n=0}^{N-1} e^{j\Omega_0 n} e^{-j\Omega n} \\ &= X_0 * W(e^{-j\Omega}) \end{aligned} \quad (8.5)$$



**Fig. 8.2** Windowed sinusoidal spectrum (power spectrum) from Fig. 14.12 in [1]

where

$$X(e^{-i\Omega}) = \begin{cases} \frac{1}{N} \frac{1 - e^{-i(\Omega - \Omega_0)N}}{1 - e^{-i(\Omega - \Omega_0)}} & (\Omega_0 \neq \Omega) \\ 1 & (\Omega_0 = \Omega) \end{cases} \quad (8.6)$$

and

$$X_0(e^{-i\Omega}) = \delta(\Omega - \Omega_0) \quad 0 \leq \Omega < 2\pi \quad (8.7)$$

$$W(e^{-i\Omega}) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-i\Omega n} = \frac{1}{N} \frac{1 - e^{-i\Omega N}}{1 - e^{-i\Omega}} \quad (8.8)$$

$$X_0 * W(e^{-i\Omega}) = \int_0^{2\pi} \delta(\Omega' - \Omega_0) \frac{1}{N} \frac{1 - e^{-i(\Omega - \Omega')N}}{1 - e^{-i(\Omega - \Omega')}} d\Omega'. \quad (8.9)$$

Figure 8.2 presents examples of the power spectra calculated from the Fourier transform  $X(e^{-i\Omega})$ . Spectral components at frequencies other than  $\Omega = \Omega_0$  can be seen from the DFT despite the original sequence being only a sinusoid with a single frequency. If the original angular frequency is  $\Omega_0 = 2\pi k/N$  ( $0 < k \leq N - 1$ ), where  $k$  and  $N$  are integers, then the value for the Fourier transformed sequence observed at  $\Omega = 2\pi l/N$  for an integer  $l$  ( $0 < l \leq N - 1$ ) becomes

$$X(e^{-i\Omega})|_{\Omega=\frac{2\pi}{N}l} = \begin{cases} 1, & l = k, \\ 0, & l \neq k. \end{cases}$$

That is, the original spectrum for the single sinusoid can be seen. The original spectrum, i.e., the true spectrum, of a finite-length record can be estimated from the spectral peak in the interpolated spectra from the DFT of the windowed sequence [1,2].

Spectral interpolation is an alternative to DFT sequence analysis. The sequence's original spectrum, which is called the true spectrum of the finite-length (finite window length) record, can be estimated using spectral peak selection from the interpolated spectra [1, 2]. Here, the original spectrum refers to the spectrum of a virtual sequence that could be obtained if a record of infinite length could be taken. Consider a target sequence expressed in the analytic form

$$x(n) = Ae^{j\frac{2\pi}{N}(k_0 + \Delta k)n} \quad (8.10)$$

where  $A$  denotes the complex magnitude including the initial phase,  $k_0, p, q$  are integers, and  $\Delta k = q/p$ . Spectral interpolation can be performed by taking the DFT of the sequence after amending the original sequence record by adding trailing zeros making its window length longer. With the record length increased to  $M$  by appending  $M - N$  zeros, the interpolated DFT sequence then becomes

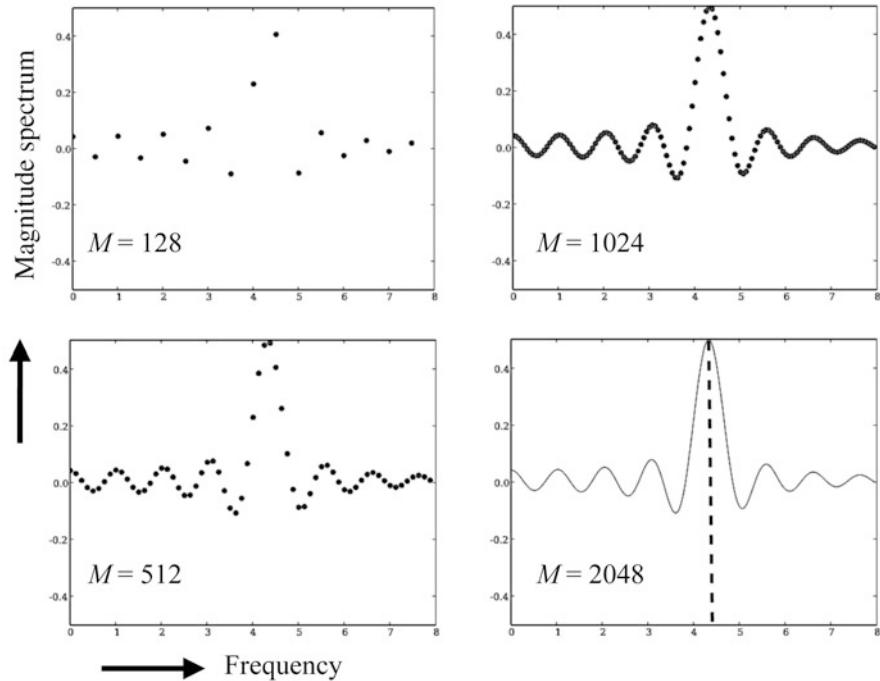
$$\begin{aligned} X(k) &= \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi k}{M}n} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} A e^{j2\pi n \frac{pk_0+q}{N \cdot p} - j2\pi n \frac{k}{M}} \\ &= \frac{A}{N} \sum_{n=0}^{N-1} e^{-j\frac{2\pi n}{M}(k - (pk_0 + q))} \end{aligned} \quad (8.11)$$

where  $M = N \cdot p$ . The original spectrum can be seen at the spectral peak where  $k = pk_0 + q$ .

Figure 8.3 on the following page confirms this spectral component of a single sinusoidal sequence estimated by spectral interpolation [1]. Specifically, the original (or true) spectrum, can be estimated by spectral peak selection from the interpolated DFT spectra [2]. That is, the spectral component of a sinusoidal sequence is well estimated by spectral interpolation, even if the observation interval is shorter than the period of the sinusoidal sequence.

### 8.1.3 Spectral Estimation of Compound Sequence by Spectral Peak Selection

Spectral estimation by interpolation can be extended to spectral analysis for compound sequences by spectral peak selection [1]. Consider now a target sequence expressed in the analytic form



**Fig. 8.3** Spectral estimation of a real sinusoidal sequence by spectral interpolation  $N \rightarrow M$ ;  $N = 64$  denotes the observation interval

$$x(n) = \sum_{k=1}^K A(k) e^{i 2\pi v(k) n} + \epsilon_K(n) \quad (8.12)$$

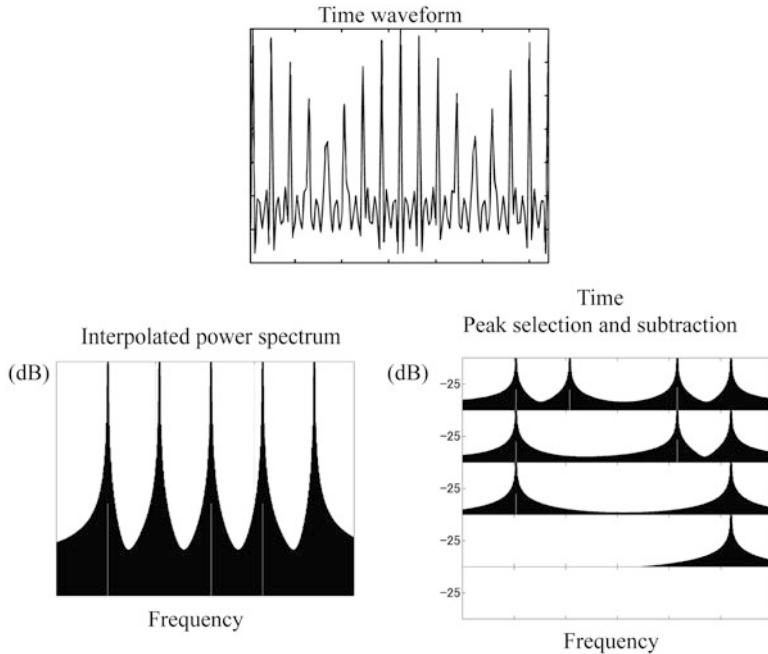
where  $A(k)$  and  $v(k)$  denote the  $k$ -th sinusoidal component's complex magnitude and frequency, respectively,  $K$  the number of dominant sinusoidal components, and  $\epsilon_K$  the residual component such as external noise. Figure 8.4 on the facing page is an example showing that the original spectrum, i.e., the true spectrum, of a finite-length record can be estimated from the spectral peak in the interpolated spectra.

Several spectral peaks corresponding to the dominant sinusoidal components are seen in the figure. The following procedure is used for spectral peak selection.

Step 1: After adding  $M - N$  trailing zeros, take the  $M$ -point DFT of the sequence in analytic form so that

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-i \frac{2\pi k}{M} n} \quad (8.13)$$

where the length of the record is  $N$ .



**Fig. 8.4** Spectral estimation of a harmonic sequence composed of five sinusoidal components (taken from Fig. 14.14 in [1]);  $N = 64$  denotes the observation interval

Step 2: Select the maximum component in the power spectrum record obtained in Step 1; that is,  $X(k_p)$  for which  $|X(k_p)|^2$  is maximum.

Step 3: Subtract the maximum component from the original sequence so that

$$e(n) = x(n) - X(k_p)e^{j\frac{2\pi}{M}k_p n} \quad n = 0, 1, \dots, N-1 \quad (8.14)$$

and set

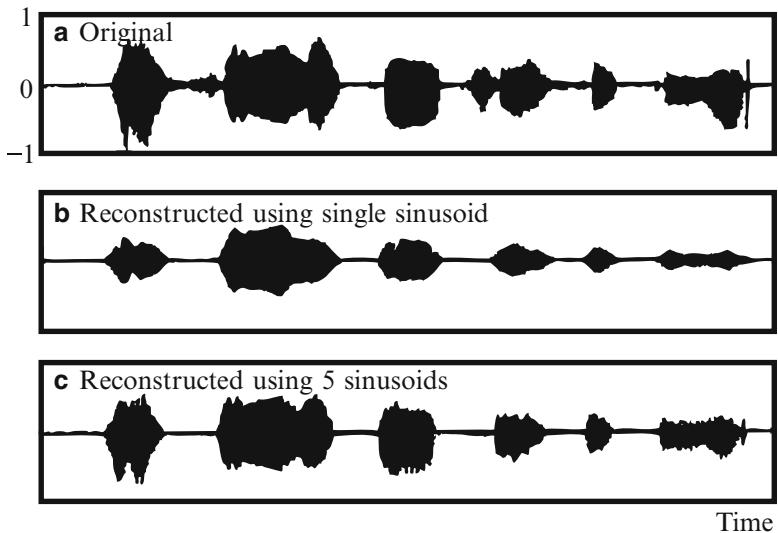
$$x(n) \leftarrow e(n) \quad n = 0, 1, \dots, N-1. \quad (8.15)$$

Step 4: Repeat steps 1–3 until

$$\sum_{n=0}^{N-1} |e(n)|^2 < E \quad (8.16)$$

where  $E$  is the allowable error.

The DFT sequence in Step 1 determines an estimate of the dominant frequency of the spectral peak that corresponds to the maximum power spectral frequency.



**Fig. 8.5** Original and reconstructed waveforms from [3] (Fig. 2) and Fig. 14.15 of [1]. (a) Original. (b) Reconstructed using single sinusoid. (c) Reconstructed using five sinusoids

The subtraction in Step 3, performed on the record of the sequence of length  $N$  excluding the added zeros, is crucial in representing the sequence without the leakage (spurious) spectrum caused by the truncation window. However, it is also possible to perform the subtraction process in the frequency domain by subtracting the spectrum of the windowed sinusoid.

Figure 8.4 on the previous page shows results obtained after five iterations for the sequence waveform plotted in Fig. 8.4 on the previous page (a) where only the real part is shown for the complex sequence. Note that the windowed spectrum, including the leakage spectrum, is removed from around every spectral peak after each subtraction.

Speech waveforms can also be represented by the spectral peak selection [3]. Figure 8.5 gives an example of a speech waveform reconstruction using only the spectral peak components. The rectangular window length for the DFT analysis is 512 samples corresponding to 32 ms, and a triangular window was used for reconstruction. Each frame starts with the last 256 data points of the previous frame to avoid discontinuities between successive frames. The envelope of the entire waveform for all frequency bands can be reconstructed mostly on a frame-by-frame basis by selecting the maximal spectral component in every frame, subject to an adequate frame length.

Note, however, the waveform reconstructed by selecting only the maximal component is no longer intelligible. This fact also implies that it is necessary to recover the frequency-band-dependent (or narrow-band) envelopes to obtain an intelligible speech waveform, e.g., every 1/4 octave bands, instead of the envelope of the entire waveform [4–6]. Although the reconstructed waveform looks intelligible,

it no longer sounds intelligible. In contrast, the waveform shown in Fig. 8.5 on the preceding page (c), which was reconstructed using five maximal components in every frame, looks similar to the waveform of Fig. 8.5 on the preceding page (b), but surprisingly it sounds almost perfectly intelligible. The narrow-band envelopes were resynthesized using at least five maximal components in every frame provided the frame length is around 32 ms.

### 8.1.4 Harmonic Sequence on Logarithmic Frequency Scale

Weierstrass functions are known as examples of continuous functions that are nowhere differentiable [7]. A Weierstrass function is formally defined by

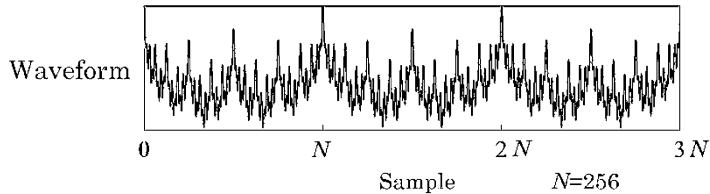
$$w(t) = \sum_{k=1}^{\infty} \alpha^k \cos(\beta^k t) \quad (8.17)$$

where  $\alpha$  is real and  $\beta$  is odd. If  $\alpha\beta > 1 + 3\pi/2$ ,  $w(t)$  is continuous but nowhere differential. Note that all frequencies of the sinusoidal components are equally spaced on the logarithmic frequency axis but not on the linear frequency scale. Hence,  $w(t)$  can be interpreted as a harmonic sequence with respect to the logarithmic frequency.

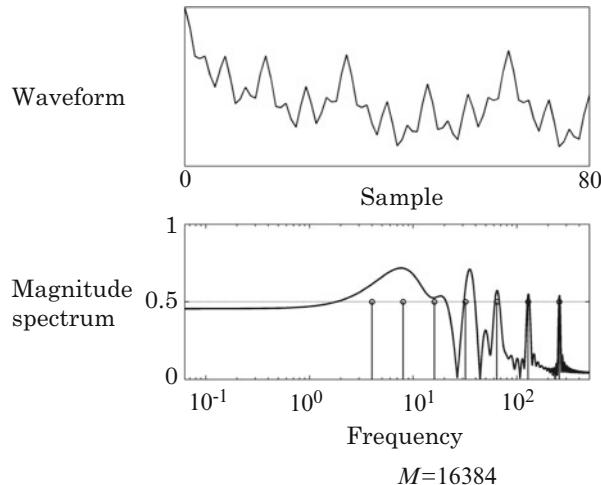
Figure 8.6 on the following page illustrates a finite-sum approximation to the Weierstrass function [7]

$$w_k(t) = \sum_{k=2}^K \cos(\beta^k t) \quad (8.18)$$

where  $\alpha = 1$ ,  $K = 8$ , and  $\beta = 2$ . After taking the DFT of the finite-sum sequence, Figure 8.7 on the following page shows the power spectral sequence that is obtained using  $N$ -sampled data points followed by a spectral interpolation from  $N \rightarrow M$ . Panel (a) presents the sampled time sequence, and panel (b) gives the magnitude spectral sequence in the positive frequency domain where the horizontal axis represents the logarithmic frequency scale. Spectral peaks for the higher frequency components, located at equally separated intervals on the logarithmic frequency scale, are evident. The non-overlapping spectral components indicate that sinusoidal components can be estimated by spectral peak selection on the logarithmic frequency scale; however, lower frequency components of the peaks are no longer well separated but overlap with each other because of effects from the length of the observation window.



**Fig. 8.6** Finite-sum approximation to Weierstrass function where  $N$  denotes period,  $K = 8$ , and  $\beta = 2$  from Fig. 12 in [7]

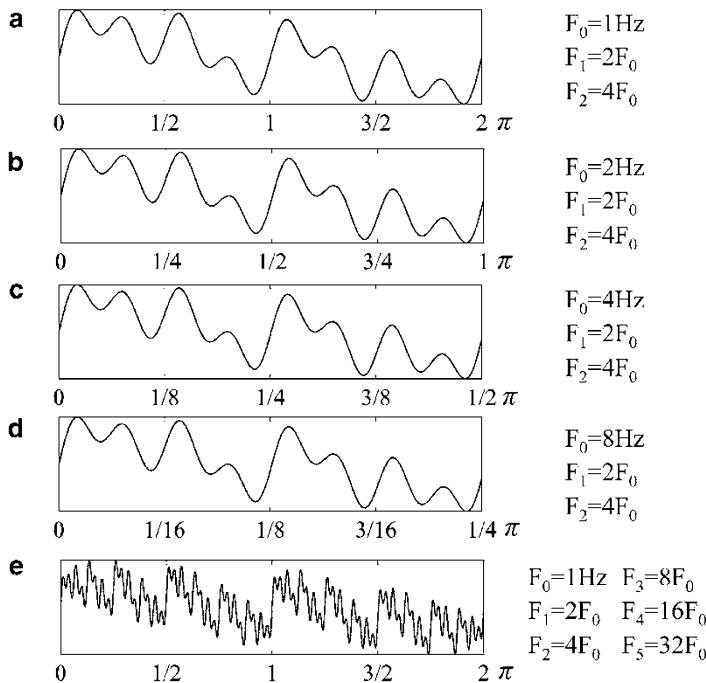


**Fig. 8.7** Spectral estimation through spectral peak selection on a logarithmic frequency scale; (a) time sequence sampled from Fig. 8.6, (b) magnitude spectral sequence obtained after spectral interpolation

### 8.1.5 Scaling Structure of Harmonic Spectral Function on Logarithmic Frequency Scale

The human ear is often modeled as a filter bank in which the center frequencies (or the band-widths of sub-band filters) are equally spaced over the logarithmic frequency scale [8]. If there is a prominent line-spectral component in every sub-band, then a very complicated sequence such as a Weierstrass function can be expected.

The harmonic structure on the logarithmic frequency scale is also called a scaling structure [7]. Suppose a sequence composed of harmonic sinusoidal components on the logarithmic frequency scale. The bottom row in Fig. 8.8 on the facing page illustrates the sum of six sinusoidal components of frequencies 1, 2, 4, 8, 16, and 32. The top row shows the sum of the first three sinusoidal components corresponding to frequencies 1, 2, and 4. Similarly, the second shows that for frequencies 2, 4, and

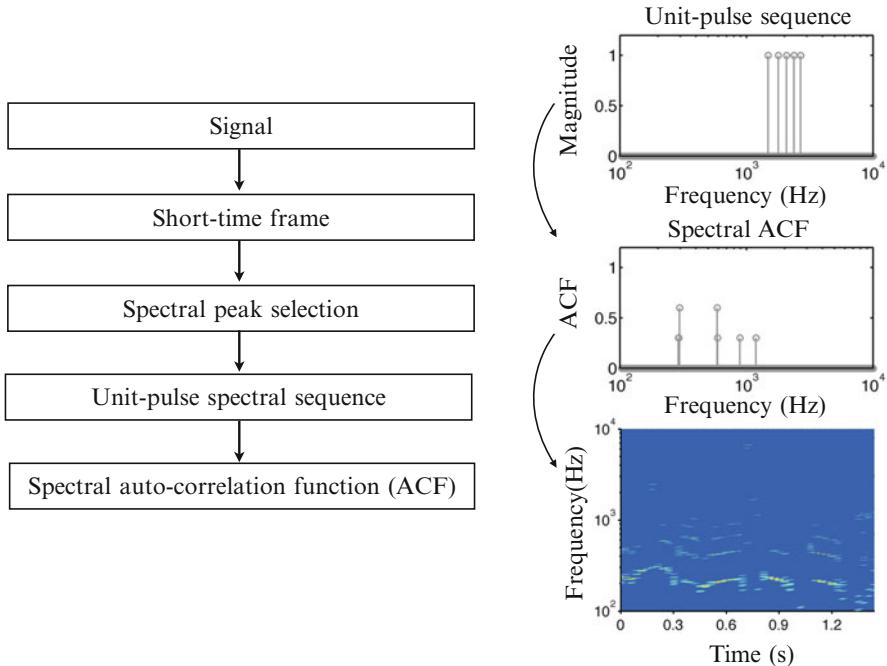


**Fig. 8.8** Scaling structure of a partial sum of harmonic sequence on logarithmic frequency scale

8. There are no differences in the two figures, although the interval of the second one is just one half of the first. The top two rows can be interpreted as amplifications of the figure in the bottom row. The sampling frequency associated with the second is 2 times higher than that for the first. The other figures, the third and fourth rows, are similarly interpreted. A scaling function or sequence, in principle, has identical waveforms, even if amplified views are taken from the original waveform. The scaling structure implies that the signal signature of the waveform is hidden in a short interval, even if a very short interval is chosen and magnified.

### 8.1.6 Auto-correlation Analysis of Musical Tones in the Frequency Domain

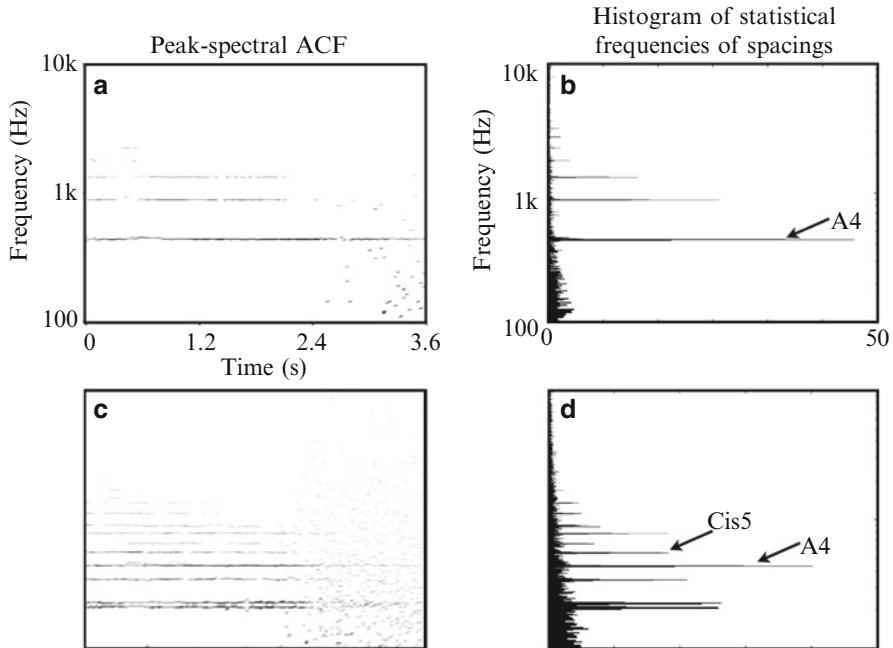
Periodic structures, each composed of a fundamental and its harmonics, reflect the resonant mechanism of the sound production process. Therefore, resonant frequencies including higher harmonics are important signatures that characterize the sound, as are the power spectral properties. Such resonant frequencies can be estimated by spectral peak selection [9]. Figure 8.9 on the following page illustrates the method of estimating the fundamental and its harmonics in the frequency domain



**Fig. 8.9** Frame-wise spectral auto-correlation analysis from Fig. 14.16 in [1]

instead of the time domain using spectral-peak selection on a frame-by-frame basis and auto-correlation analysis [1, 9]. A spectral sequence composed of unit pulses is obtained using spectral peak selection by discarding the magnitude information. If the auto-correlation analysis is performed on the unit spectral sequence along the frequency axis, a histogram of the frequency spacing for the dominant peaks can be derived. Normalization of the power spectral peaks is important so that power spectral effects other than the harmonic structure can be avoided. Hence, the periodicity of the harmonic structure, which is closely related to the fundamental frequency, can be well estimated by the auto-correlation analysis in regard to the unit-pulse-like sequence composed of the frequencies of the spectral peaks.

Figure 8.10 on the facing page shows samples of an auto-correlation analysis of musical tones without their fundamentals on the frequency axis following the procedure described in Fig. 8.9 [9]. Despite missing fundamentals, the fundamental frequencies and the harmonics can be estimated from the histograms, which are obtained by auto-correlation analysis for the selected spectral unit-pulse sequence in the frequency domain. All of the figures were obtained in a frame-by-frame manner every 30 ms using six spectral peaks selected in each frame. On the left, the frame-wise auto-correlation sequences are taken along the frequency axis with regard to the normalized power spectral unit-pulse sequence selected from the short-term (frame-wise) power spectral sequence. The horizontal axis represents the time corresponding to the order of the observation frames.



**Fig. 8.10** Samples of frame-wise auto-correlation for piano tones (taken from Fig. 14.17 in [1] and [9]); left column: frame-wise auto-correlation sequences obtained using the procedure given in Fig. 8.9 on the preceding page, right column: averaged auto-correlation sequences for all frames, upper row: for single note A4, lower row: for compound tone A4 and Cis5

In contrast, the right column presents averages over all frames. The average of the auto-correlation sequence for a unit-pulse sequence can be interpreted as the histogram of statistical frequencies of spacings between the spectral peaks. The upper row gives a result for a single note, whereas the lower displays that for a compound tone of two notes. Panel (a) shows a harmonic structure with missing fundamental in the frequency domain; however, the fundamental frequency corresponding to tone A4 can be read from both of the frame-dependent auto-correlation sequences [panel (a)] and their average [panel (b)]. Similarly, panel (c) shows the result of two tones composed of A4 and Cis5 without the fundamental frequency. Interestingly, frequencies can even be seen one octave (or even two octaves) below that for A4; [see panel (a)]. The existence of this component two octaves lower can be interpreted from a mathematical point of view, by recalling that the ratio of the fundamental frequencies of A4 and Cis5 is 4/5. However, such a low frequency seems unlikely as a candidate frequency of the fundamental from a perceptual point of view. The most probable fundamental frequency, that of A4, is that which is determined by the mode in panel (d) showing the frame-averaged auto-correlation sequences.

Like the auto-correlation analysis in the time domain, the auto-correlation analysis in the frequency domain is useful, as demonstrated by the figures, in estimating the spectral harmonic structures of time sequences and in identifying their fundamental frequencies. However, the spectral auto-correlation analysis in the frequency domain does not work well in identifying the fundamental frequencies, particularly when a harmonic structure contains vacant harmonics; for example, structures containing no even-(or odd-)numbered harmonic components. Nevertheless, harmonic structure analysis using spectral auto-correlation in the frequency domain might be promising in practical situations, because vacancies for all even (or odd) numbered harmonics are quite unlikely in general.

## 8.2 Clustered Line Spectral Modeling

### 8.2.1 *Windowed Sinusoidal Function and Its Spectral Sequence*

The spectral components might be estimated by spectral peak selection. However, this does not work well for a pair of adjacent sinusoidal components for which the frequencies are closely spaced [1].

Consider a compound sinusoidal sequence composed of two sinusoidal frequencies

$$x(n) = A_1 e^{j \frac{2\pi}{N} (k_0 + \Delta k_1)n} + A_2 e^{j \frac{2\pi}{N} (k_0 + \Delta k_2)n} = x_1(n) + x_2(n) \quad (8.19)$$

where  $k_0, p, q_1, q_2$  are all positive integers, with  $\Delta k_1 = q_1/p$  and  $\Delta k_2 = q_2/p$ . Figure 8.11 on the facing page shows an example of such a waveform. When the observation window length is long enough, the two spectral peaks are seen well separated (Fig. 8.11 on the facing page). In contrast, if the window length becomes too short, those spectral peaks are no longer separated (Fig. 8.12 on the facing page [1]).

By taking the Fourier transform for the compound sequence, the transforms give

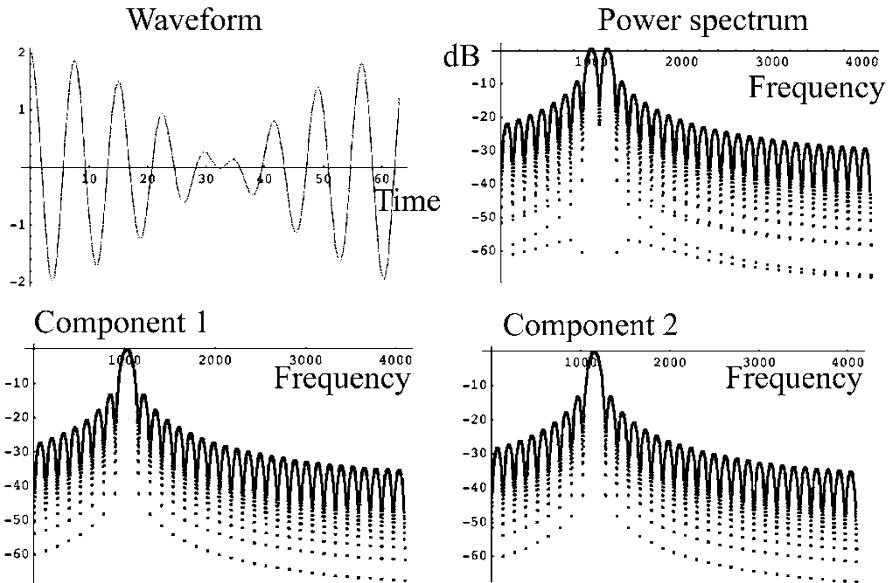
$$X(k_1) = X_1(k_1) + W(k_1 - k_2)X_2(k_2) \quad (8.20)$$

$$X(k_2) = W(k_2 - k_1)X_1(k_1) + X_2(k_2) \quad (8.21)$$

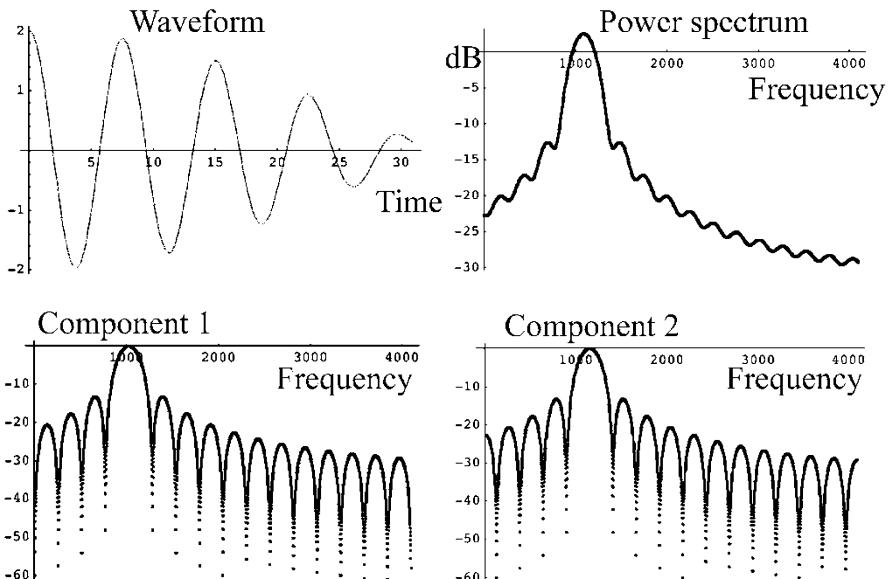
at two sufficiently interpolated frequency bins in the discrete form; here

$$X(k_1) = \left. \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi k}{M} n} \right|_{k=k_1} = X(k)|_{k=k_1} \quad (8.22)$$

$$X(k_2) = X(k)|_{k=k_2} \quad (8.23)$$

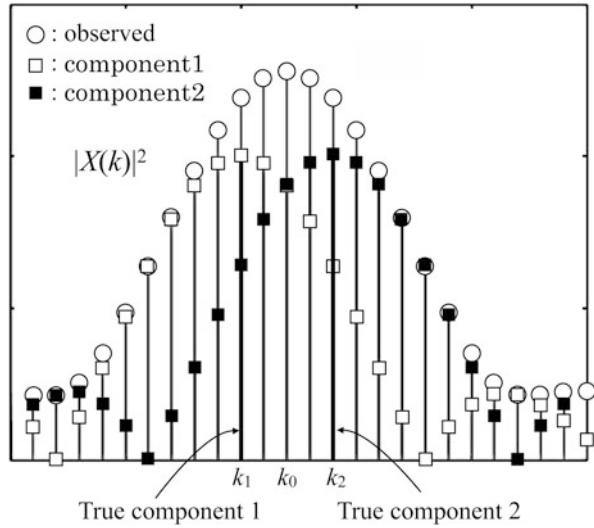


**Fig. 8.11** Separated peaks are observed if the window length is sufficiently long from Fig. 14.21 in [1]



**Fig. 8.12** Spectral properties for two truncated sinusoids with nearly the same frequencies obtained when window length is too short to distinguish the two frequencies from Fig. 14.22 in [1]

**Fig. 8.13** Overlap of the leakage spectra of two clustered sinusoids truncated by rectangular window from Fig. 14.23 in [1]



$$W(k_2 - k_1) = \frac{1}{N} \sum_{n=0}^{N-1} w(n) e^{-i \frac{2\pi k}{M} n} \Big|_{k=k_2-k_1} \quad (8.24)$$

$$M = N \cdot p \quad (8.25)$$

and  $w(n)$  denotes the window function applied to the sequence. This spectral representation can be interpreted graphically (Fig. 8.13[1]). Specifically, the spectrum of the compound sequence is produced by the overlapped leakage spectral components. Consequently, if a record of the spectrum is observed for at least two frequency bins, the spectral magnitude and phase for the two components can be estimated by solving the simultaneous equations on the frequency domain because the leakage spectral characteristics are determined according to the window function [2].

### 8.2.2 Formulation of CLSM

Consider a sequence of record length  $N$  and its interpolated spectrum obtained using an  $M$ -point DFT after adding trailing zeros. Assume that the sequence is composed of  $K$  clustered sinusoidal components around the peak  $k = k_p$  such that

$$x(n) = \sum_{k=1}^K A(k) e^{i 2\pi v(k) n} + \epsilon_K(n) \quad (8.26)$$

where  $A(k)$ , and  $\nu(k)$  denote the complex magnitude and frequency for the  $k$ -th frequency bin, respectively, and  $\epsilon_K(n)$  is the residual noise or modeling error. To represent the sequence by  $P$  sinusoidal components clustered between  $k = k_{p-m}$  and  $k = k_{p-m+P-1}$ , these clustered components can be estimated to within least-squares-error (LSE) criterion using a set of linear equations established as spectral observations at  $L$  frequency points between  $k_{p-l}$  and  $k_{p-l+L-1}$ ; that is [2]:

$$\mathbf{x}_{\text{observe}} = W \mathbf{x}_{\text{signal}} \quad (8.27)$$

where

$$\begin{pmatrix} X(k_{p-l}) \\ \vdots \\ X(k_{p-l+L-1}) \end{pmatrix} = \mathbf{x}_{\text{observe}} \quad (8.28)$$

denotes the spectrum observed at  $L$  frequency points,

$$\begin{pmatrix} X_s(k_{p-m}) \\ \vdots \\ X_s(k_{p-m+P-1}) \end{pmatrix} = \mathbf{x}_{\text{signal}} \quad (8.29)$$

denotes the  $P$  spectral components for the sequence where  $L > P$  and  $l > m$ , and

$$m = \begin{cases} \frac{P-1}{2} & P : \text{odd}, \quad l = \frac{L-1}{2}, \quad L : \text{odd}, \\ \frac{P}{2} & P : \text{even}, \quad l = \frac{L}{2}, \quad L : \text{even}. \end{cases} \quad (8.30)$$

The matrix  $W$  is given by

$$W = \begin{pmatrix} W_{NM}(k_{p-l} - k_{p-m}) & \cdots & W_{NM}(k_{p-l} - k_{p-m+P-1}) \\ \vdots & \ddots & \vdots \\ W_{NM}(k_{p-l+L-1} - k_{p-m}) & \cdots & W_{NM}(k_{p-l+L-1} - k_{p-m+P-1}) \end{pmatrix} \quad (8.31)$$

where

$$W_{NM}(q) = \left. \frac{1}{N} \sum_{n=0}^{N-1} w(n) e^{-j \frac{2\pi k n}{M}} \right|_{k=q} \quad (8.32)$$

for the window function  $w(n)$ . The spectral components of the signal can be estimated by finding the LSE solutions [10]

$$\hat{\mathbf{x}}_{\text{signal}} = (W^T W)^{-1} W^T \mathbf{x}_{\text{observe}}. \quad (8.33)$$

### 8.2.3 LSE Solution of the Simultaneous Equations

The LSE solution of the simultaneous equations is developed as follows [10]. Given a set of simultaneous linear equations

$$Ax = \mathbf{b} \quad (8.34)$$

a set of unique solutions exists if the matrix  $A$  is square and is composed of independent column vectors. When the matrix  $A$  is rectangular ( $N$  rows and  $M$  columns) and  $N < M$ , solutions exist but are not unique. This is because the set of equations is incomplete, i.e., their number  $N$  is smaller than the number of unknowns [10, 11]. In contrast, if  $N > M$ , no solution exists, but LSE solutions are obtained instead [10]. This LSE solution  $\hat{\mathbf{x}}$  minimizes the squared error that is defined by

$$|\mathbf{e}|^2 = |\mathbf{b} - A\hat{\mathbf{x}}|^2 \quad (8.35)$$

where  $|\mathbf{e}|^2$  denotes the square norm of a vector. That is, the LSE solution  $\hat{\mathbf{x}}$  solves the equation

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \quad (8.36)$$

where  $\mathbf{b} = \hat{\mathbf{b}} + \mathbf{e}$ , instead of the equation  $Ax = \mathbf{b}$ .

The linear equation  $Ax = \mathbf{b}$  can be written as a linear combination of the column vectors  $\mathbf{v}$  of the matrix  $A$ :

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \dots + x_M\mathbf{v}_M = \mathbf{b} \quad (8.37)$$

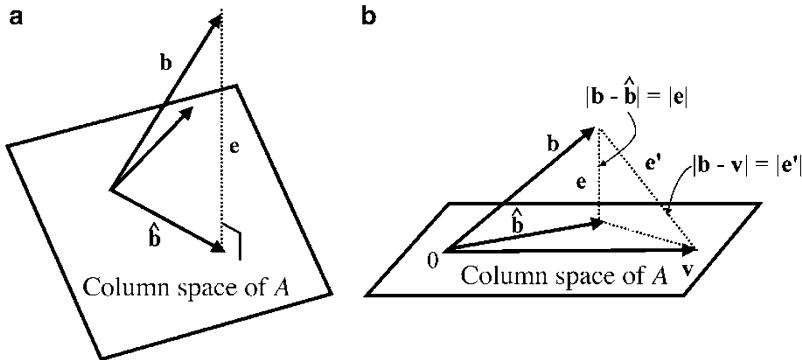
where

$$A = (\mathbf{v}_1 \quad \mathbf{v}_2 \cdots \mathbf{v}_M) \quad (8.38)$$

$$\mathbf{x} = (x_1 \quad x_2 \cdots x_M)^T \quad (8.39)$$

$$\mathbf{b} = (b_1 \quad b_2 \cdots b_N)^T. \quad (8.40)$$

When  $\mathbf{b}$  is a vector in the column space, the vector of unknowns  $\mathbf{x}$  gives the solution vector for the simultaneous equation. Here  $\mathbf{x}^T$  denotes taking the transpose of  $\mathbf{x}$ . In contrast, when the vector  $\mathbf{b}$  is not located in the column space, it cannot be expressed as a linear combination of column vectors. Figure 8.14 on the facing page shows the orthogonally projected vector  $\hat{\mathbf{b}}$  of the vector  $\mathbf{b}$  on the column space [12].



**Fig. 8.14** Orthogonal projection onto column space: (a) orthogonally projected vector; (b) orthogonally projected vector and the least-squares-error criterion from Fig. 14.24 in [1]

This projected vector meets the LSE criterion. Specifically, the squared norm of the error is minimized when  $\mathbf{e} \perp \hat{\mathbf{b}}$ . The LSE solution vector  $\mathbf{x}$  satisfies the linear equation  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ .

The orthogonal relationship  $\mathbf{e} \perp \hat{\mathbf{b}}$  can be rewritten using the inner product[11]

$$\hat{\mathbf{b}}^T \mathbf{e} = 0 \quad (8.41)$$

or equivalently

$$\mathbf{v}_1^T \mathbf{e} = \mathbf{v}_2^T \mathbf{e} = \dots \mathbf{v}_M^T \mathbf{e} = 0 \quad (8.42)$$

where  $\mathbf{e} = \mathbf{b} - \hat{\mathbf{b}}$ . Substituting  $\mathbf{e}$  into Eq. 8.42, a set of linear equations

$$A^T \hat{\mathbf{b}} = A^T \mathbf{b} \quad (8.43)$$

is obtained where  $N > M$ . Thus, recalling Eq. 8.36 on the preceding page

$$A^T A \hat{\mathbf{x}} = A^T \mathbf{b} \quad (8.44)$$

is obtained where the matrix  $A^T A$  is square and symmetric. When the square matrix is non-singular, the LSE solution is

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}. \quad (8.45)$$

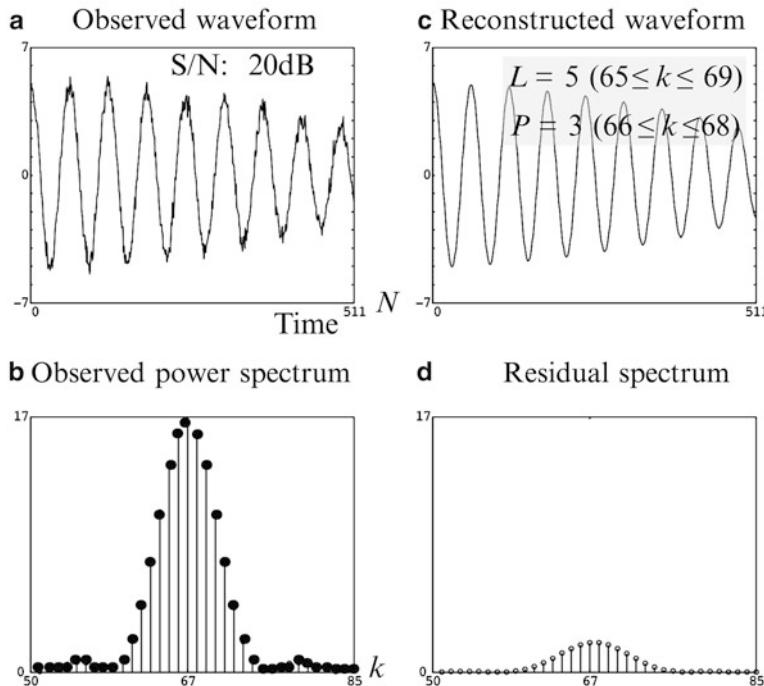
Finding the linear regression line is a typical method of generating the LSE solutions [12].

### 8.2.4 CLSM Examples

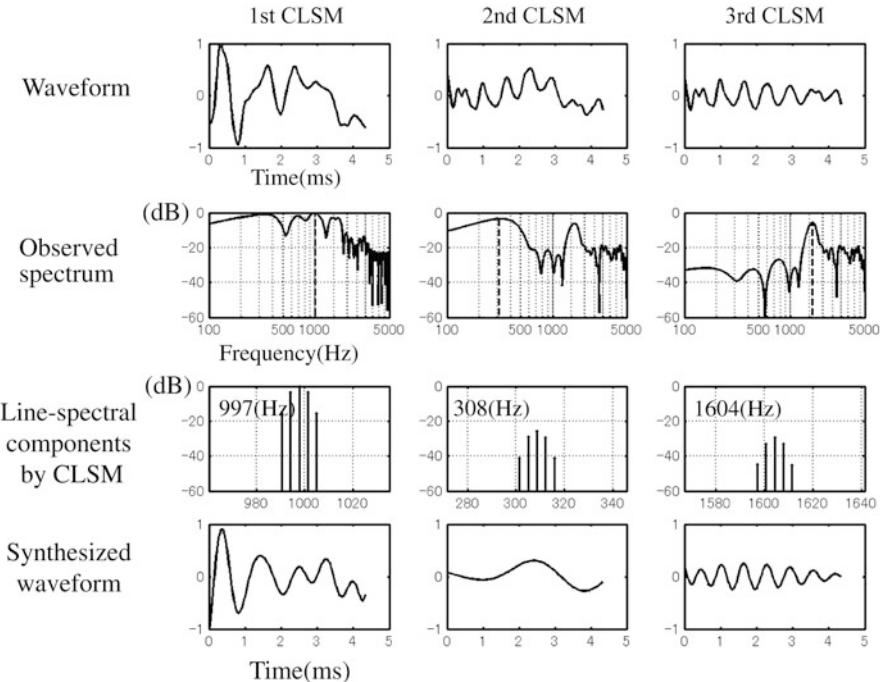
The clustered line spectral modeling, or CLSM for short, is developed into an iterative method for the sinusoidal representation of compound sequences based on the least-squares-error criterion in the frequency domain [1, 2]. Recalling that the time envelope of a sequence is characterized by closely located line spectral components, CLSM may be a good means to represent a temporal sequence using a finite number of sinusoidal components in short frames.

Figure 8.15 is an example of CLSM, where  $A_k = 1$ ,  $K = 5$ ,  $N = 512$ ,  $M = 4,096$ , and the signal-to-noise-ratio is 20 dB [2]. The waveform is reconstructed, where  $L=5$  between  $k = k_p - 2 = 65$  and  $k = k_p + 2 = 67$  for  $P = 3$  sinusoidal components between  $k = k_p - 1 = 66$  and  $k = k_p + 1 = 68$ . The CLSM approach can be iterated to accommodate multiple dominant spectral peaks.

CLSM is a means to represent the temporal envelope of a sequence that might be an informative temporal signature using spectral components for which the frequencies are closely located around the dominant spectral peaks. Figure 8.16



**Fig. 8.15** CLSM example for a compound signal  $x(n) = \sum_{k=1}^5 e^{j2\pi v(k)n} \cdot w(n)$ ,  $w(n) = 1$  ( $0 \leq n \leq N - 1$ ),  $N = 512$ , and  $v(k) = (8 + k/8)/512$  from Fig. 14.25 in [1] and Fig. 2 in [2]. **(a)** Observed waveform. **(b)** Observed power spectrum. **(c)** Reconstructed waveform. **(d)** Residual spectrum



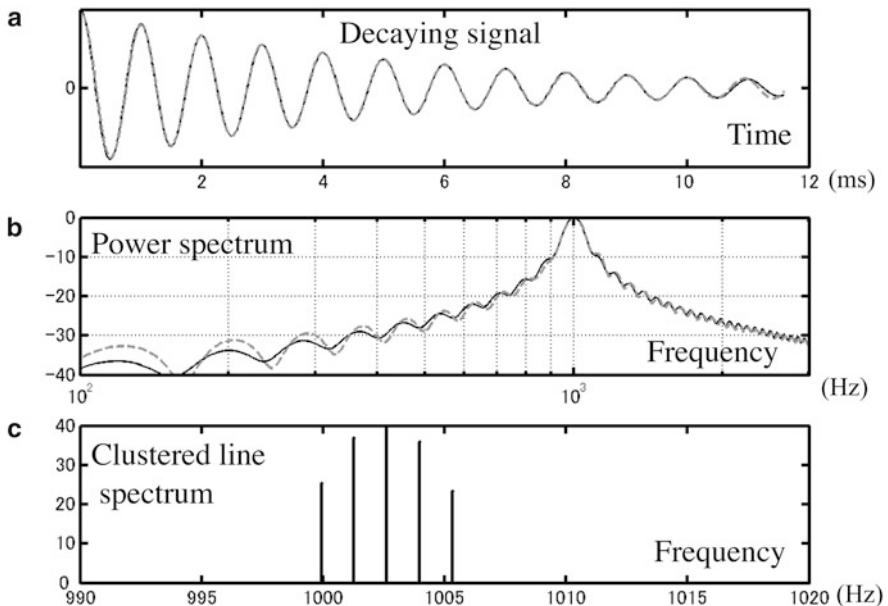
**Fig. 8.16** CLSM representation of spoken vowel ‘a’ from Fig. 14.26 in [1] and Fig. 3 in [13]

is an example for the spoken vowel ‘a’. The panels from top to bottom in the figure give the waveform to be analyzed, the power spectral components, the line-spectral components (magnitude) extracted by CLSM where  $P = 5$  and  $L = 7$ , and the synthesized waveforms by CLSM. The top-left panel shows a cycle of the waveform; the left panel second down panel shows the dominant spectral peak around 1 kHz. The line spectral components were obtained by applying CLSM to the dominant peak,(left panel third down). The subsequent synthesized waveform (bottom left panel) resembles the envelope of the entire waveform of the top left panel.

The top-center panel shows the residual component left after the first CLSM at around 1 kHz. By applying a second CLSM approach to the second dominant peak around 300 Hz, the synthesized waveform (bottom-center panel) was obtained. Subsequently, a third dominant spectral peak is left around 1,600 Hz (right panel second down).

The right panels display results for the third trial of the CLSM approach to the second residual. This dominant component can be represented by CLSM, and thus only a few residual components are left.

An exponentially decaying sinusoid might be expressed by CLSM. Suppose a decaying sinusoidal waveform representing the resonant impulse response of a single-degree-of-freedom system. Figure 8.17 on the following page is an example

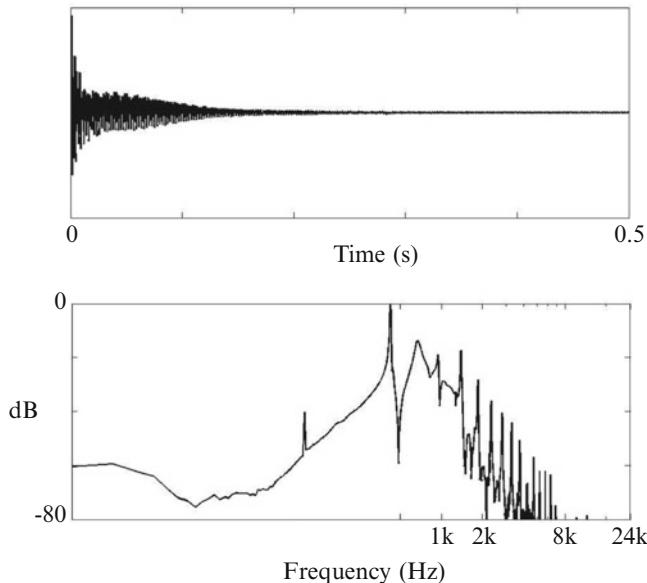


**Fig. 8.17** CLSM analysis for decaying signal from Fig. 14.27 in [1] and Fig. 1 in [13]

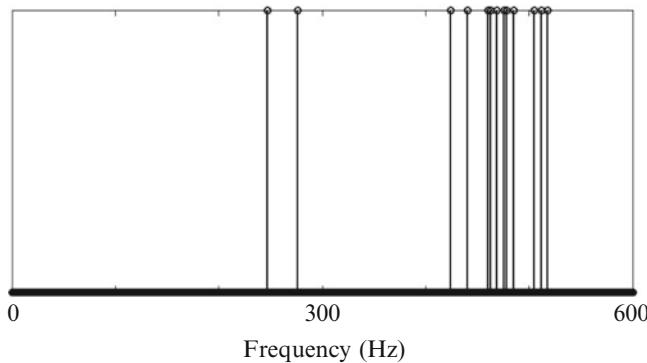
of the CLSM approach to the decaying signal. The top graph shows the waveform; in the center is the power spectrum of the waveform as well as that for the synthesized signal by CLSM. The decaying envelope can be represented by five clustered spectral components (bottom panel).

Note that the number of sinusoidal components constructing the dominant spectral peak in a target sequence is unknown. Nevertheless, the required number of components  $P$  for CLSM can be estimated in practice by repeating the CLSM process for the dominant peaks so that the residual energy of the sequence becomes arbitrarily small.

The vibrations of piano strings are typical examples of decaying and transient waveforms [14]. Figure 8.18 on the facing page (a) shows an amplified waveform up to 0.5 s after a single piano-key (A4) was hammered. Panel (b) illustrates the power spectral record of the vibration at a sampling frequency of 48 kHz. The fundamental and its first 15 harmonics are observed in the power spectral data, however, the intervals between adjacent spectral peaks are not equally separated (Fig. 8.19 on the facing page[14]). In addition, a decay curve of the piano-string vibration, which is composed of an initial quick decay process followed by one slower [15], does not take an exponentially decaying form (Fig. 8.20 on page 184). These results indicate not only the inharmonicity of the spectral peaks, but the temporally decaying envelope must be closely related to the clustered spectral components around every spectral peak.

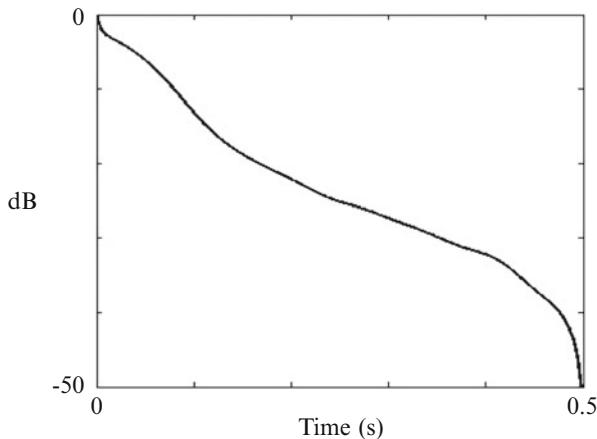


**Fig. 8.18** Time-waveform of piano-string vibration and its frequency-power spectrum under the sampling frequency of 48 kHz: (a) acceleration waveform and (b) power spectrum from Fig. 2 in [14]

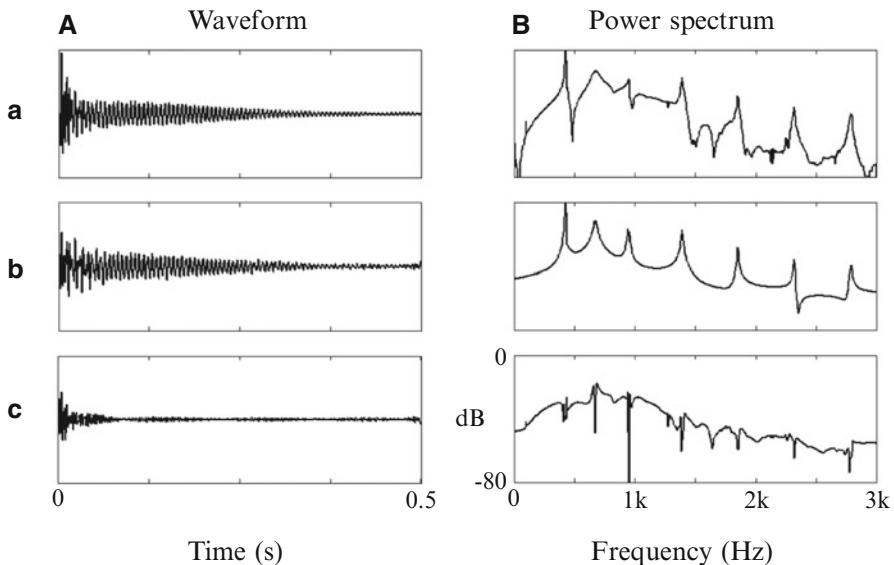


**Fig. 8.19** Intervals of the spectral components between adjacent spectral peaks from Fig. 3 in [14]

CLSM is applied to the 0.2 s down-sampled waveform and its spectrum (first row of Fig. 8.21 on the following page). The data are converted to a down-sampled version at 6 kHz sampling frequency. The whole record is composed of 1,200 observation data points followed by 3,600 zeros for spectral interpolation.

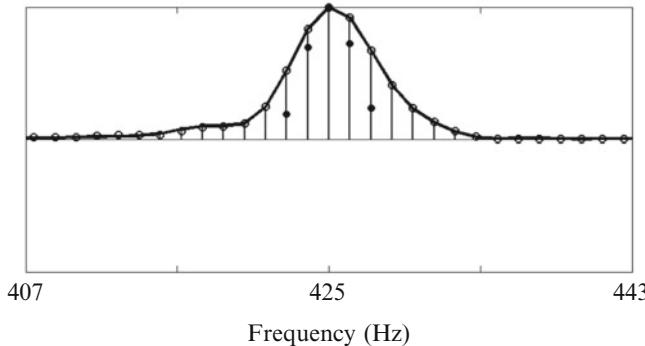


**Fig. 8.20** Decay curve of the piano-string vibration from Fig. 11.10 in [1] and Fig. 4 in [14]



**Fig. 8.21** Results of CLSM where (a) observed, (b) synthesized, and (c) residual (error) after CLSM from Fig. 5 in [14]. (A) Waveform; (B) power spectrum

Figure 8.22 on the facing page shows the CLSM analysis for the first spectral peak corresponding to the fundamental frequency. Here, five sinusoidal components are assumed and seven observation data are used in solving the CLSM simultaneous equations. By repeating the CLSM approach with seven dominant spectral components, Fig. 8.21 (b) and (c) are obtained. Figure 8.21 (b)-A and -B present the

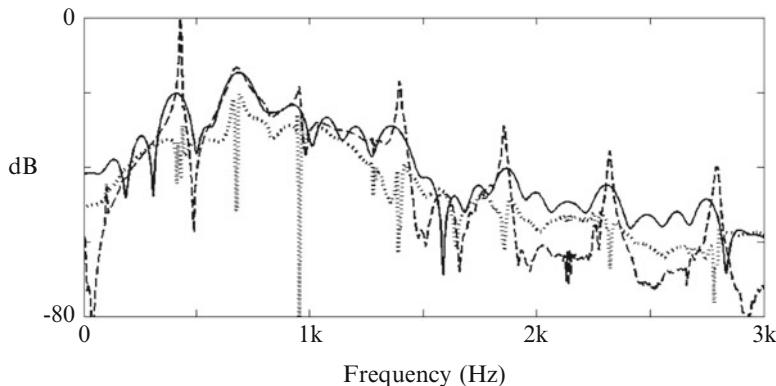


**Fig. 8.22** Example of CLSM for the first spectral peak where *open circles* indicate observed power spectral sequence, *solid circles* are estimated power spectra for sinusoidal components, and the *solid line* shows the synthesized power spectral function from Fig. 6 in [14]

synthesized waveform and its power spectral function. The decaying characteristic of vibration can be represented by the CLSM approach, and the seven dominant peaks (including the entire spectral envelope) are well represented as well as the temporal envelope of the waveform including its fine structure. However, looking in detail at the error shown by panel (c)-A reveals that the error produced by CLSM is concentrated in the initial transient portion of the waveform.

CLSM, in principle, can provide a representation of the dominant spectral components that are produced by the resonant mechanism of the target sequence. The transient portions might be important for signal analysis, in particular, for musical sound analysis, as are the spectral harmonic analysis in the frequency domain. Therefore, the CLSM approach might not be a good way to represent the transient portion of sequences such as the initial parts of impulse response records.

Figure 8.23 on the following page depicts the power spectral characteristics of the residual (error) by the dotted line; similarly, the solid line depicts that for the initial portion (within 10 ms), whereas the broken line corresponds to that for the original vibration data as shown in Fig. 8.21 on the preceding page (a)-B. There is no significant difference in the power spectral characteristics between the error and initial portion of the vibration data. In addition, the initial portion already determines the entire spectral envelope (including troughs) of the vibration data. Thus, the initial portion or the residual of CLSM still conveys a significant signature of the sequence of interest. A transient sequence, such as the initial portion of a response, with a brief record length could be characterized by the zeros or spectral troughs, as seen in Fig. 8.23 on the following page rather than the poles resulting from the resonance frequencies. The next chapter deals with a method of representing a brief sequence in the time domain.



**Fig. 8.23** Power-spectral characteristics for initial (within 10 ms) and entire vibration record with CLSM residual (error) where *solid line*: initial part, *broken*: whole of record, *dotted*: error as shown in Fig. 8.21 on page 184 (c)-B from Fig. 7 [14]

## 8.3 Prediction of Compound Sinusoidal Sequences

### 8.3.1 Compound Sinusoidal Sequence and Almost Periodic Function

Prediction, estimation or extrapolation outside an observation interval is an attractive and interesting topic in the waveform analysis. It is impossible to predict random sequences in a deterministic sense, because the non-deterministic properties are random in nature. A periodic sequence can be decomposed into sinusoidal components for which the ratios of the corresponding frequencies are rational numbers. In contrast, a function that contains pairs of sinusoidal components with frequency ratios that are irrational numbers is no longer periodic, and thus it is called an almost periodic (or quasi-periodic) function [16].

An almost periodic function is predictable in principle, despite a lack of periodicity. This is because its spectral components are given in a deterministic sense, if they can be well estimated using a record observed in an interval of finite length. However, the spectral function of a sequence of a discrete variable with a finite-length record, which could be handled on a computer, can only be given at discrete frequency bins. Thus, such a sequence is essentially expressed as a periodic sequence, so that the spectral function can be represented as a discrete sequence with line spectral components where any spectral pair in the components can be represented by a ratio of rational numbers. A compound sinusoidal sequence, which is periodic as long as it is represented by a discrete and finite basis, might be a good model for representing periodic or almost periodic functions. Therefore, if the spectral components can be well estimated for the compound sinusoidal functions, it might be possible to predict a compound sequence, even if the period is very long as for almost-almost periodic functions.

### 8.3.2 Sinusoidal Compound Sequences and Spectral Functions

A compound sinusoidal sequence might be predictable by estimating the spectral components using the observed record within an observation interval. The Fourier transform of a discrete sequence  $h(n)$  is always written as a compound sinusoidal function in the frequency domain such that

$$H(e^{-i\Omega}) = \sum_{n=0}^{N-1} h(n)e^{-in\Omega} \quad (8.46)$$

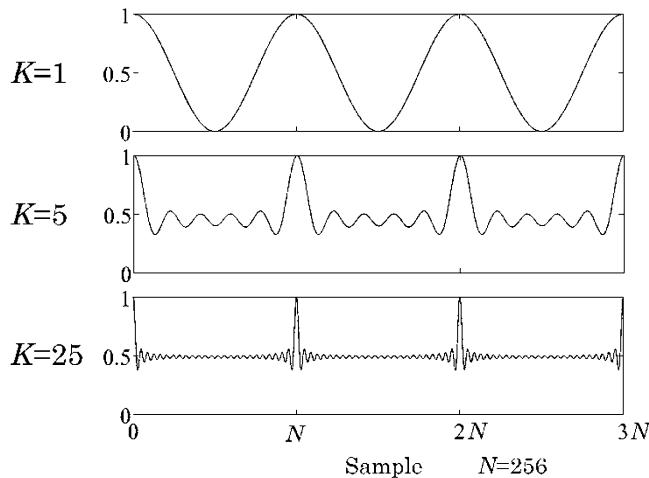
subject to  $N$  being a finite integer. It implies that the spectral function might be extended into the entire frequency plane, even if the spectral function is observed in only a limited region taken from  $0 \leq \Omega < 2\pi$ . Specifically, even if the spectral function is band limited to  $0 \leq \Omega < \Omega_H < 2\pi$ , for example, the function can be extrapolated to  $0 \leq \Omega < 2\pi$  by estimating the compound sinusoidal function in the frequency domain.

### 8.3.3 Spectral Peak Selection and Extensions Outside the Observation Interval

A single sinusoidal component can be estimated by its spectral parameters, such as the magnitude, phase, and frequency, even if the observation interval is limited to a finite length. A sinusoidal sequence is then predictable or can be extended outside the observation interval. This is also true for periodic compound sinusoidal sequences, provided the spectral peak selection works well. This is the significant difference between the spectral peak selection and conventional DFT. In principle, predictions of a sequence are impossible by DFT.

Figure 8.3 on page 166 demonstrates that a sinusoidal sequence can be estimated and extended outside the observation interval. Figure 8.4 on page 167 presents an example in which the sequence can be extended outside the observation interval using spectral peak selection. The spectral peak selection can also be applied to non-harmonic sequences. Section 8.7 on page 170 shows that spectral components can be well estimated by spectral peak selection for a harmonic sequence on the logarithmic frequency scale, provided the spectral functions do not overlap with each other. The reconstructed sequence can be extended (predicted) outside the observation interval.

However, it might not be possible if the spectral components are too densely arranged to be separately picked up by peak selection. That is, extrapolation



**Fig. 8.24** Examples of envelopes given by Eq. 8.47 where  $K = 1, 5, 25$ ,  $\Omega_c = 2\pi/N$ , and  $N$  denotes period for panels (a), (b), and (c), respectively

including the temporal envelope does not work well for such a sequence even if the sequence might be modeled by a compound sinusoidal sequence. This requirement implies the sequence composed of clustered spectral components would not be well predicted. Suppose that a sequence can be expressed as

$$y(n) = \sum_{k=1}^K (1 + \cos k \Omega_m n) \cos \Omega_c n \quad (8.47)$$

in a narrow frequency band where  $\Omega_m$  denotes the fundamental angular frequency of amplitude modulation,  $\Omega_c$  the central angular frequency of the frequency band or that for the spectral peak, and  $K$  the number of frequency components (or harmonics) for the envelope. Figure 8.24 shows samples of the envelopes generated using Eq. 8.47. The frequency band becomes wider or the number of higher harmonics increases for the periodic envelope, as  $K$  increases. In particular, a large number of frequency components (harmonics) must be estimated to make the sequence predictable outside the observation interval when the sequence might have a pulse-like periodic envelope rather than a slowly varying periodic one. CLSM might afford an approach to such cases; however, the frequency range of the temporal envelope might be necessary as prior knowledge for the observed sequence to be predicted.

## References

1. M. Tohyama, *Sound and Signals* (Springer, 2011)
2. M. Kazama, K. Yoshida, M. Tohyama, Signal representation including waveform envelope by clustered line-spectrum modeling. *J. Audio Eng. Soc.* **51**(3), 123–137 (2003)
3. M. Kazama, M. Tohyama, Estimation of speech components by ACF analysis in a noisy environment. *J. Sound Vib.* **241**(1), 41–52 (2001)
4. T. Houtgast, H.J.M. Steeneken, R. Plomp, A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* **77**(3), 1069–1077 (1985)
5. R. Drullman, Temporal envelope and fine structure cues for speech intelligibility. *J. Acoust. Soc. Am.* **97**(1), 585–592 (1995)
6. M. Kazama, S. Gotoh, M. Tohyama, T. Houtgast, On the significance of phase in the short term Fourier spectrum for speech intelligibility. *J. Acoust. Soc. Am.* **127**(3), 1432–1439 (2010)
7. M.R. Schroeder, *Fractals, Chaos, Power Laws* (W.H. Freeman and Company, 1991)
8. R. Meddis, M.J. Hewitt, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *J. Acoust. Soc. Am.* **89**(6), 2866–2882 (1991)
9. Y. Hara, M. Matsumoto, K. Miyoshi, Method for estimating pitch independently from power spectrum envelope for speech and musical signal. *J. Temp. Design Arch. Environ.* **9**(1), 121–124 2009
10. D.C. Lay, *Linear Algebra and Its Applications* (Addison-Wesley, 1994)
11. S. Lang, *Linear Algebra* (Springer, 1987)
12. M. Tohyama, T. Koike, *Fundamentals of Acoustic Signal Processing* (Academic Press, 1998)
13. O. Yasojima, Y. Takahashi, M. Tohyama, Resonant bandwidth estimation of vowels using clustered-line spectrum modeling for pressure speech waveforms, in *Int. Symp. Signal Processing and Information Technology* (2006), pp. 589–593
14. T. Hasegawa, M. Tohyama, Analysis of spectral and temporal waveforms of piano-string vibration. *J. Audio Eng. Soc.* **60**(4), 237–245 (2012)
15. G. Weinreich, Coupled piano strings. *J. Acoust. Soc. Am.* **62**(6), 1474–1484 (1977)
16. N. Wiener, *The Fourier Integral and Certain of Its Applications* (Dover, 1958)

# Chapter 9

## Modeling for Zeros in Complex Time and Frequency Plane

The poles, in principle, correspond to the resonant frequencies, whereas the zeros are created by local interaction between the responses of two adjacent poles. The occurrences of the zeros can be formulated using the residues of the adjacent poles and the remainder function that approximates the sum of responses from other poles. Clustered line-spectral modeling (CLSM) synthesizes the pole/zero responses in the frequency domain with some modeling error. An example of piano-vibration analysis by CLSM uncovers a significant zero in the modeling error from the source signature, which determines an initial portion of the time sequence for the response record, but mostly determines the entire spectral envelope of the response. Complementarity under complex conjugation holds between the complex-time and frequency planes. Like CLSM in the frequency domain, clustered time sequence modeling (CTSM) can be formulated on the complex-time domain in accordance with complementarity. CTSM of piano-string vibration analysis extracts the source signature from the initial portion of the vibration corresponds to the residual error by CLSM. Adjacent pairing time-pulse modeling (APTM) is formulated for spectral trough selection as well as the spectral peak selection on the frequency plane. The significant zeros generated by the piano-string vibration hence can be identified using APTM.

### 9.1 Sinusoidal Modeling and Zeros for Transfer Functions

#### 9.1.1 Frequency Characteristics for Single-Degree-of-Freedom System

Transfer functions are complex functions defined on a complex frequency plane. They are represented by poles and zeros. A vibration system whose impulse response is represented by decaying sinusoidal function

$$h(t) = A e^{-\delta_0 t} \sin(\omega_d t + \phi) \quad (9.1)$$

is called a single-degree-of-freedom system. Here  $\omega_d$  is the angular frequency for the damped free oscillation that is equal to the eigenfrequency if the decaying factor (or damping constant)  $\delta_0$  were equal to zero. Note that  $2\delta_0$ , sometimes called the damping constant, can be written in terms of the reverberation time  $T_R$

$$2\delta_0 = \frac{\ln 10^6}{T_R} \cong \frac{13.8}{T_R}. \quad (1/\text{s}) \quad (9.2)$$

The free oscillation of the single-degree-of-freedom system can be obtained as a solution of the differential equation

$$M \frac{d^2x(t)}{dt^2} + R \frac{dx(t)}{dt} + Kx(t) = 0 \quad (\text{N}) \quad (9.3)$$

which can be interpreted as the free oscillation of a mass and a spring. The frequency of the free oscillation is derived as the solution of the quadratic equation

$$\omega_s^2 M - i\omega_s R - K = 0. \quad (9.4)$$

Specifically, the complex frequencies of the free oscillation are

$$\omega_s = \pm \sqrt{\omega_0^2 - \delta_0^2} + i\delta_0 = \pm \omega_d + i\delta_0 \quad (1/\text{s}) \quad (9.5)$$

where

$$\omega_0 = \sqrt{K/M}, \quad \delta_0 = R/2M. \quad (1/\text{s}) \quad (9.6)$$

By taking the Fourier transform of the impulse response

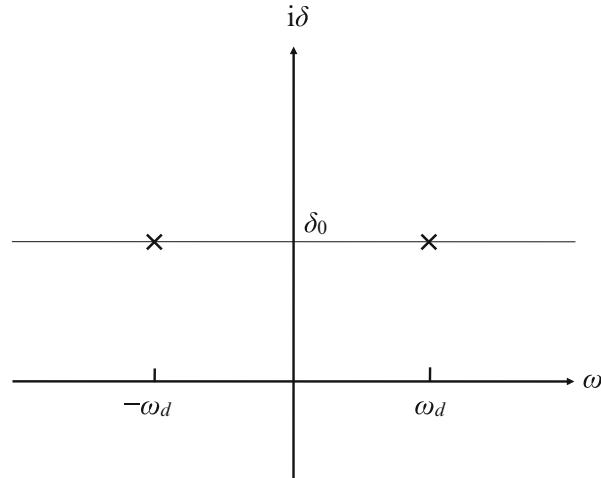
$$H(\omega) = \frac{A}{4\pi} \left( \frac{e^{-i\phi}}{\omega - \omega_{s_1}} - \frac{e^{i\phi}}{\omega - \omega_{s_2}} \right) \quad (9.7)$$

is derived; this is called the frequency characteristic of the vibrating system, where

$$\omega_{s_1} = -\omega_d + i\delta_0 \quad (9.8)$$

$$\omega_{s_2} = +\omega_d + i\delta_0. \quad (9.9)$$

In addition, by allowing the frequency to become complex  $\omega_s = \omega + i\delta$ , the function of  $\omega_s$  is called the transfer function defined on the complex frequency plane. However,  $\omega_{s_1} = \omega_{p_1}$  and  $\omega_{s_2} = \omega_{p_2}$  are denoted by the poles where the



**Fig. 9.1** Example of a pair of poles on the complex frequency plane from Fig. 13.10 in [1]

transfer functions are not defined. The poles are also called singularities of the transfer function. The poles are located above the real frequency line (axis) for sinusoidal vibrations represented using  $e^{i\omega t}$  as shown in Fig. 9.1. The distance between the poles and real frequency line corresponds to the damping constant; hence distance increases with damping. In contrast, if the damping is small, the distance is also short, and thus the poles are located very close to the real frequency line.

The magnitude of the frequency characteristics is called the magnitude frequency response, whereas the angle of the frequency characteristics is called the phase frequency response. The magnitude response is a function of frequency, and takes its maximum at the frequency, called the resonance frequency, closest to the eigenfrequency. The frequency characteristics can be approximated as

$$\begin{aligned} H(\omega) &\cong \frac{A}{4\pi} \left( \frac{(e^{-i\phi} - e^{i\phi})\omega_d - (\omega_{p2}e^{-i\phi} - \omega_{p1}e^{i\phi})}{(\omega - \omega_{p1})(\omega - \omega_{p2})} \right) \\ &= \frac{A}{4\pi} \frac{N(\omega)}{D(\omega)} \end{aligned} \quad (9.10)$$

subject to  $\omega \cong \omega_d$ . Therefore, the resonance frequency at which the magnitude is a maximum is given by

$$\omega_M = \sqrt{\omega_d^2 - \delta_0^2} = \sqrt{\omega_0^2 - 2\delta_0^2} \cong \omega_0 \quad (\text{rad/s}) \quad (9.11)$$

indicating that the denominator becomes the minimum at the frequency. Consequently, there are three different frequencies that represent the single-freedom

-of-vibration system: eigenfrequency  $\omega_0$ , frequency of free oscillation  $\omega_d$ , and resonance frequency  $\omega_M$ , where  $\omega_0 > \omega_d > \omega_M$  holds in general [2].

The frequency characteristics around the resonance frequency are sometimes simply called the resonance response. The resonance response can be written as

$$H(\omega) = \frac{A}{4\pi} H_N(\omega) H_D(\omega) \quad (9.12)$$

$$H_D(\omega) = \frac{1}{(\omega - \omega_{p1})(\omega - \omega_{p2})}. \quad (9.13)$$

The half-power bandwidth can be defined. By setting

$$|H_D(\omega_M)|^2 \cong \frac{1}{4\delta_0^2\omega_M^2} \quad (9.14)$$

the frequency  $\omega_B$  at which the squared magnitude becomes

$$|H_D(\omega_B)|^2 = \frac{1}{2}|H_D(\omega_M)|^2 \quad (9.15)$$

is given by

$$\omega_B \cong \omega_M \pm \delta_0 \quad (9.16)$$

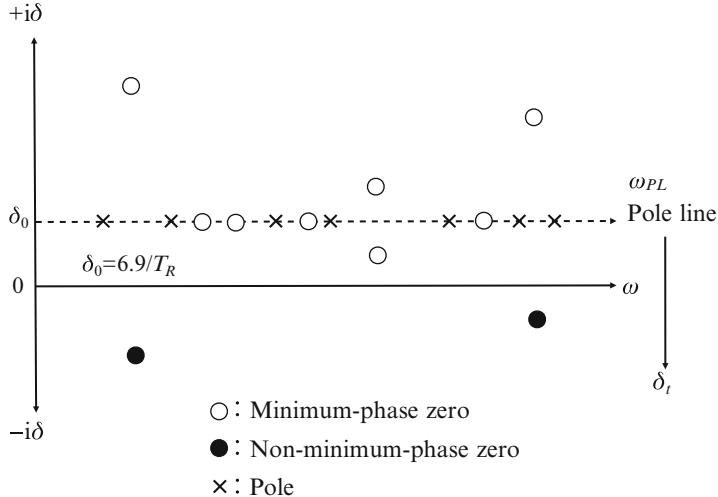
where

$$\begin{aligned} |H_D(\omega)|^2 &= \frac{1}{(\omega^2 - \omega_d^2 - \delta_0^2)^2 + 4\delta_0^2\omega^2} \\ &= \frac{1}{(\omega^2 - \omega_0^2)^2 + 4\delta_0^2\omega^2} \end{aligned} \quad (9.17)$$

and  $\delta_0$  is called the half-power bandwidth. The modal expansion form of the transfer function is just the superposition of the responses of the single-degree-of-freedom systems subject to  $\delta_N \cong \delta_0$  where  $\delta_N$  denotes the modal damping constant [1].

### 9.1.2 Residues and Zeros of Transfer Function

The transfer function of a single degree-of-freedom system can be characterized by the poles. The transfer function of a multi-degrees-of-freedom system, as in the room transfer function, can be written as a superposition of the responses of single-degree-of-freedom systems [1]. However, the transfer function of a multi-degrees-of-freedom system can contain the zeros as well as the poles. The occurrence of zeros depends on the sign of the residues of the poles [3,4].



**Fig. 9.2** Pole-zero pattern in complex-frequency plane from Fig. 13.11 in [1]

Set the  $N$ -th pole such that

$$\omega_{p_N} = \omega_{N_d} + i\delta_N \cong \omega_N + i\delta_N \quad (9.18)$$

where  $\omega_N > \delta_N > 0$  and  $\omega_{N_d} \cong \omega_N$  are assumed,  $\omega_N$  denotes the angular eigenfrequency without sound absorption, and  $\omega_{N_d}$  the angular frequency of the free oscillation. Figure 9.2 is an image of the distribution for the poles and zeros in the complex frequency plane. The horizontal axis shows the real frequency, whereas the vertical corresponds to the imaginary part of the complex frequency, i.e., the damping constant. The poles are located above the real frequency axis; the zeros are distributed above and below the frequency axis. The line connecting the poles is called the pole line, and is parallel to the real frequency axis provided the damping constant is independent of frequency. The transfer function is symmetric with respect to the pole line.

Consider next the occurrence of the zeros between two adjacent poles on the pole line [3,4]. Define the transfer function for the two poles  $\omega_A < \omega < \omega_B$  as

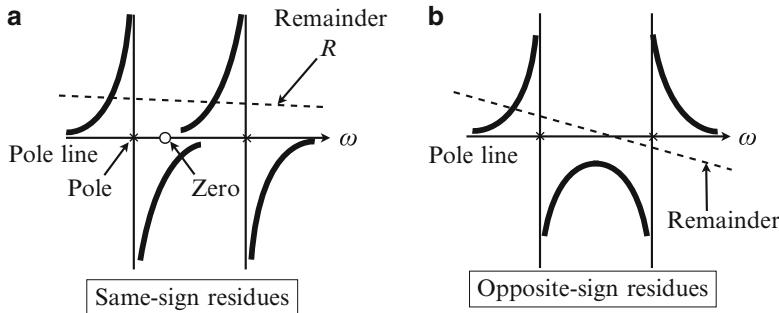
$$H(\omega_{PL}) = \frac{A}{\omega_{PL} - \omega_{PA}} + \frac{B}{\omega_{PL} - \omega_{PB}} \quad (9.19)$$

where

$$\omega_{PA} = \omega_A + i\delta_0 \quad (9.20)$$

$$\omega_{PB} = \omega_B + i\delta_0 \quad (9.21)$$

$$\omega_{PL} = \omega + i\delta_0 \quad (9.22)$$



**Fig. 9.3** Possible appearance of a zero in the interval between two adjacent poles, depending on relative signs of the residues from [5] (Fig. 4) and from Fig. 13.12 in [1]. (a) Single zero. (b) No zero

and  $A$  and  $B$  are called the residues for the respective poles and are assumed to be real numbers. Figure 9.3 is a schematic providing the condition for the appearance of a zero. If the poles have residues of the same sign, a zero appears on the pole line; no zeros occur if the residues have opposite signs [3–5].

The transfer function for a multi-degrees-of-freedom system can be expressed as a superposition of the resonance and out-of-resonance responses [1, 3, 4]. Now reconsider the occurrence of zeros between two adjacent poles for a multi-degrees-of-freedom system. For that purpose, define the transfer function as

$$H(\omega_{PL_s}) = \frac{A}{\omega_{PL_s} - \omega_A} + \frac{B}{\omega_{PL_s} - \omega_B} + R(\omega_{PL_s}) \quad (9.23)$$

where the complex frequency is extended away from the pole line into the complex frequency plane as

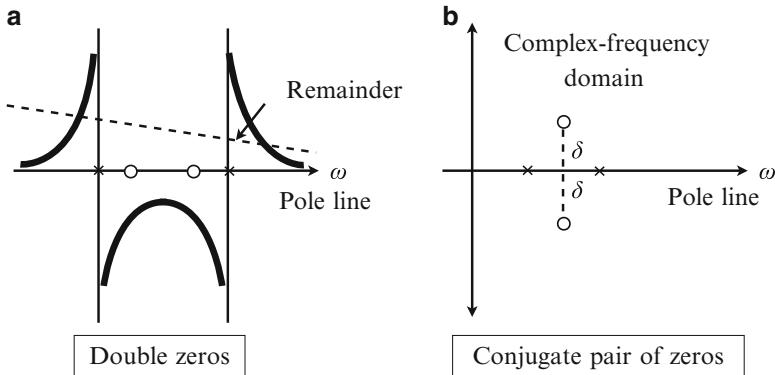
$$\omega_{PL_s} = \omega_{PL} \pm i\delta_t \quad (9.24)$$

where  $\delta_t$  denotes the perpendicular distance from the pole line.  $R(\omega_{PL_s})$  is called the remainder function assuming that [3, 4]

$$R(\omega_{PL_s}) \cong \text{const} \quad (\omega_A < \omega < \omega_B). \quad (9.25)$$

Returning to Fig. 9.3, there is a zero on the pole line between two adjacent poles with same-sign residues, despite a nonzero remainder function. In contrast, three instances occur for poles with opposite-sign residues, specifically, no zero, double zero, and a symmetric pair of zeros (see Fig. 9.4 on the facing page). These zero locations can be formulated as follows [6].

By assuming the remainder function is almost constant, the transfer function can be approximated by [3]



**Fig. 9.4** Zeros from opposite-sign residues and remainder from [5] (Fig. 5) and Fig. 13.13 in [1].  
**(a)** Double zero. **(b)** Pair of symmetric zeros

$$H(\omega) \cong \frac{A}{\omega - \omega_A} + \frac{B}{\omega - \omega_B} + R \quad (9.26)$$

where  $\omega_{PL_s}$  is simply written as  $\omega$ . First, consider that  $R \cong 0$ . A zero

$$\omega_z = \omega_0 + \Delta\omega \frac{A - B}{A + B} \quad (9.27)$$

is obtained as a solution of the equation

$$H(\omega_z) = \frac{A}{\omega_z - \omega_A} + \frac{B}{\omega_z - \omega_B} = 0 \quad (9.28)$$

where

$$\begin{aligned} \omega_0 &= \frac{\omega_A + \omega_B}{2}, & \Delta\omega &= \frac{\omega_B - \omega_A}{2} \quad (\omega_B > \omega_A) \\ \omega_B &= \omega_0 + \Delta\omega, & \omega_A &= \omega_0 - \Delta\omega. \end{aligned}$$

If  $A$  and  $B$  are residues with the same sign

$$\left| \frac{A - B}{A + B} \right| < 1 \quad (9.29)$$

holds and consequently the zero is located on the pole line between the poles.

In contrast, consider that  $A$  and  $B$  are with opposite sign and consider the zero on the pole line for the equation

$$\frac{A}{\omega_z - \omega_A} + \frac{B}{\omega_z - \omega_B} + R = 0. \quad (9.30)$$

This equation can be rewritten as

$$\frac{A}{\hat{\omega} + \Delta\omega} + \frac{B}{\hat{\omega} - \Delta\omega} + R = 0 \quad (9.31)$$

by introducing variables

$$\hat{\omega} = \omega_z - \omega_0, \quad \omega_z - \omega_A = \hat{\omega} + \Delta\omega, \quad \omega_z - \omega_B = \hat{\omega} - \Delta\omega. \quad (9.32)$$

The solutions of the quadratic equation above are given by

$$\omega_z = \omega_0 + \frac{-(A + B) \pm \sqrt{(A + B)^2 + 4R \cdot (R \cdot (\Delta\omega)^2 + (A - B)\Delta\omega)}}{2R}. \quad (9.33)$$

For simplicity, consider that  $|A| = |B| = A > 0$ . Assuming that the residues have the same sign, the zeros are given by

$$\omega_{z_1} = \omega_0 + \sqrt{\frac{A^2}{R^2} + (\Delta\omega)^2} + \frac{-A}{R} \quad (9.34)$$

$$\omega_{z_2} = \omega_0 - \sqrt{\frac{A^2}{R^2} + (\Delta\omega)^2} + \frac{-A}{R} \quad (9.35)$$

one of which resides on the pole line between the poles [3, 4, 6]. Specifically, if  $A/R > 0$ , then  $\omega_{z_1}$  is the corresponding zero; if  $A/R < 0$ ,  $\omega_{z_2}$  is the zero. This is because

$$-\Delta\omega < \sqrt{\frac{A^2}{R^2} + (\Delta\omega)^2} + \frac{-A}{R} < \Delta\omega \quad (9.36)$$

holds when  $A/R > 0$ .

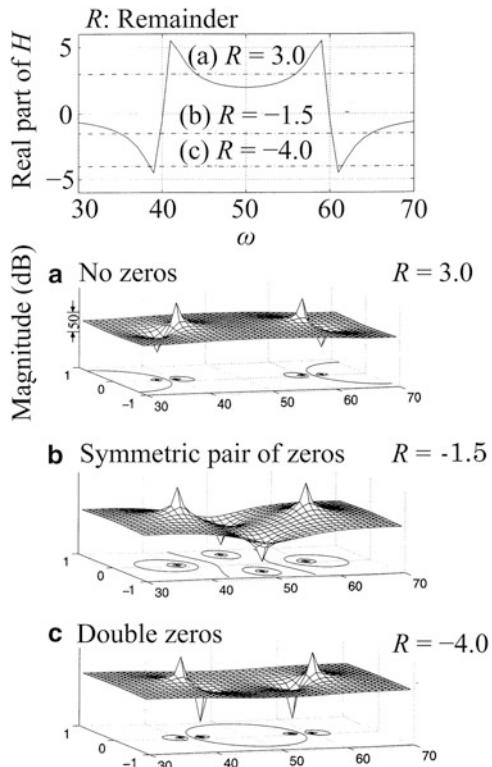
Figure 9.5 on the facing page exemplify the three distinct instances for opposite-sign residues. As  $|A| = |B| > 0$ , the zeros are re-expressed as

$$\omega_z = \omega_0 \pm \Delta\omega \sqrt{1 + \frac{2A}{\Delta\omega R}} \quad (9.37)$$

and enables a classification: there are

- (a) no zeros if  $R > 0$
- (b) a symmetric pair of zeros if  $R < 0$  and  $1 + \frac{2A}{\Delta\omega R} < 0$  that is given by a pair of complex numbers such that

**Fig. 9.5** Occurrence of zeros for opposite-sign residues from [7] (Fig. 7.4.4) and Fig. 13.14 in [1]. (a) No zeros. (b) Symmetric pair of zeros. (c) Double zeros



$$\omega_z = \omega_0 \pm i\Delta\omega \sqrt{-\left(1 + \frac{2A}{\Delta\omega R}\right)} \quad (9.38)$$

and

(c) double zero on the pole line if  $R < 0$  and

$$1 > 1 + \frac{2A}{\Delta\omega R} > 0 \quad (9.39)$$

that coincide if  $1 + \frac{2A}{\Delta\omega R} = 0$ . If  $A < 0$ , a similar analysis holds with the sign of  $R$  inverted [6].

Both minimum and non-minimum-phase zeros can be contained in the transfer function. Indeed a non-minimum-phase zero is produced as one of a pair of zeros, which are located equi-distant from the pole line. This is because the transfer function has a symmetrical form in the complex domain with respect to the pole line, assuming real residues. One member of this pair of zeros should be non-minimum phase in a slightly damped system, because the pole line runs just above (below) the real-frequency axis assuming an  $e^{i\omega t}$  ( $e^{-i\omega t}$ ) time dependency. Note that,

because of the counter-balancing phase behavior from the symmetry of the pair of zeros, no clear phase jump is observed on the real frequency axis near the non-minimum-phase zero when the damping of the transfer function is very small. It therefore produces phase characteristics consistent with minimum-phase behavior. The symmetric location of the pair of conjugate-like zeros equidistance above and below the pole line, approximately equal to the real frequency axis, cancels their phase effects [6].

### 9.1.3 Sinusoidal Modeling and Zeros for External Source

Sound or vibration, in which the source information is conveyed, is generally observed using a vibration system. Source signature analysis could be a powerful tool for condition monitoring and diagnostics, even if the entire source waveform cannot be estimated [8, 9]. A significant source signature might be represented by zeros for a pulse-like source waveform. However, because the transfer function is also expressed by poles and zeros, an important question is how to distinguish the zeros representing the source signature from those representing the reverberant path information [9].

The occurrence of the transfer function zeros between an adjacent pair of poles is formulated by introducing the remainder function developed in the previous subsection. The source characteristics could also be reflected in a remainder function that might not always be a slow-varying function. The zeros between a pair of adjacent poles can be estimated by solving

$$\begin{aligned} H(z^{-1}) &\cong \frac{A}{1 - az^{-1}} + \frac{B}{1 - bz^{-1}} + R(z^{-1}) \\ &= H_p(z^{-1}) + R(z^{-1}) = 0 \end{aligned} \quad (9.40)$$

where  $a$  and  $b$  denote the pair of adjacent poles,  $A$  and  $B$  are the residues of the poles, and  $R(z^{-1})$  represents the response from the modes other than those owing to the two poles. The  $R(z^{-1})$  is assumed to be a slow-varying function between the poles. On the one hand, a zero is generated when  $A$  and  $B$  have the same sign. On the other hand, there are three possibilities when  $A$  and  $B$  have opposite signs, specifically, no zeros generated, two zeros generated on the pole line between the pole pair, and two zeros generated creating a complex conjugate pair across the pole line.

Consider that the response record  $y(n)$  for pulse-like source  $x(n)$  is obtained using a reverberant transfer function. Taking the  $z$ -transform of the observation response

$$Y(z^{-1}) = X(z^{-1})H(z^{-1}). \quad (9.41)$$

The  $X(z^{-1})$  represents the source characteristics

$$X(z^{-1}) = \sum_{n=0}^{N-1} x(n)z^{-n} \quad (9.42)$$

which is obtained by setting  $x(0) = 1$ . The  $H(z^{-1})$  determines the transfer function of the reverberant space. Thus, the observation record can be rewritten as

$$\begin{aligned} Y(z^{-1}) &= H(z^{-1}) + \left[ \sum_{n=1}^{N-1} x(n)z^{-n} \right] H(z^{-1}) \\ &= H(z^{-1}) + R(z^{-1}). \end{aligned} \quad (9.43)$$

That is, the remainder function  $R(z^{-1})$  is necessary for representing the source signature, and the modeling error for the synthesized transfer function can be used to identify the zeros representing the source characteristics.

Source information can be found in the remainder function, which is the residual when the transfer function is subtracted from the observation record. Although the transfer function is not likely to be known in practical situations, it might be possible to reconstruct the transfer function using CLSM [1, 9, 10]. The magnitude and phase of those components are formulated using a least-squares-error (LSE) criterion in the frequency domain for the modal response. If  $N$  points of the observed recorded signal are synthesized using  $M < N$  sinusoids, the vector of CLSM solutions can be obtained using the LSE criterion

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}. \quad (9.44)$$

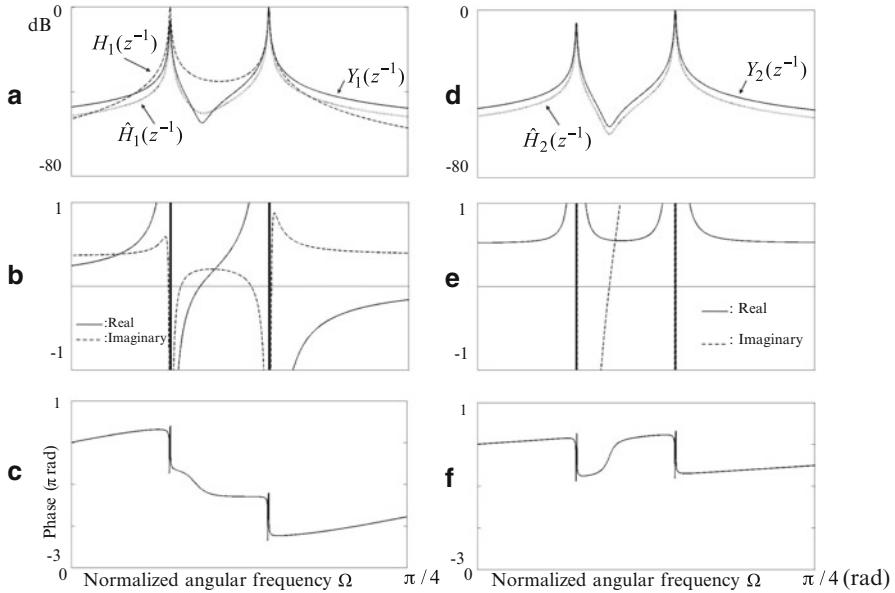
where  $\mathbf{b}$  is the vector of the observation record, and  $A$  is an  $N$  by  $M$  matrix in which the row vectors represent frequency-shifted spectra of a rectangular window function with window length  $L$ .

Figure 9.6 on the following page is an example of discriminating between the zeros representing the source properties

$$Y_1(z^{-1}) = X_1(z^{-1})H_1(z^{-1}) \quad (9.45)$$

and those of the transfer function

$$Y_2(z^{-1}) = H_2(z^{-1}). \quad (9.46)$$



**Fig. 9.6** Numerical examples of zeros representing sound source and transfer function, where  $z = e^{-i\Omega}$ : (a) and (d) power spectrum of observed responses; (b) and (e) remainder function; (c) and (f) phase of remainder function from Fig. 1 in [9]

Here

$$H_1(z^{-1}) = \frac{1}{1 - az^{-1}} + \frac{-1}{1 - bz^{-1}} \quad (9.47)$$

$$H_2(z^{-1}) = \frac{1}{1 - az^{-1}} + \frac{1}{1 - bz^{-1}} \quad (9.48)$$

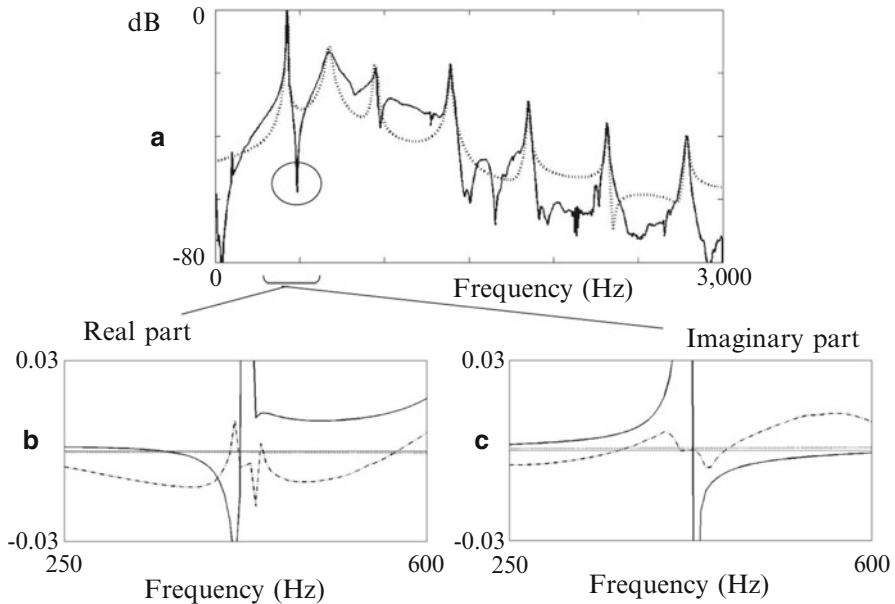
$$a = 0.998e^{2\pi i(440/6000)} \quad (9.49)$$

$$b = 0.998e^{2\pi i(880/6000)} \quad (9.50)$$

$$X_1(e^{-i\Omega}) = 1 - 0.9716e^{2\pi i(880/6000)}. \quad (9.51)$$

Panels (a) and (d) show the power spectra of the observed responses ( $Y$ ) (solid lines) and the synthesized ones  $\hat{H}$  using CLSM (dotted lines). There is no clear difference between the solid lines; however, the transfer function  $H_1(z^{-1})$  has no zeros whereas  $H_2(z^{-1})$  has a single zero between the two poles. Although the synthesized curves for the two responses are similar, the errors are substantially different around the zeros. This suggests that the error can be used to distinguish the two types of zeros.

Panels (b) and (e), depict the real part (solid line) and imaginary part (dashed line) of the modeling error,  $R_1 = Y_1 - \hat{H}_1$  and  $R_2 = Y_2 - \hat{H}_2$ , i.e., the remainder function. Panels (c) and (f), depict the phases of the modeling error. There is no



**Fig. 9.7** Example of piano-string vibration analysis, (a) solid line: observed record, dotted line: synthesized response by CLSM, (b) and (c) close-ups of real and imaginary parts, solid: synthesized two-pole response; dotted: remainder function from out-of-frequency band; dash-dotted: CLSM modeling error from Fig. 2 in [9]

phase jump in (c) meaning that the modeling error is not null at the spectral zero. Consequently, the single zero in panel (a) must represent the source characteristics. That is, the remainder function, or modeling error, is needed to create the zero representing the source characteristics. In contrast, the dotted line in panel (d) accurately represents the response of  $Y_2(z^{-1})$  around the zero in the pole pair interval because the modeling error is almost null in the interval between the two poles, as illustrated by the phase of the error in panel (f). In short, the single zero in panel (d) could be identified as a zero representing the path information.

Figure 9.7 shows the results of a waveform analysis of a piano-string vibration. The solid line in panel (a) represents the power spectral record captured using an acceleration pickup for a single tone of an up-right piano as the note A4 was played. The vibration of the string was recorded at about  $L/16$ , where  $L$  is the string length between the ends. There are dominant peaks and dips that correspond to the poles and zeros of the response record.

CLSM was iterated 7 times, corresponding to the seven spectral dominant peaks. The dotted line shows the superposition of the CLSM results; most of the dominant peaks are well fitted. Most of particular interest are the double zeros appearing around 2,500 Hz. This means that the responses from the lower frequency intervals than that for 2,500 Hz might be needed to create the zeros that constitute a slowly varying remainder function. The other zeros not created by the CLSM fitting imply

the source-related characteristics that would be observable at the recording position. That is, the condition of the hammer position can be reflected in the zeros or the modeling error of a not-slowly-varying remainder function. Panels (b) and (c) respectively show close-ups of the real and imaginary parts for the response of the lowest interval of the pole pairs. The solid lines show only the responses summed over the two-pole fittings whereas the dotted lines depict the slowly varying remainder function composed of responses outside the band. The dash-dotted lines show the modeling error that creates a zero by cancelling the solid line, i.e., the error needed for the zero to be a component of the not-slowly-varying remainder function. The remainder function needed for creating the double zeros is smaller than those for the zeros representing the sound source. This shows that the zero arises from the sound source rather than from the transfer function.

CLSM with the remainder function might be a possible approach to distinguish the zeros representing a sound source from those representing the transfer function. It is based on the residue-sign model for modeling zeros corresponding to a slow-varying remainder function, as described in the previous subsection. The source characteristics are formulated as a remainder function (or modeling error), subject to the source characteristics given by a finite impulse response. The modeling error that produces the zeros representing the source characteristics is not always a slow-varying function.

The source characteristics represented by a brief impulse response (or zeros on the frequency plane) can also be modeled in the time domain. The next section describes the modeling of the pulse-like waveform in the time domain based on the correspondence between the complex time and frequency planes.

## 9.2 Clustered Time-Sequence Modeling (CTSM)

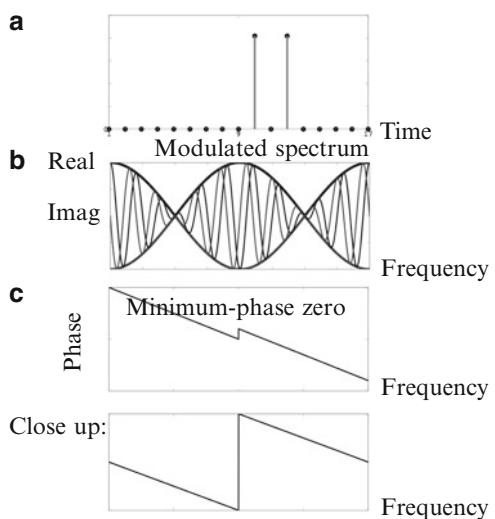
### 9.2.1 Correspondence Between Time and Frequency Regions

The Fourier transform of a single pulse is a sinusoidal function in the frequency plane in the complex form. If there are two pulses in the time domain, its Fourier transform is the modulated complex sinusoidal sequence (Fig. 9.8 on the facing page). Specifically, the zeros are produced in the frequency domain. These zeros can be interpreted as spectral troughs caused by early echoes generated in the room acoustics. They are minimum phase as shown in Fig. 9.8c on the facing page.

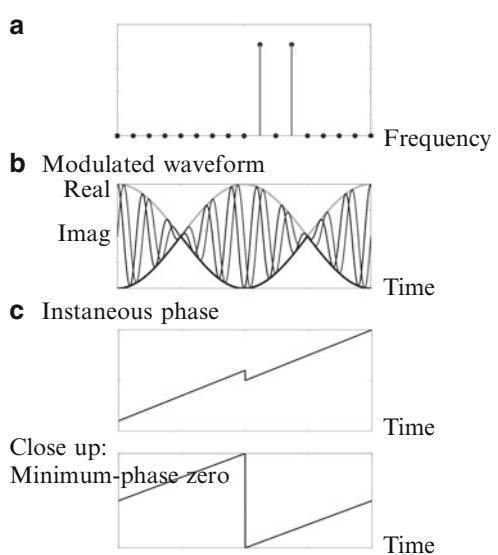
In contrast, taking the inverse Fourier transform for the two line-spectral components, the modulated time waveform is obtained (Fig. 9.9 on the facing page). The zeros can be defined in the time region, and those zeros are interpreted as minimum phase in the complex time domain, as shown by the sign changes in the real and imaginary parts of the carrier in Figs. 9.10 on page 206 and 9.11 on page 206. These results show the complementarity of time and frequency as conjugate variables.

The phase change induced by the minimum-phase zeros on the complex time domain can be represented by the positive instantaneous frequency corresponding

**Fig. 9.8** Pair of unit pulse (a) and its modulated spectrum (b) and phase spectra (c) from Fig. 14.31 in [1]



**Fig. 9.9** Pair of two line-spectral components (a) and its modulated waveform (b) with instantaneous phase (c) (taken from Fig. 14.34 [1])



to the representation of the analytic signal. The non-minimum-phase zeros can be produced in the complex time domain as well as in the frequency region. Figure 9.12 on page 207 illustrates another example of the Fourier transform of a pair of line-spectral components. It looks similar to the curves plotted in Fig. 9.9; the sign changes in the carrier are different as shown in Fig. 9.13 on page 207. Specifically, the zeros are interpreted as non-minimum-phase zeros (Figs. 9.14 on page 208 and 9.15 on page 208). Such non-minimum-phase zeros correspond to the negative instantaneous frequencies.

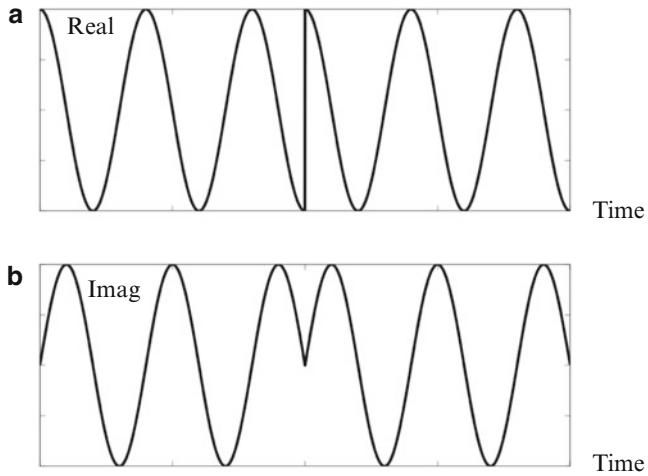


Fig. 9.10 Real (a) and imaginary (b) parts for minimum-phase carrier from Fig. 14.35 [1]

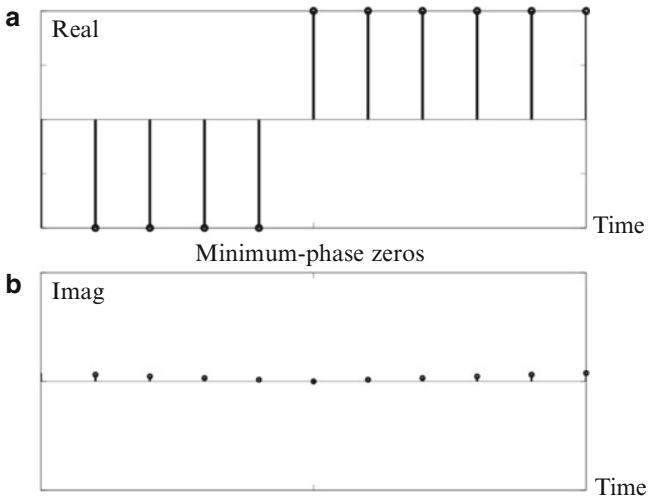
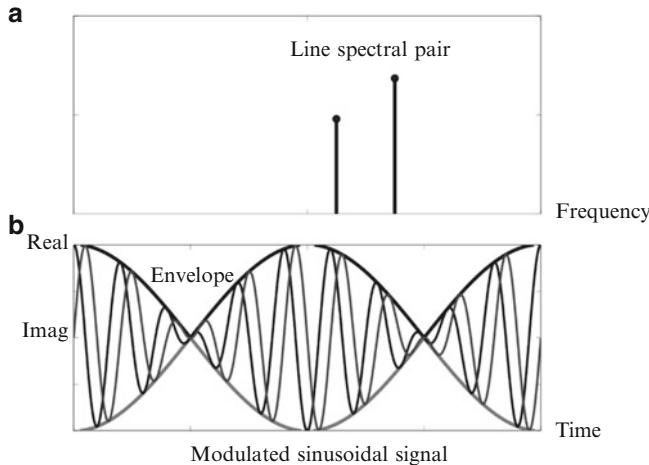
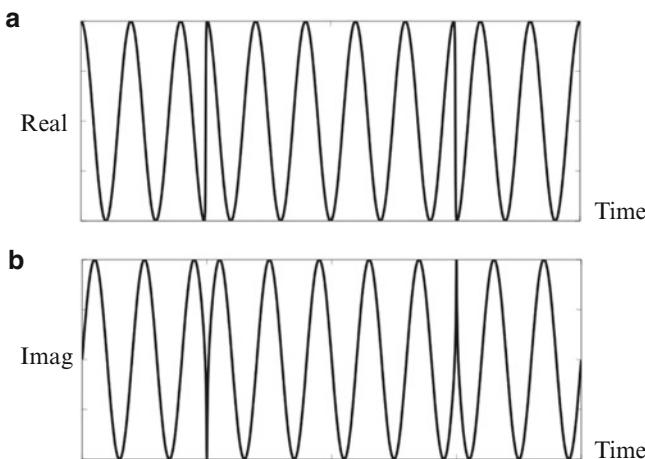


Fig. 9.11 Sign changes for the real (a) and imaginary (b) part of the minimum-phase carrier from Fig. 14.36 [1]

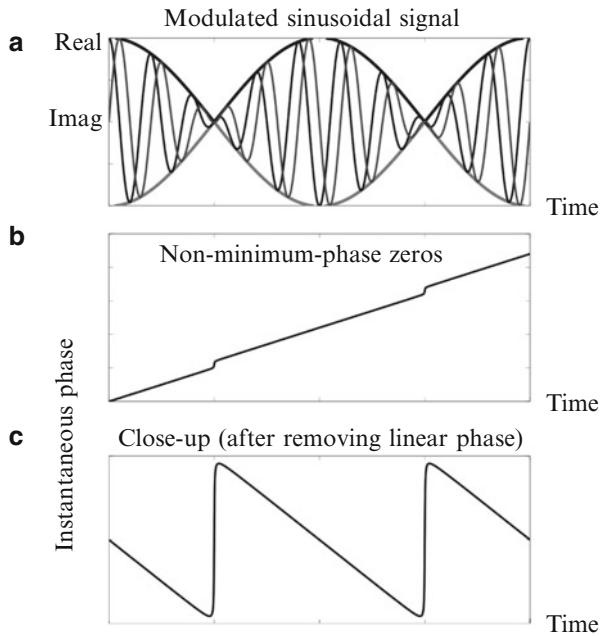


**Fig. 9.12** Asymmetric line-spectral pair (**a**) and modulated (analytic) sinusoidal signal (**b**) from Fig. 14.37 in [1]

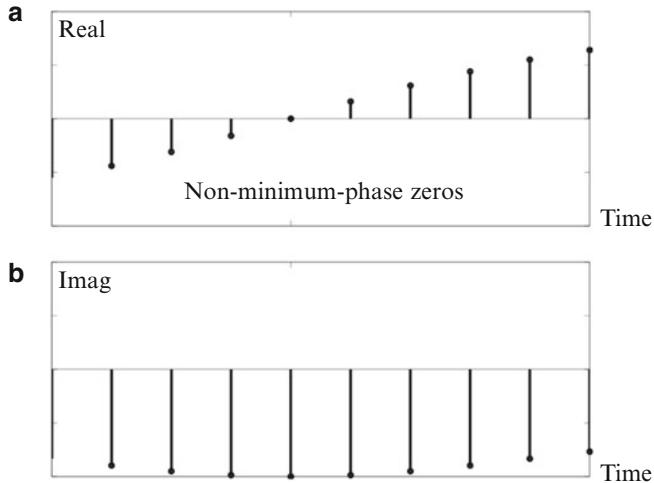


**Fig. 9.13** Real (**a**) and imaginary (**b**) parts for non-minimum phase carrier from Fig. 14.38 in [1]

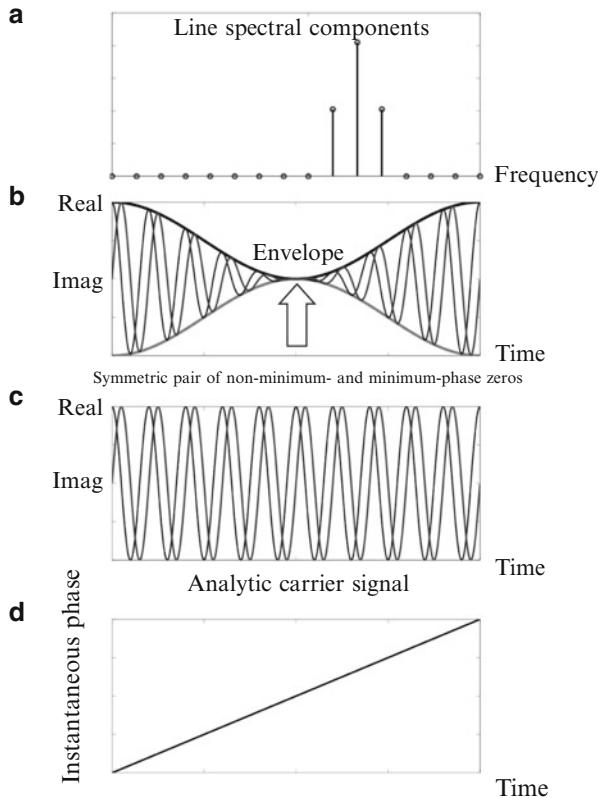
An example is presented in Fig. 9.16 on page 209, where the three line-spectral components are arranged in the frequency plane. If this symmetric arrangement of the pulse-like components is available in the time region, the linear-phase characteristic can be obtained. The same thing happens in the time domain. The inverse Fourier transform for the line-spectral sequence shows the actual linear phase in the complex time domain. The zeros might be located as symmetric pairs with respect to the real time axis in the complex time domain. Consequently, phase



**Fig. 9.14** Modulated (analytic) waveform (a) and non-minimum phase instantaneous phase with (b) and without (c) linear-phase component from Fig. 14.39 in [1]



**Fig. 9.15** Sign change of real (a) and imaginary (b) parts for non-minimum phase carrier from Fig. 14.40 in [1]

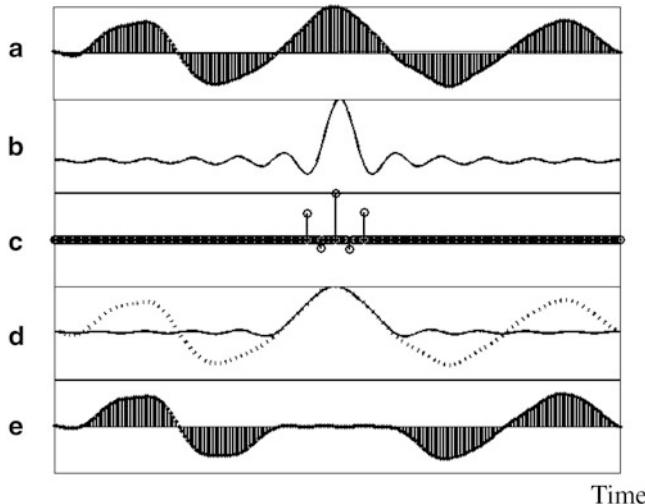


**Fig. 9.16** Three symmetric sinusoidal components (a) and amplitude-modulated analytic waveform (b), its carrier (c), and instantaneous phase (d) from Fig. 14.41 in [1]

effects from the symmetrically located zeros cancel each other, and thus the linear phase can be seen.

Both the sequence and spectrum can be represented by the complex variables. The magnitude and phase are applied to the spectrum, and similarly the envelope and instantaneous phase are assigned to the complex sequences. Recall that the magnitude or phase spectral components can be converted to each other for the minimum-phase sequences. The same thing is possible for the minimum-phase complex-time sequences. That is, the envelope and carrier part can be converted into each other if the complex time sequence is minimum phase. However, from experimental studies, sound, as in speech, mostly seems to be non-minimum phase in the complex-time domain [11].

Filtering with the filtered impulse response and windowing with the windowed spectral function make a corresponding pair between the time and frequency planes. The effect of filtering in the frequency domain can be seen by the smearing of sequences in the time domain, and by the smearing of spectral responses because of



**Fig. 9.17** Schematic for CTSM; (a) sample of the time waveform, (b) impulse response of the narrow-band filter, (c) CTSM solution vector as clustered time series, (d) synthesized response, and (e) residual error from Fig. 14.42 [1, 12]

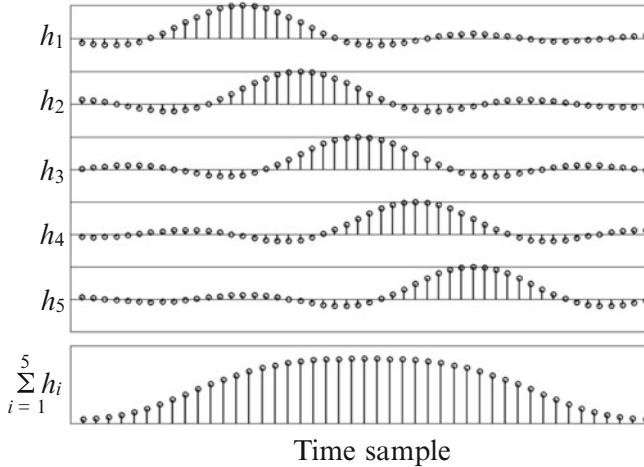
windowing in the time domain. Actually, the CLSM approach is based on spectral leakage as a result of time windowing of the sequence. The same type of approach might be able to represent a narrow-band sequence in accordance with signal smearing by filtering in the frequency plane. Inspired by CLSM, the term CTSM has been given to this time-sequence analysis.

### 9.2.2 *Formulation of CTSM*

CTSM is a method for representing a transient sequence in a short time period as an output response from a narrow-band filter to an input sequence composed of clustered time pulses[1,12]. Therefore, CTSM is formulated in the time region based on the same type of principle that formulates the CLSM in the frequency plane [1, 10, 12].

In accordance with the correspondence between the time and frequency domains, the spectral leakage arising from windowing the target sequence can be interpreted as the impulse response from narrow-band filtering. Specifically, the window-length used for CLSM in the time region corresponds to the bandwidth of the filtering for CTSM. Thus, the overlap of the leakage spectra that is the basis of the CLSM represents a superposition of the impulse response records in the time domain for the CTSM [13].

Figure 9.17 explains the CTSM approach graphically. Assume that a sequence is written as a superposition of impulse responses from filtering



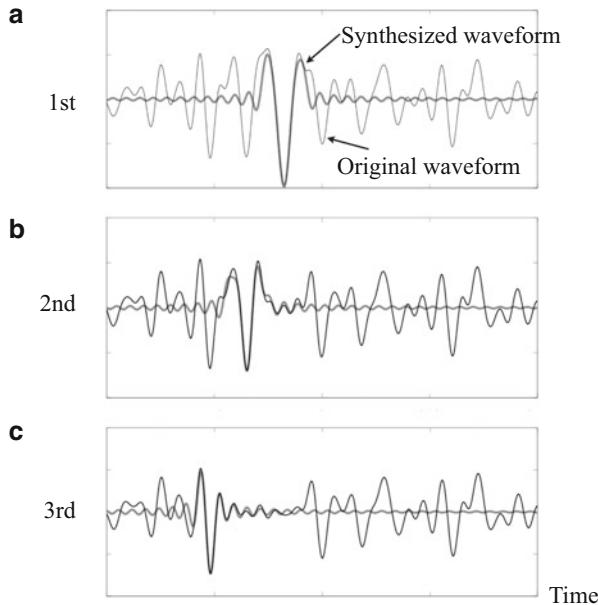
**Fig. 9.18** Superposition of impulse response records for CTSM from Fig. 14.43 in [1, 12]

$$x(n) = \sum_{m=1}^M a(m)h(n - l_m) \quad (9.52)$$

where  $l_m$  denotes the time-shift for the impulse response  $h(n)$ . By taking  $L$  observation points around the peak ( $L > M$ ) in the sequence, the clustered time series  $a(m)$  can be obtained as the LSE solutions for  $L$  simultaneous equations similar to the CLSM approach but in the time domain. Figure 9.17 on the preceding page shows a sample of a time waveform, the impulse response for the narrow-band filtering, the solution with respect to the dominant peak expressed as the clustered time series, the synthesized response by narrow-band filtering the solution of the sequence, and the residual component. Figure 9.18 depicts a magnification of the superposition of the impulse responses corresponding to the overlap of the leakage spectra for CLSM in the frequency domain. By applying the CTSM approach to the residual, the second dominant peak can be characterized by the second clustered time sequence. By repeating the process so that the residual becomes sufficiently small, the transient response can be represented by CTSM (Fig. 9.19 on the following page).

### 9.2.3 Source Signature Analysis by CTSM

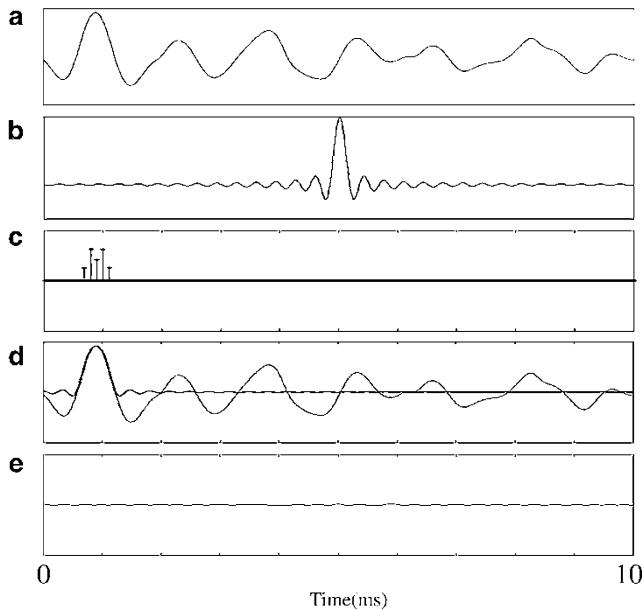
Figure 9.20 on page 213 is an example of the CTSM approach to the initial transient portion of the vibration where the CLSM approach might not be good [12]. Figure 9.20a on page 213 displays the initial portion between 0 and 10 ms of the vibration record. Moreover, Figure 9.20b on page 213 displays the impulse



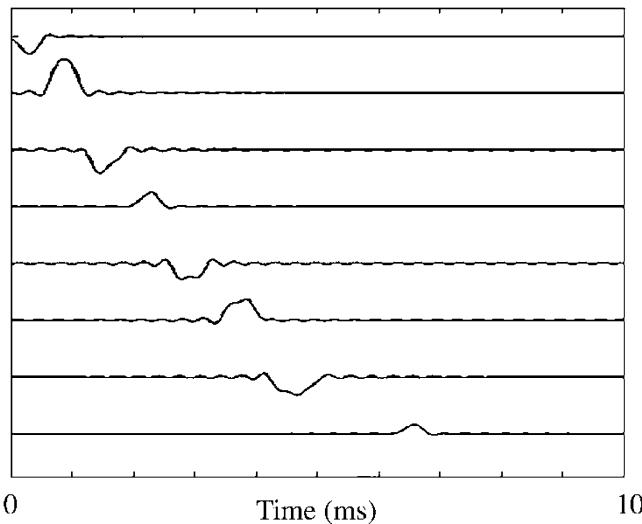
**Fig. 9.19** Repeated CTSM from Fig. 14.44 in [1, 12]

response for narrow-band filtering (lower than 3 kHz). Recall that narrow-band filtering corresponds here to time-windowing for CLSM, but should be interpreted more correctly as over-sampling rather than filtering, corresponding to the time-windowed sequence being expressed by interpolated spectral components obtained by the DFT with zero-padding in the CLSM approach. Panel (c) presents the clustered time sequence that produces the first dominant peak by filtering. This solution was obtained by assuming five pulses and observing the waveform at seven points around the peak. Panel (d) shows the synthesized waveform and the original one; by repeating this process twenty times on the residual that is defined by subtraction of the synthesized waveform from the original, the residual [panel (f)] remains.

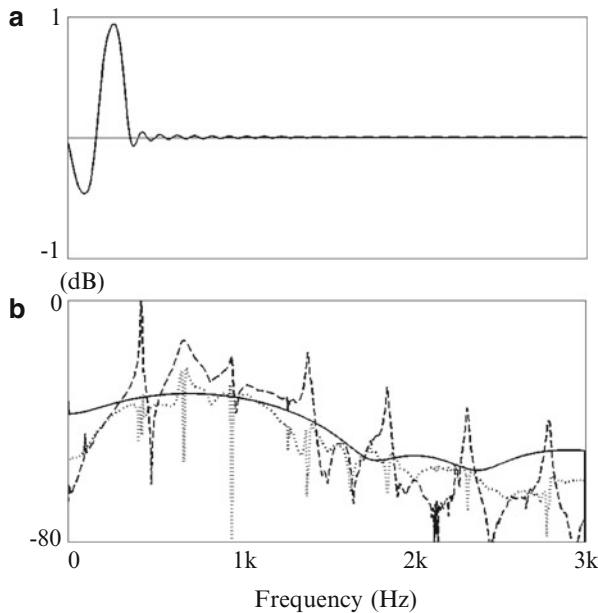
Synthesized waveforms by CTSM could be used to identify the direct and cyclic waveforms that reflect the sound source properties [12]. The direct wave travels on both sides of the string and approaches the observation point. However, the observed waveform is composed of the direct waveform and the waveform reflected from the boundaries. The first and second waveforms in Fig. 9.21 on the facing page can be assumed to be the direct waveform from the source. The third and fourth can then be assumed to be the reflected waves from the right and left boundaries. Figure 9.22a on page 214 shows the estimated direct waveform by CTSM, and panel (b) shows the power spectrum by a solid curve. The broken line shows the power spectrum of the entire observed waveform given in Fig. 9.23 on page 214. When the two curves are compared, it is clear that the direct wave reflects the properties of the



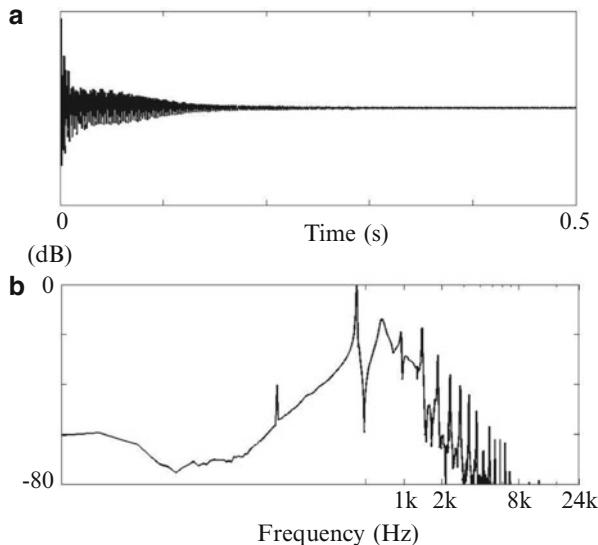
**Fig. 9.20** CTSM example for string vibration analysis: (a) initial portion of the string vibration record (0–10 ms), (b) impulse response of low-pass filtering lower than 3 kHz, (c) clustered time sequence obtained for the first dominant peak (b), (d) synthesized waveform (*thick*) with the original (*thin*), (e) residual left after 20 repetitions from Fig. 14.45 in [1] and [12]



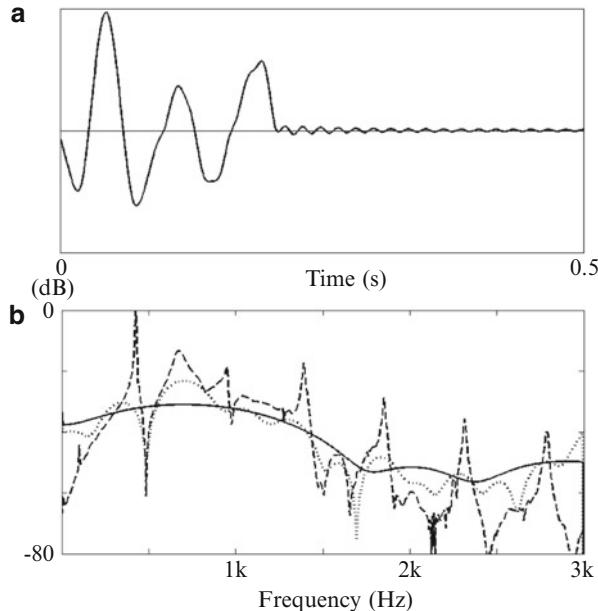
**Fig. 9.21** Dominant samples of synthesized waveforms by CTSM in time order from Fig. 13 in [12]



**Fig. 9.22** Estimated direct wave with its power spectrum: (a) waveform of estimated direct wave, (b) solid line: power spectral function of (a), broken: original record, dotted: residual of CLSM from Fig. 14 in [12]



**Fig. 9.23** Time-waveform of piano-string vibration (a) and its power spectrum (b) where sampling frequency is 48 kHz from Fig. 2 in [12]

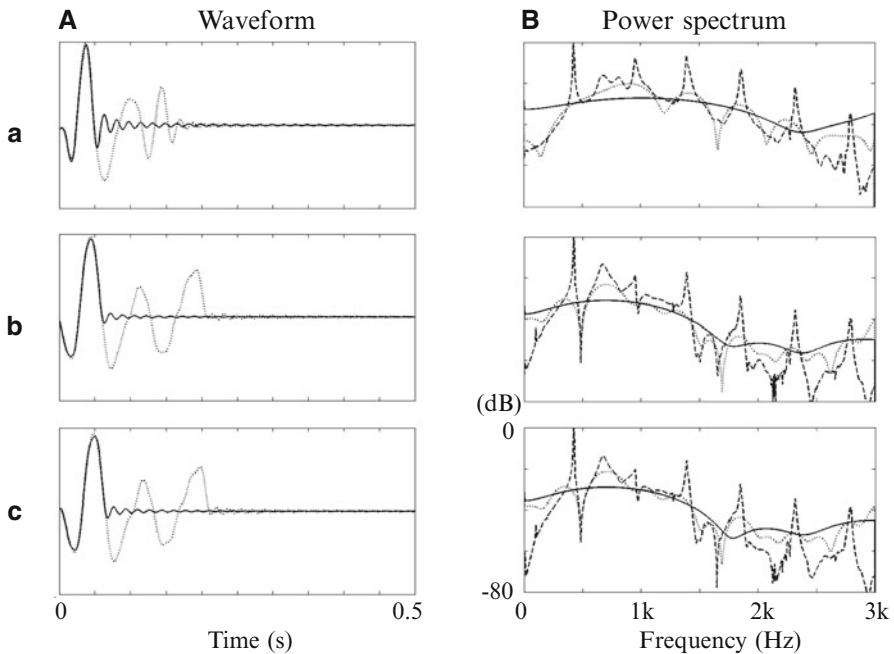


**Fig. 9.24** Estimated cyclic wave: (a) waveform of estimated cyclic wave and (b) power spectra; dotted line: power spectral function of (a), broken: original record, solid: direct wave from Fig. 15 in [12]

over-all frequency characteristics of the original waveform. In addition, the spectral characteristics of the residual error given by the dotted line seem similar to that for the direct waveform.

A single cycle of a periodic wave characterizes a musical sound as the sound source does. A cyclic wave is estimated from the synthesized wave using CTSM. A single-cycle waveform might ideally be composed of the direct and reflected waveforms. As a direct wave, the cyclic waveform is estimated using the CTSM results shown in Fig. 9.24 (including the reflected waves at the right and left edges). The estimated cyclic waveform is shown in Fig. 9.24a. Panel (b) shows the power spectrum (dotted line) of the estimated cycle of the periodic wave. The synthesized cycle of the waveform produces the dominant troughs that indicate the original spectrum (broken line). The single cycle of the periodic waveform characterizes the entire spectral envelope including the troughs that could not be produced by the direct wave (solid line). This result corresponds to that in the previous subsection in that the first deep trough (zero) arises from the source characteristics.

Figure 9.25 shows the waveforms and power spectrum obtained by plucking the piano string instead of being hammer struck, whereas panels (b) and (c) show two different samples of the same string, hammer struck by the same player. The result determines that the power spectrum of the estimated direct wave expresses the overall characteristics of the original waveform, and that the cyclic wave

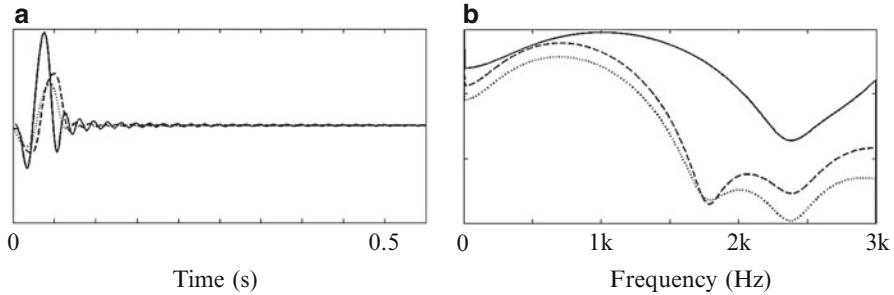


**Fig. 9.25** Waveform (A) and power spectrum (B) of the direct wave and cycle wave for the three types of sound source (a) plucked string, (b) hammered string 1, (c): hammered string 2 where *solid*: estimated direct wave, *dotted*: estimated cycle wave, *broken*: original record is given in Fig. 9.23 on page 214 from Fig. 16 in [12]

reproduces the dominant troughs for a different sound source. Here, the troughs of the cyclic wave are dependent on the observation point but the direct wave is independent of the observation point, so the direct wave must be used to compare the sound-source characteristic in three experiments. Comparing the results for the piano-string samples in Fig. 9.26 on the facing page shows that the overall sound-source characteristic are similar, although a slight difference can be seen around the troughs above frequency 1.5 kHz. In contrast, the plucked string shows a significant difference from the other two in that the components above 1 kHz are much larger.

The direct and cyclic waveforms, the source signature of which is hidden in both, were estimated using CTSM. As illustrated above, the overall characteristics of the original waveform can be expressed by the estimated direct wave, and the dominant troughs for the frequency characteristics can be reproduced from the estimated cyclic wave. The initial portion of the time waveform, which is difficult to analyze using CLSM, includes important source information. The source signature of the direct wave can be observed despite the vibrating system using CLSM and CTSM.

In general, the source signature can be identified by the zeros in the frequency plane, provided the source waveform can be represented by a brief sequence, in



**Fig. 9.26** Comparisons of the estimated direct wave for the three conditions shown in Fig. 9.25 on the preceding page: (a) waveforms of estimated direct wave, and (b) power spectrum of (a) for the three types, *solid*: plucked string, *dotted*: hammer-struck string 1, *broken*: hammer struck 2 from Fig. 17 in [12]

particular, pairs of adjacent time pulses. The next section describes the modeling of zeros created by adjacent time-pulse pairs [14].

## 9.3 Adjacent Pairing Time-Pulse Modeling of Zeros

### 9.3.1 *Formulation of Adjacent Pairing Time-Pulse Modeling (APTM)*

The source waveform is characterized by the troughs arising from zeros that are independent of the poles and widely distributed over the frequency plane. The poles are produced by the sinusoidal functions in the time domain. In contrast, the zeros are given by a set of pairs of adjacent pulses with a single unit of time delay in the time domain. The  $z$ -transform of the time sequence or the transfer function can be represented by zeros, provided the time record has a finite length. The transfer function is factorized so that

$$\begin{aligned} H(z^{-1}) &= K(1 - a_1 z^{-1})(1 - a_2 z^{-1}) \cdots (1 - a_{N-1} z^{-1}) \\ &= \sum_{n=0}^{N-1} h(n) z^{-n} \end{aligned} \quad (9.53)$$

where  $h(n)$  denotes the time sequence (or impulse response) of length  $N$  and  $K$  is a constant. The zeros are given by the roots of the polynomial that define the transfer function, and thus, a single zero can be specified by a pair of time pulses with a unit single delay. It suggests that every zero can be locally identified on the complex frequency plane by solving a set of linear equations.

Consider a time record composed of a pair of adjacent time pulses and with a single-unit delay. Assuming its  $z$ -transform is

$$H(z^{-1}) = 1 - a_1 z^{-1} \quad (9.54)$$

the unknown vector  $\mathbf{a} = (1 \ a_1)^T$  satisfies the linear equation

$$\mathbf{b} = A\mathbf{a} \quad (9.55)$$

where

$$\mathbf{b} = (b_0 \ b_1 \ \dots \ b_{L-1})^T \quad (9.56)$$

denotes the vector composed of the observational recorded response around the spectral trough on the complex frequency plane, and  $A$  is the observational frequency matrix

$$A = \begin{pmatrix} 1 & e^{-i\Omega_0} \\ 1 & e^{-i\Omega_1} \\ \vdots & \vdots \\ 1 & e^{-i\Omega_{L-1}} \end{pmatrix} \quad (9.57)$$

where  $\Omega$  denotes the normalized angular frequency, and  $L \geq 2$ . Thus, the solution is similarly obtained to that from Eq. 9.44 on page 201 based on the least-squares-error criterion [14]

$$\hat{\mathbf{a}} = (A^T A)^{-1} A^T \mathbf{b}. \quad (9.58)$$

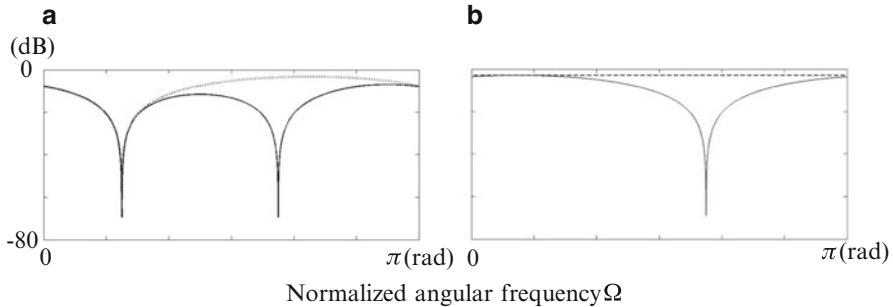
### 9.3.2 Spectral Trough Tracking

The zeros of the transfer function generally create the noticeable spectral troughs in the frequency characteristics. Taking the logarithm of the transfer function in Eq. 9.53 on the previous page yields

$$\hat{H}(z^{-1}) = \log_e H(z^{-1}) = \log_e \sum_{k=1}^{N-1} (1 - a_k z^{-1}) + K'. \quad (9.59)$$

The process for solving the linear equation can be repeated for each zero with respect to the residual so that [14]

$$\hat{H}_i(z^{-1}) - \hat{H}_{i+1}(z^{-1}) = \hat{H}_{i+2}(z^{-1}) \quad (9.60)$$



**Fig. 9.27** Iterative identification of the zeros; (a) *solid*: frequency response by two zeros, *dotted*: identification of zero at lower frequency, (b) *solid*: residual response after subtraction of the first-zero-response in panel (a), *dotted*: identification of second zero, *dashed*: residual from Fig. 3 in [14]

where

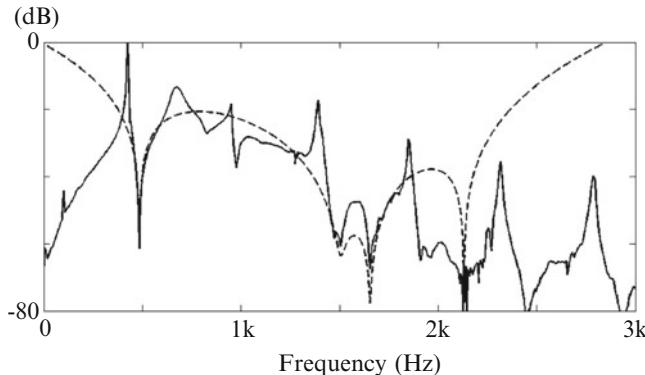
$$\hat{H}_0(z^{-1}) = \hat{H}(z^{-1}). \quad (9.61)$$

Figure 9.27 presents an example where two zeros are contained in the  $z$ -transform, the power spectral characteristics of which is shown by the solid line in panel (a). The dotted line in panel (a) is estimated for the first zero, and the solid line is the residual in panel (b). The dotted line in panel (b) is a synthesized one for the higher zero. No residual response depending on the frequency is given by the dashed line in panel (b), after iteratively solving the two zeros.

### 9.3.3 Representation of Source Signature by Spectral Trough Selection

Figure 9.28 on the following page shows the results of spectral trough selection for a piano-string vibration as well as the solid line [from Fig. 9.7 on page 203 (a)]. The broken line shows the estimated response by the first to fourth zeros, and for reference the solid line illustrates the original frequency response. Repeating the estimating process for zero1–zero4, the frequency response from the four zeros is well synthesized. The troughs representing the source characteristics are well selected.

Source signatures according to the zeros could be informative in condition monitoring and diagnostics. However, a topic suitable for research would be to identify the zeros arising from the source signature separately from those for the transfer function in an environment where sound or vibration is observed. The transfer function of the vibrating system is generally expressed using poles and zeros, and thus the source characteristics can be reflected into the modeling error for



**Fig. 9.28** Identification of zeros for piano-string vibration shown in Fig. 9.7 on page 203; *solid*: original response, *broken*: identified response for four zeros from Fig. 4 in [14]

a transfer function synthesized using CLSM from a signal processing point of view. The example of a piano-string vibration shows that CLSM, CTSM, and APTM are possible approaches to separate the zeros owing to the source waveform separately from those of the transfer function. The zeros representing the source signature are widely distributed and independent of the poles of the transfer function on the frequency domain. A pole corresponds to the sinusoidal wave in the time domain, whereas a zero can be identified by a pair of time pulses with a single unit of time delay. A pair of time pulses is obtained by solving the linear equations for each zero in the frequency plane.

## References

1. M. Tohyama, *Sound and Signals* (Springer, 2011)
2. J. Blauert, N. Xiang, *Acoustics for Engineers* (Springer, 2008)
3. R.H. Lyon, Progressive phase trends in multi-degree-of-freedom systems. *J. Acoust. Soc. Am.* **73**(4), 1223–1228 (1983)
4. R.H. Lyon, Range and frequency dependence of transfer function phase. *J. Acoust. Soc. Am.* **76**(5), 1435–1437 (1984)
5. M. Tohyama, Response statistics of rooms, in *Encyclopedia of Acoustics*, vol. 2, no. 77, ed. by M.J. Crocker (Wiley, 1997), pp. 913–923
6. M. Tohyama, R.H. Lyon, Zeros of a transfer function in a multi-degree-of-freedom system. *J. Acoust. Soc. Am.* **86**(5), 1854–1863 (1989)
7. M. Tohyama, T. Koike, *Fundamentals of Acoustic Signal Processing* (Academic Press, 1998)
8. R.H. Lyon, *Machinery Noise and Diagnostics* (Butterworth, 2000)
9. T. Hasegawa, M. Tohyama, Separation of zeros for source signature identification under reverberant path condition. *J. Acoust. Soc. Am.* **130**(4), EL271–EL275 (2011)
10. M. Kazama, K. Yoshida, M. Tohyama, Signal representation including waveform envelope by clustered line-spectrum modeling. *J. Audio Eng. Soc.* **51**(3), 123–137 (2003)
11. R. Kumaresari, A. Rao, On minimum/maximum/all-pass decomposition in time and frequency domains. *IEEE Trans. SP* **48**, 2973–2976 (2000)

12. T. Hasegawa, M. Tohyama, Analysis of spectral and temporal waveforms of piano-string vibration. *J. Audio Eng. Soc.* **60**(4), 237–245 (2012)
13. H. Nakajima, M. Tanaka, M. Tohyama, Signal representation and inverse filtering using recursive vector projection (in Japanese with English abstract). *J. IEICE Jpn.* **J83-A**(4), 353–360 (2000)
14. T. Hasegawa, M. Tohyama, Source signature identification by using pole/zero modeling of transfer function. *Inter Noise* **2011**, 431616 (2011)

Erratum to

## Waveform Analysis of Sound

Mikio Tohyama

© Springer Japan 2015

M. Tohyama, *Waveform Analysis of Sound*, Mathematics for Industry 3,  
DOI 10.1007/978-4-431-54424-1

DOI 10.1007/978-4-431-54424-1\_10

Chapter 3

In Chapter 3 **Temporal and Spectral Characteristics of Discrete Sequence** in page 30, figure 3.1 and its caption has been updated as below:

Page	Correct																					
p. 30, Fig. 3.1	<p>Decomposition of sequence (period = 8) into odd and even sequences from Fig. 14.3 in [7]</p> <table style="width: 100%; text-align: center;"> <tr> <td style="width: 33%;">Sequence</td> <td style="width: 33%;">Causal</td> <td style="width: 33%;">Non-causal</td> </tr> <tr> <td> </td> <td> <math>=</math> </td> <td> <math>+</math> </td> </tr> <tr> <td></td> <td style="text-align: center;"><math>+</math></td> <td style="text-align: center;"><math>+</math></td> </tr> <tr> <td style="vertical-align: bottom;">even</td> <td> <math>=</math> </td> <td> <math>+</math> </td> </tr> <tr> <td></td> <td style="text-align: center;"><math>+</math></td> <td style="text-align: center;"><math>+</math></td> </tr> <tr> <td style="vertical-align: bottom;">odd</td> <td> <math>=</math> </td> <td> <math>+</math> </td> </tr> <tr> <td></td> <td style="text-align: center;"><math>+</math></td> <td style="text-align: center;"><math>+</math></td> </tr> </table>	Sequence	Causal	Non-causal		$=$	$+$		$+$	$+$	even	$=$	$+$		$+$	$+$	odd	$=$	$+$		$+$	$+$
Sequence	Causal	Non-causal																				
	$=$	$+$																				
	$+$	$+$																				
even	$=$	$+$																				
	$+$	$+$																				
odd	$=$	$+$																				
	$+$	$+$																				

The online version of this chapter can be found at  
[http://dx.doi.org/10.1007/978-4-431-54424-1\\_3](http://dx.doi.org/10.1007/978-4-431-54424-1_3)

The online version of this book can be found at  
<http://dx.doi.org/10.1007/978-4-431-54424-1>

**Chapter 6**

In Chapter 6 **Transfer Function of Linear Systems** in page 126, equation (6.93) has been updated as below:

Page	p. 126 line 10 (eqn. (6.93)) $\sim$ line 11
Error	$+ (2\Re[a])^2 - 2\Re[a](1 +  a ^2)z^{-1} +  a ^2 z^{-2}]$ .
Correct	$+ (2\Re[a])^2 - 2\Re[a](1 +  a ^2)z^{-1} +  a ^2 z^{-2}] \text{ for } Z = e^{j\Omega}$ .

The online version of this chapter can be found at

[http://dx.doi.org/10.1007/978-4-431-54424-1\\_6](http://dx.doi.org/10.1007/978-4-431-54424-1_6)

**Chapter 7**

In Chapter 7 **Sampling Theorem and Discrete Fourier Transform** in page 143, equations (7.20) and (7.21) have been updated as below:

Page	p. 143 eqn. (7.20)
Error	$x(t) = \frac{1}{2}A_0 + \sum_{k=1}^{\infty} (A_k \cos kx + B_k \sin kx)$
Correct	$x(t) = \frac{1}{2}A_0 + \sum_{k=1}^{\infty} (A_k \cos kt + B_k \sin kt)$
Page	p. 143 eqn. (7.21)
Error	$x_N(t) = \frac{1}{2}A_0 + \sum_{k=1}^N (A_k \cos kx + B_k \sin kx)$
Correct	$x_N(t) = \frac{1}{2}A_0 + \sum_{k=1}^N (A_k \cos kt + B_k \sin kt)$

The online version of this chapter can be found at

[http://dx.doi.org/10.1007/978-4-431-54424-1\\_7](http://dx.doi.org/10.1007/978-4-431-54424-1_7)

# Index

## A

- all-pass filter, 33
- analytic expression
  - of the power spectral density function, 53
- analytic signal, 205
- angular frequency
  - normalized angular frequency, 22
- auto-correlation sequence over time domain, 81

## B

- beat, 94
- blind source separation, 63

## C

- causal sequence
  - causal spectral representation, 26
- causal system, 16
- center frequency of band, 39
- cepstrum, 123
  - all-pass cepstrum for complex conjugate pair of non-minimum-phase zeros, 124
  - cepstral decomposition into magnitude and phase parts, 122
  - cepstral decomposition into minimum-phase and all-pass parts, 122
  - cepstral sequence, 116
  - complex cepstral sequence, 116
  - complex cepstrum, 116
  - decomposition of cepstrum, 122
  - even part of minimum-phase cepstrum, 118

- extraction of minimum-phase cepstrum from cepstral sequence, 123
- for complex conjugate pair of minimum-phase zeros, 124
- for complex conjugate pair of non-minimum-phase zeros, 124
- logarithmic expression of transfer function, 116
- logarithmic magnitude and phase functions, 118
- magnitude cepstrum, 117, 118
- minimum-phase cepstrum, 117, 118
- minimum-phase phase cepstrum, 118
- odd part of minimum-phase cepstrum, 118
- power series expansion for logarithmic expression of transfer function, 117
- relationship between the logarithmic magnitude and phase, 117
- closed loop
  - gain, 21
  - impulse response, 21
- CLSM, 216
- clustered time-sequence modeling, 210
- coherent range, 69
- coherent region, 67
- complementarity of time and frequency, 204
- complex function
  - magnitude of complex function, 26
  - phase angle of complex function, 26
- complex time domain, 204, 205, 207
- compound
  - sound, 87
  - wave, 87
- contour integration, 18

convergence  
     of series, 19  
     region of convergence, 17  
 convolution, 14  
     auto-convolution of complex spectral function, 81  
 correlation, 15  
     auto-correlation, 15  
     auto-correlation function over the frequency domain, 81  
     auto-correlation of envelope, 61  
     auto-correlation of flat-spectral sequence, 88  
     auto-correlation of periodic sequence, 32  
     auto-correlation of unit sequence, 52  
     auto-correlation sequence for compound tone, 88  
     auto-correlation sequence for output response through single-echo system, 100  
     auto-correlation sequence for pair of sinusoidal components, 91  
     auto-correlation sequence for response of all-pass system corresponding to input sound, 121  
     auto-correlation sequence of white noise, 75  
     causal auto-correlation, 52  
     complex spectral correlation, 81  
     cross-correlation between a pair of squared envelopes, 79  
     cross-correlation coefficient, 80  
     cross-correlation coefficients between two squared envelopes, 80  
     early portion of auto-correlation sequence, 69  
     frame-wise auto-correlation, 50  
     frame-wise auto-correlation, 49  
     one-side auto-correlation, 52  
     phase correlation function, 34  
     short-term auto-correlation, 49, 50  
     spectral correlation function, 34  
     statistical representation of phase correlation, 36  
     sub-band auto-correlation for direct sound, 61  
     sub-band auto-correlation for direct sound envelope, 61  
     cross-correlation, 15  
 correlation and convolution, 15  
 corresponding pair between time and frequency planes, 209  
 CTSM, 210, 216

**D**  
 damping constant, 192, 193, 195  
 decimation, 152, 158  
 definition, 60  
 Deutlichkeit, 60  
 direct sound, 58  
     direct waveform, 216  
     estimated cyclic waveform, 216  
     estimated direct waveform, 216  
     spatial range of direct sound, 81  
     sub-band component of direct sound, 60  
 Dirichlet kernel, 130, 141, 144  
     analytic expression of Dirichlet kernel, 157

**E**  
 early reflections, 58  
 eigenfrequency, 58, 194  
 energy ratio  
     critical distance of sound field, 60  
     of the direct and other reflected sounds, 59  
     squared average of sum of direct and reflected sound, 60  
 envelope, 33  
     angular frequency, 90  
     angular frequency of squared envelope, 78  
     ensemble average of squared envelope, 78  
     entire spectral envelope, 216  
     for 90-degree phase-shift amplitude modulation, 45  
     frequency of squared envelope, 78  
     instantaneous envelope, 26  
     instantaneous envelope and phase for real sequence, 27  
     of the auto-correlation, 61  
     of the power spectral density function, 53  
     spectral energy of envelope, 72  
     spectral envelope, 52, 69, 216  
     spectrum, 42  
     spectrum of temporal envelope, 72  
     temporal signal dynamics, 72

**F**  
 factorization, 19  
     for transfer function composed of zeros inside the unit circle, 102  
     for transfer function composed of zeros outside the unit circle, 103  
     for transfer function for single echo system, 101

- of transfer function for complex conjugate pair of zeros inside the unit circle, 102
- of transfer function for complex conjugate pair of zeros outside the unit circle, 103
- theorem, 19
- feedback
  - negative feedback system, 22
- filter
  - bank, 140
  - filtering function, 39
  - frequency response of ideal low-pass filter, 127
  - ideal low-pass filter, 126, 132
  - impulse response for  $k$ -th sub-band filtering function, 39
  - impulse response of ideal low-pass filter, 127
  - impulse response of sub-band filter for DFT filter-bank, 140
  - linear phase low-pass filter, 132
  - low-pass filtering, 126
  - perfect reconstruction filter, 40, 41
  - spectral function for ideal low-pass filter by Fourier transform, 127
  - sub-band, 39
  - sub-band filter, 140
  - sub-band filtering function, 39
  - sub-band spectrum, 39
  - sub-band time-sequence, 39
  - truncated spectral function for ideal low-pass filter, 130
- filter bank, 39
- Fouier
  - continuous periodic function expressed by Fourier series expansion, 143
  - partial sum, 143
  - partial sum and Dirichlet kernel, 143
- Fourier
  - coefficient, 144
  - coefficient for sampling function, 145
  - completeness of Fourier transformation, 32
  - derivatives of Fourier transform for real and discrete sequence, 47
  - DFT, 138, 163
  - discrete Fourier transform, 138, 139
  - discrete Fourier transform of sequence, 138
  - Frame-wise Fourier transform, 37
  - frame-wise Fourier transform, 39
  - Gibbs's phenomenon, 132
  - inverse discrete Fourier transform, 139
  - inverse Fourier transform, 23
- inverse Fourier transform for causal sequence, 25
- inverse Fourier transform for continuous periodic function, 138
- inverse Fourier transformation for discrete sequence, 23
- inverse Fourier transformation of finite length of sequence, 23
- partial sum and Dirichlet kernel, 144
- real and imaginary parts of Fourier transform, 23, 24
- sampled sequence in Fourier series expansion, 145
- series expansion, 22
- spectral sequence of sampled time sequence, 146
- sub-band sequence by discrete Fourier transform, 140
- transform for cepstral sequence, 117
- transform for discrete sequence, 138
- transform for minimum-phase cepstrum, 117
- transform of auto-correlation sequence, 31
- transform of even sequence, 24
- transform of odd sequence, 24
- transform of sequence weighted by triangular window, 47
- transform of squared sequence, 33
- transform of squared sequence and spectral correlation sequence, 34
- transformation extended into complex frequency plane, 27
- transformation of finite length of sequence, 23
- Fourier transform, 22
  - of causal auto-correlation sequence, 52
  - of negative-slope triangular window, 48
  - of positive-slope triangular window, 48
  - of short-term auto-correlation sequence, 50
- Fourier transformation, 22
- free oscillation, 192
- frequency
  - angular frequency of damped free oscillation, 192
  - complex frequency, 191, 192
  - complex frequency of the free oscillation, 192
  - instantaneous frequency, 204
  - negative frequency, 26, 149
  - negative instantaneous frequency, 205
  - of free oscillation, 194
  - positive frequency, 26
  - positive instantaneous frequency, 204
  - resonance frequency, 193, 194

frequency characteristic, 192

frequency characteristics, 22

function

- almost periodic function, 186

- even function, 24

- odd function, 24

- polar form of complex function, 26

- quasi-periodic function, 186

fundamental, 33

- frequency, 31, 87

- missing fundamental, 33

- period, 31

## G

generating function, 14

- expanded into series, 17

- for auto-correlation sequence, 16

- generating function for correlation sequence, 15

group delay, 34, 42, 46

- function, 34

- owing to complex conjugate pair of zeros inside unit circle, 107

- owing to complex conjugate pair of zeros outside unit circle, 108

group delay of real and discrete sequence, 47

group velocity, 46

## H

half-power bandwidth, 194

histogram of statistical frequencies of spacings, 173

howling, 20

## I

impulse response, 16, 191

informational masking, 86

inharmonicity, 182

inner product, 179

instantaneous magnitude, 33, 81

instantaneous magnitude and phase, 26

interpolation, 141, 158, 165

- interpolated continuous function from which sequence  $x(n)$  is assumed to be sampled, 141

- interpolated continuous spectral function from which spectral sequence  $X(k)$  is assumed to be sampled, 142

inverse system, 21

## L

least-squares-error criterion, 177

linear combination, 178

linear system, 16

LSE, 177, 178

## M

magnitude

- instantaneous magnitude, 26

magnitude response, 193

modulation

- amplitude modulation and phase difference, 45

- angular frequency of modulation envelope, 90

- carrier of amplitude modulation, 43

- complex modulation index, 81

- complex modulation transfer function, 81

- envelope of amplitude modulation, 43

- index, 46, 78

- instantaneous magnitude and phase with sinusoidal phase difference, 45

- modulated sequence, 78

- multi-sinusoidal modulation, 91

- multiple amplitude modulation, 44

- sinusoidal-amplitude modulation, 42

- sinusoidally amplitude modulation and group delay for sinusoidal sequence, 43

multi-degrees-of-freedom system, 194, 196

## O

octave band, 39

orthogonal, 179

orthogonal projection, 179

orthogonality, 23

- for the complex sinusoidal functions, 138

- of complex sinusoidal function, 23

## P

peak location for envelope, 44

period, 87

periodic, 87

- continuous periodic function composed of finite number of sinusoids, 147

- fundamental and fundamental frequency, 88

- harmonics and harmonic frequencies, 88

- missing fundamental, 89

- period analysis, 32
  - period and fundamental frequency, 88
  - sequence, 31, 87, 138
  - sequence represented by periodic unit-pulse train, 31
  - single cycle of periodic sequence, 31, 32
  - unit-pulse train, 31
  - periodic spectral sequence, 138, 139
  - phase
    - accumulated phase, 112
    - analytic expression of phase frequency response of transfer function composed of complex conjugate pair of zeros inside unit circle, 106
    - analytic expression of phase frequency response of transfer function composed of complex conjugate pair of zeros outside unit circle, 108
  - delay, 46
  - derivative of phase spectrum, 47
  - difference between source and receiver, 46
  - dispersive, 46
  - envelope under linear phase condition, 44
  - frequency response, 105
  - frequency response of phase, 105
  - geometric expression of phase response
    - for complex-conjugate pair of zeros inside unit circle and double poles at the origin, 112
  - geometric expression of phase response for complex-conjugate zero for  $\Phi_a$ , 111
  - geometric expression of phase response for single zero in unit circle, 111
  - instantaneous phase, 26
  - jump due to inside zero, 107
  - jump of  $\pi$ , 107
  - jump owing to outside zero, 109
  - linear phase, 44, 89, 126, 207
  - linear phase in complex time domain, 207
  - linear phase response owing to reciprocal pairs of zeros, 109
  - linear phase response owing to reciprocal zero pairs, 126
  - linear phase trend, 126
  - minimum phase, 199
  - minimum phase in complex time domain, 204
  - minimum-phase complex-time sequence, 209
  - minimum-phase property, 116
  - minimum-phase system, 116
  - non-minimum phase in complex time domain, 205
  - phase spectrum for real sequence, 47
  - principal value of phase, 114
  - total accumulated phase for poles and zeros, 114
  - unwrapped phase, 114
  - velocity, 46
  - wrapped phase, 114
  - phase response, 193
  - pitch, 31
  - place theory, 89
  - pole, 16, 96, 185, 192, 194, 195
    - for transfer function, 96
  - pole line, 195
  - poles and zeros, 191
  - polynomial
    - in complex domain, 19
  - power series, 19
  - power series expansion for logarithmic function, 102
  - power spectral density function over the frequency domain, 81
  - power spectrum
    - for periodic sequence, 32
- R**
- real frequency line (axis), 193
  - remainder function, 196
  - residue, 194, 196
  - residue theorem, 18, 114
  - resonance response, 194
  - resonant frequency, 129
  - response
    - frequency response, 22
  - reverberation, 58
    - decay curve, 75
    - response, 74
  - reverberation time, 130, 192
- S**
- sampling, 13
    - aliasing, 151
    - frequency, 22
    - function, 145
    - period, 13, 22
    - theorem, 148
    - down-sampling, 158
    - frequency, 13
    - up-sampling, 158
  - scaling
    - function, 171
    - structure, 171

- sequence  
 analytic representation for sinusoidal sequence, 26  
 analytic representation of a real sequence, 26  
 analytic sequence, 26  
 causal and stable, 16  
 decomposition of causal sequence into even and odd sequences, 25  
 even sequence, 24  
 harmonic sequences, 92  
 imaginary part of analytic sequence, 26  
 odd sequence, 24  
 right-hand sequence, 16  
 unit-pulse sequence, 16
- series  
 truncation of series, 19
- singing, 20
- single echo  
 frequency response for single echo impulse response, 98  
 impulse response, 96  
 power spectral density function for single echo impulse response, 98  
 transfer function for single echo impulse response, 96
- single-degree-of-freedom system, 192
- singularity, 193
- source signature, 216
- source signature analysis, 216
- spectral correlation, 34
- spectrum  
 continuous spectrum, 22  
 continuous spectrum of discrete sequence, 22  
 discrete sequence representation of continuous spectrum, 23  
 leakage spectral characteristics, 176  
 line spectrum, 32  
 logarithmic power-spectral density function, 104  
 macroscopic characteristics of spectrum, 69  
 magnitude and phase of a spectrum, 29  
 magnitude spectrum, 30  
 of squared sequence, 34  
 original spectrum, 165  
 phase spectrum, 30  
 power spectral density function, 30  
 power spectral density function of output response, 99  
 power spectral density of single cycle of sequence, 32  
 representation of causal spectral function, 26
- short-term power spectrum, 49  
 spectral fine structure, 69  
 spectral peak, 164  
 Spectral peak selection, 219  
 spectral peak selection, 165, 171  
 squared magnitude spectrum for periodic sequence, 32  
 squared magnitude spectrum sampled from its density function, 32  
 true spectrum, 164, 165
- square norm, 178
- steady state  
 frequency characteristics, 58
- T**
- time theory, 89
- time-reversed  
 complex spectrum of time-reversed speech, 86  
 sequence, 86  
 speech, 86
- transfer function, 16, 95, 191, 192  
 all-pass component of transfer function, 119  
 all-pass transfer function, 119  
 decomposition of impulse response into minimum-phase and all-pass responses, 120  
 for even sequence, 126  
 for reciprocal zero pairs, 126  
 for recursive single zero system, 21  
 for recursive system, 20  
 inverse transfer function, 21  
 magnitude of all-pass response for single zero transfer function, 120  
 minimum-phase component of transfer function, 119  
 of recursive system including ideal inverse system, 22  
 room transfer function, 194
- transient  
 auto-correlation of transient response, 62  
 build-up response, 62  
 build-up response to sequence, 62  
 power spectral response for build-up process, 62  
 running correlation, 62  
 transient response, 62
- transpose, 179
- U**
- unit disc, 17

**W**

- wave number, 46
- Weierstrass function, 169
- white noise, 63
- window
  - even triangular sequence, 52
  - function, 164
  - negative-slope triangular window, 48
  - positive-slope triangular window, 48
  - power spectrum of triangular windowed sequence, 48
  - time windowing, 38
  - triangular time window, 47

**Z**

- z-transform, 16, 18
  - convergence of z-transform, 17
  - inverse z-transform, 18
  - of auto-correlation sequence, 31
  - of periodic sequence, 32
  - of single cycle of sequence, 32
  - of unit-pulse train, 31
- zero, 96, 185, 194, 195, 207, 216
  - cancellation by interference, 99
  - decomposition of pole-zero plot, 119
  - double zero, 196, 199
  - for transfer function, 96
  - in time region, 204
  - minimum phase zeros on complex time domain, 204
  - minimum-phase zero, 199
  - non-minimum phase zero in complex time domain, 205
  - non-minimum-phase zero, 199
  - power spectral density function for single zero and its symmetric location, 119
  - single-zero system, 21
  - spectral trough, 216

- spectral trough selection, 219
- symmetric pair of zeros, 196, 198, 207
- zeros
  - decomposition of the square of the magnitude of transfer function into reciprocal complex-conjugate pair of zeros, 102
  - occurrence of zeros, 194
  - on the complex frequency domain, 105
  - power series expansion for logarithmic expression of power spectral density function, 104
  - power series expansion for logarithmic expression of transfer function for complex conjugate pair of zeros inside the unit circle, 103
  - power series expansion for logarithmic expression of transfer function for zeros inside the unit circle, 103
  - power series expansion for logarithmic expression of transfer function for zeros outside the unit circle, 104
  - time series representation for logarithmic expression of transfer function for complex conjugate pair of zeros inside the unit circle, 103
  - time series representation for logarithmic expression of transfer function for zeros inside the unit circle, 103
  - transfer function for conjugate pair of zeros inside outside circle, 107
  - transfer function for conjugate pair of zeros inside unit circle, 106
  - zeros and logarithmic power-spectral density function, 104
  - reciprocal complex-conjugate pair of zeros, 98
  - symmetric pair of zeros, 98
  - symmetric zeros, 98