

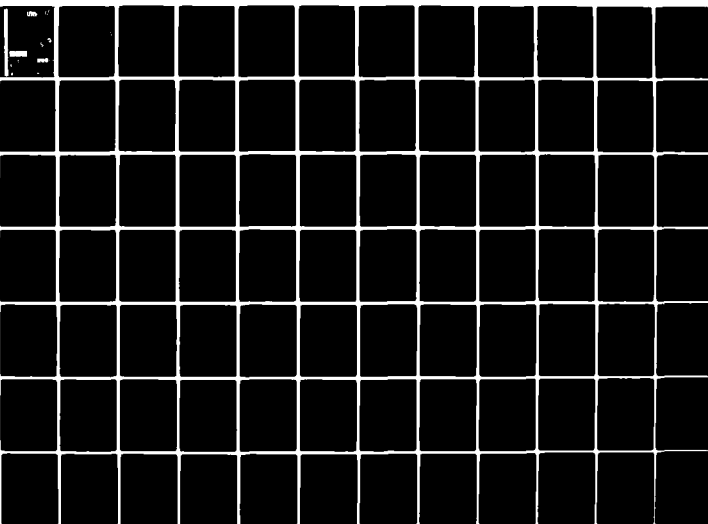
AD-A087 670

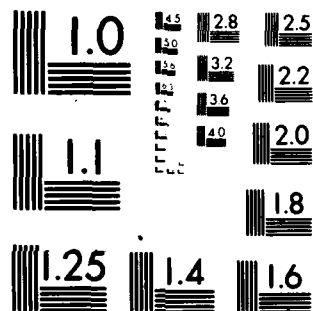
MASSACHUSETTS INST OF TECH CAMBRIDGE RESEARCH LAB OF--ETC F/G 5/8
PHASE ESTIMATION WITH APPLICATION TO SPEECH ANALYSIS-SYNTHESIS.(U)
NOV 79 T F QUATIERI N00014-75-C-0951
TR-491 NL

UNCLASSIFIED

1-2

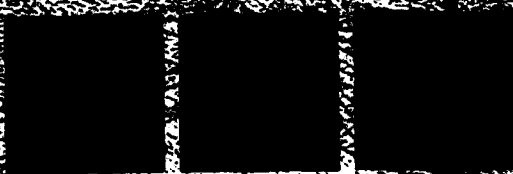
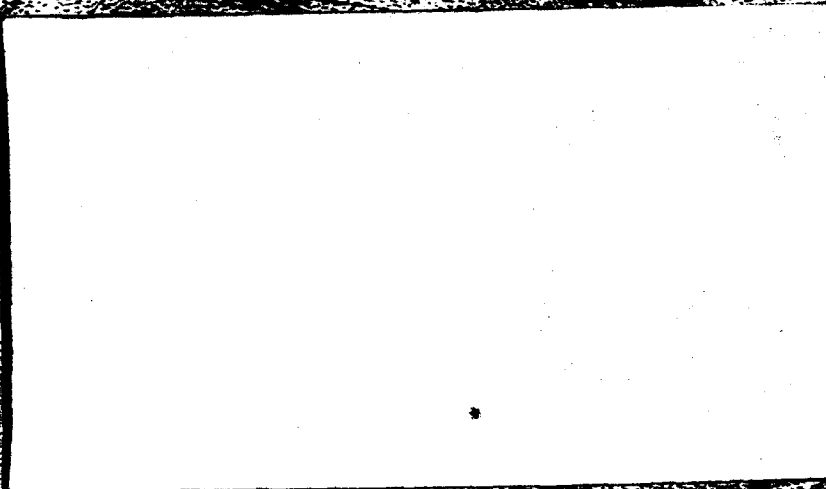
AD-A087 670





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ADA087670



MASSACHUSETTS INSTITUTE OF TECHNOLOGY



9 D. thesis

1

Phase Estimation with Application to Speech Analysis-Synthesis

DTIC

13

Thomas F. Quatieri, Jr.

TECHNICAL REPORT 491

11 Nov 1979

12 14L

Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, Massachusetts 02139

14 TR-491

DTIC
ELECTE
AUG 7 1980
D
C

THE RUTH H. HOOKER
TECHNICAL LIBRARY

FEB

THE RUTH H. HOOKER
TECHNICAL LIBRARY

NAVAL RESEARCH LABORATORY

FEB 1980

NAVAL RESEARCH LABORATORY

15

This work was supported by the Advanced Research Projects Agency monitored by ONR under Contract N00014-75-C-0951/NR 049-328.

This document has been approved for public release and sale; its distribution is unlimited.

304050

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	AD-A087620	
4. TITLE (and Subtitle)		5. TYPE OF REPORT & PERIOD COVERED
PHASE ESTIMATION WITH APPLICATION TO SPEECH ANALYSIS-SYNTHESIS		Technical Report
6. PERFORMING ORG. REPORT NUMBER		
7. AUTHOR(s)		8. CONTRACT OR GRANT NUMBER(s)
Thomas F. Quatieri, Jr.		N00014-75-C-0951
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Research Laboratory of Electronics/ Massachusetts Institute of Technology Cambridge, MA 02139		NR 049-328
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE
Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22217		November 1979
13. NUMBER OF PAGES		14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
137		Office of Naval Research Information Systems Program Code 347 Arlington, Virginia 22217
15. SECURITY CLASS. (of this report)		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
Unclassified		
16. DISTRIBUTION STATEMENT (of this Report)		
Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
Homomorphic Phase Speech		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)		
See other side		

DD FORM 1473

1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. ABSTRACT

This dissertation addresses the problem of estimating the phase of the frequency response of mixed phase signals and systems. A number of techniques are applied to estimation of the phase frequency response of the speech production tract from quasi-periodic speech segments. Methods of phase estimation are categorized as indirect or direct. A subset of the indirect procedures yield a closed form solution for retrieving the phase from the magnitude of a mixed phase frequency response and a priori knowledge about the corresponding signal. Linear iterative algorithms are also developed for retrieving the phase from the magnitude, and, similarly, the magnitude from the phase, with a causality or finite duration constraint imposed on the desired signal. The iterative algorithm for magnitude retrieval provides an alternative to the Hilbert transform for obtaining the magnitude from the phase of a minimum phase signal, but without the need of an unwrapped phase. In addition, it serves as the major component within a new phase unwrapping algorithm which does not require modulo 2π considerations. An alternate indirect strategy changes a phase estimation problem to one of magnitude estimation by modifying a quasi-periodic waveform so that the desired impulse response takes on a minimum phase characteristic. Direct approaches rely on harmonic samples of a frequency response, or of the principal value of its phase. Specifically, time-domain windowing and frequency-domain interpolation are applied to a quasi-periodic waveform in estimating the unwrapped phase at harmonics. Mixed phase estimates from these direct approaches are incorporated within homomorphic and spectral envelope speech analysis-synthesis systems with a second-order improvement in quality over their minimum phase counterparts.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

PHASE ESTIMATION WITH APPLICATION TO SPEECH ANALYSIS-SYNTHESIS

by

Thomas F. Quatieri, Jr.

Submitted to the Department of Electrical Engineering and Computer Science,
on November 20, 1979, in partial fulfillment of the requirements for the
degree of Doctor of Science.

ABSTRACT

This dissertation addresses the problem of estimating the phase of the frequency response of mixed phase signals and systems. A number of techniques are applied to estimation of the phase of the frequency response of the speech production tract from quasi-periodic speech segments. Methods of phase estimation are categorized as indirect or direct. A subset of the indirect procedures yield a closed form solution for retrieving the phase from the magnitude of a mixed phase frequency response and a priori knowledge about the corresponding signal. Linear iterative algorithms are also developed for retrieving the phase from the magnitude, and, similarly, the magnitude from the phase, with a causality or finite duration constraint imposed on the desired signal. The iterative algorithm for magnitude retrieval provides an alternative to the Hilbert transform for obtaining the magnitude from the phase of a minimum phase signal, but without the need of an unwrapped phase. In addition, it serves as the major component within a new phase unwrapping algorithm which does not require modulo 2π considerations. An alternate indirect strategy changes a phase estimation problem to one of magnitude estimation by modifying a quasi-periodic waveform so that the desired impulse response takes on a minimum phase characteristic. Direct approaches rely on harmonic samples of a frequency response, or of the principal value of its phase. Specifically, time-domain windowing and frequency-domain interpolation are applied to a quasi-periodic waveform in estimating the unwrapped phase at harmonics. Mixed phase estimates from these direct approaches are incorporated within homomorphic and spectral envelope speech analysis-synthesis systems with a second-order improvement in quality over their minimum phase counterparts.

THESIS SUPERVISOR: Alan V. Oppenheim

TITLE: Professor of Electrical Engineering

Accession For	
NTIS GSA&I	☒
DDC TAB	☐
Unma. old	☐
Ja. ...	☐
By	
Dist	
A	

ACKNOWLEDGEMENTS

I am most deeply indebted to Professor Alan V. Oppenheim for his enthusiasm, encouragement, and keen intuitive insights throughout this study. His guidance and uncompromising standards contributed greatly to both my intellectual and personal growth and had a major impact on determining the outcome of this work.

I extend my warmest thanks to my parents who encouraged and supported my studies at M.I.T. To them I am most grateful for instilling within me curiosity.

I am also indebted to Professor Jae S. Lim and Dr. Victor Zue for serving as thesis readers and for carefully commenting on this manuscript.

A deserved thanks is given to members of the M.I.T. Digital Signal Processing group for many meaningful discussions and for creating an enjoyable working environment. Appreciation is also extended to Gary Kopec for providing the signal processing language "FANLZ" which allowed spontaneous implementation of ideas which otherwise may have been forever lost. Dr. Ashok K. Gupta must likewise be acknowledged for providing stimulating discussion on phase estimation in modulation theory.

Funding during my doctoral studies was furnished by the Advanced Research Projects Agency (ARPA). I am indebted to them for their support and for making possible interaction with members of the ARPA speech community throughout the United States.

Finally, I would like to acknowledge Betty Larsen for translating a handwritten scrawl into a beautifully typewritten version.

TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT	1
ACKNOWLEDGEMENTS	2
TABLE OF CONTENTS	3
LIST OF FIGURES AND TABLES	6
CHAPTER 1. INTRODUCTION	8
1.1 Scope of Thesis	9
1.2 Phase in Speech	13
1.3 Outline of Chapters	15
CHAPTER 2. A FRAMEWORK FOR PHASE ESTIMATION IN SPEECH ANALYSIS	18
2.1 Deconvolution of a Quasi-Periodic Waveform	19
2.2 Minimum, Maximum, and Mixed Phase Sequences	21
2.2.1 Definitions	21
2.2.2 Hilbert Transform Relations for Minimum Phase Sequences	22
2.2.3 The Unwrapped Phase Function	24
2.3 Sequences with Rational Z-Transforms	26
2.3.1 Minimum and Maximum Phase Rational Z-Transforms	27
2.3.2 Magnitude-Phase Relations for Mixed Phase Rational Z-Transforms	28
CHAPTER 3. PHASE RETRIEVAL FROM MAGNITUDE	32
3.1 Phase Retrieval from Magnitude with Constraints	34
3.1.1 Constraints on the Sequence	34
3.1.1.1 Infinite Length Sequences and the Pade Approximation	35
3.1.1.2 Knowledge of $h(0)$	36
3.1.2 Constraints on the Phase Function	37
3.2 An Iterative Procedure to Retrieve Phase from Magnitude	38
3.2.1 Theory	39
3.2.2 The DFT Realization	44
3.2.2.1 Phase Retrieval from Samples of the Magnitude	44

	<u>Page</u>
3.2.2.2 Iteration Based on Samples of the Magnitude	45
3.2.3 Examples	46
3.3 Phase Estimation from Magnitude by Transformation	47
3.3.1 Conditions For a Minimum Phase Property	50
3.3.1.1 The Nyquist Criterion	51
3.3.1.2 The Positive Real Constraint	55
3.3.1.3 A Sufficient Condition on $h(0)$	57
3.3.2 Transformations For a Minimum Phase Sequence	59
3.3.2.1 Exponential Weighting	60
3.3.2.2 Addition of a Reference Signal	61
CHAPTER 4. MIXED PHASE DECONVOLUTION	63
4.1 Phase Estimation by Homomorphic Deconvolution	65
4.1.1 Theory	65
4.1.2 The Direct Approach	68
4.1.3 The Heuristics of Unwrapped Phase Sensitivity	68
4.1.4 Magnitude-Only Deconvolution by Transformation: An Indirect Approach	70
4.1.5 Examples	74
4.2 Phase Unwrapping by Phase-Only Signal Reconstruction	75
4.2.1 Magnitude Retrieval from Phase with Constraints	77
4.2.1.1 Causality and Phase Continuity	77
4.2.1.2 The Finite Length Constraint	78
4.2.2 An Iterative Procedure to Retrieve Magnitude from Phase	79
4.2.2.1 Theory	79
4.2.2.2 The DFT Realization	84
4.2.2.3 Examples	84
4.2.3 A New Phase Unwrapping Algorithm	88
CHAPTER 5. ESTIMATION OF THE UNWRAPPED PHASE FROM HARMONIC SAMPLES	90
5.1 Techniques of Phase Tracking	91
5.1.1 Linear Interpolation in the Frequency Domain	93
5.1.1.1 Examples	99

	<u>Page</u>
5.1.2 Windowing in the Time Domain	101
5.1.2.1 Examples	104
5.1.3 Comments on Comparing Magnitude and Phase Estimation by Windowing	106
5.2 On the Problem of Alignment	107
5.2.1 Alignment Ambiguity	107
5.2.2 Phase Tracking without Alignment	108
CHAPTER 6. APPLICATION OF PHASE TRACKING TO SPEECH ANALYSIS-SYNTHESIS	110
6.1 Bandwidth-Pitch Period Constraints	111
6.2 Pre- and Post-Alignment	116
6.3 Homomorphic Speech Analysis-Synthesis	118
6.3.1 Minimum Phase Analysis-Synthesis	118
6.3.2 Mixed Phase Analysis	119
6.3.3 Short-Time Reconstruction	120
6.3.4 Informal Listening Tests	121
6.4 Spectral Envelope Speech Analysis-Synthesis	125
6.4.1 Minimum Phase Analysis-Synthesis	125
6.4.2 Phase Envelope Estimation by Linear Interpolation	126
6.4.3 Short-Time Reconstruction	126
6.4.4 Informal Listening Tests	127
CHAPTER 7. SUMMARY AND FUTURE RESEARCH	128
7.1 Summary	128
7.2 Suggestions for Future Research	129
REFERENCES	133
BIOGRAPHICAL NOTE	136

LIST OF FIGURES AND TABLES

<u>Fig. No.</u>	<u>Caption</u>	<u>Page</u>
3.1	Iterative algorithm to recover phase from magnitude	40
3.2	(a) Convergence of $\log[M_k(\omega)]$ in example 3.1 (b) Convergence of $\theta_k(\omega)^k$ in example 3.1	48
3.3	(a) Convergence of $\log[M_k(\omega)]$ in example 3.2 (b) Convergence of $\theta_k(\omega)^k$ in example 3.2	49
3.4	(a) Contour within the z-plane required by the Nyquist criterion (b) Polar plot of $H(z)$ in (3.24)	52
4.1	(a) Polar plot of $E(\omega)$ in (4.9) with a small disturbance (no origin encirclement) (b) Same as (a) with a large disturbance (origin encirclement)	71
4.2	(a) Continuous unwrapped phase of $H(\omega)$ from example 4.1 (b) Estimate of (a) by the direct approach (c) Estimate of (a) by the indirect approach	76
4.3	(a) Unwrapped phase (with a linear phase component) of $H(\omega)$ from example 4.2 (b) Estimate of (a) in the presence of noise by the indirect approach (c) Estimate of (a) in the presence of noise by the direct approach	76
4.4	Iterative algorithm to recover magnitude from phase	80
4.5	(a) Convergence of $\log[M_k(\omega)]$ in example 4.3 (b) Convergence of $\theta_k(\omega)^k$ in example 4.3	86
4.6	(a) Convergence of $\log[M_k(\omega)]$ in example 4.4 (b) Convergence of $\theta_k(\omega)^k$ in example 4.4	87
5.1	(a) Polar plot of $\hat{H}(\omega, n_0)$ for $\Delta\theta_n(\omega_k) < \pi$ (b) Polar plot of $\hat{H}(\omega, n_0)$ for $\Delta\theta_n(\omega_k) > \pi$	96
5.2	(a) Unwrapped phase of a second order all-pass system function (b) Estimate of (a) in example 5.1 with $n_0=0$ and $P=130$ (c) Same as (b) with $P=129$	100

LIST OF FIGURES AND TABLES
(continued)

<u>Fig. No.</u>	<u>Caption</u>	<u>Page</u>
5.2	(d) Same as (b) with $n_0 = -6$ and $P=150$ (e) Same as (d) with $n_0 = -7$	100
5.3	(a) $\hat{H}_r(\omega, n_0)$ derived from the window $w_M(n)$ in (5.20) (b) Approximate unwrapped phase corresponding to (a)	102
5.4	(a) Unwrapped phase of a system function with two complex pole pairs and one complex zero pair (b) Estimate of (a) in example 5.2 with a 100 point Hamming window where $n_0=0$ (c) Same as (b) with a 390 point Hamming window (d) Error function for estimate of (a) in example 5.2 with a 100 point Hamming window where $n_0=-20$	105
6.1	Unwrapped phase of $1-a\exp[j\omega]$	112
6.2	Mixed phase homomorphic analysis-synthesis system	122
6.3	(a) Original speech segment (b) Synthetic speech segment derived from the system in Fig. 6.2	123

Table

6.1	Computed half-power bandwidths required for phase tracking	115
6.2	Mean half-power bandwidths (B_1, B_2, B_3) for the first three resonances of vocal tract configurations. Three male subjects (S_1, S_2, S_3) used eight vowel configurations with two glottal conditions (after House and Stevens[11])	115

CHAPTER 1

INTRODUCTION

In many physical situations, we encounter the problem of recovering a signal which is not directly accessible to measurement, but for which partial knowledge of its Fourier transform can be determined. This partial knowledge may, under certain conditions, be sufficient to recover the entire Fourier transform. For example, when a signal is known a priori to be minimum or maximum phase, the Fourier transform can be recovered within a sign or scale factor from only its magnitude or phase[22].

Within this dissertation, we shall, in particular, consider techniques which use various kinds of partial information in estimating the phase of the Fourier transform of mixed phase discrete-time signals and systems. A number of these techniques are applied specifically to the analysis of voiced speech. That is, we consider estimation of the phase of the frequency response of a signal which consists of the convolution of the combined vocal and nasal tract impulse response, and the glottal wavelet[5]. A simplified model of voiced speech sounds such as vowels, nasalized vowels, and nasal consonants is given by the convolution of this response with a train of equally spaced impulses representing the periodicity of the vocal cord excitation function[5]. This type of representation, i.e. a "quasi-periodic" waveform, is also characteristic of other waveforms found for example in biology, music, and many other acoustic disturbances.

We shall apply a subset of our procedures for phase estimation to two speech analysis-synthesis systems: (i) the homomorphic system proposed by Oppenheim[23], and (ii) the spectral envelope system proposed by Paul[26].

The original schemes introduced either a zero or minimum phase impulse response derived from a magnitude estimate only.

Although many of the phase estimation procedures we consider are developed in the context of speech analysis, our techniques have more general implications and potential use in other areas where phase estimation is required. For example, a number of results may be extended to two-dimensional signals and systems, and thus are possibly applicable to such areas as image or seismic signal processing.

1.1 Scope of Thesis

Most deconvolution techniques for recovering an impulse response from a quasi-periodic waveform are directed primarily to estimating the magnitude of the desired frequency response. Linear prediction[2,15] and homomorphic filtering[23,24], for example, have been applied quite successfully to magnitude estimation.

The phase estimation procedures we shall consider are roughly classified as either "indirect" or "direct". Our class of indirect schemes capitalizes on either knowing the desired magnitude or being capable of deriving an accurate estimate of it by conventional deconvolution techniques. Clearly, when the desired signal is minimum or maximum phase, the phase can be obtained by applying a Hilbert transform to the logarithm of the given magnitude. More generally, we shall use various kinds of a priori information about the desired signal along with the magnitude to unambiguously determine a mixed phase function. In particular, imposing causality or a finite length constraint on the signal and specifying a few samples of the phase, or the first few points of a discrete-time sequence,

in some cases, is sufficient to uniquely characterize the entire phase function. For example, as we shall see, for a certain class of causal sequences only the initial value of the sequence is necessary.

An alternate indirect means of recovering phase from magnitude converts the phase estimation problem to a magnitude estimation problem by modifying the speech waveform so that the desired impulse response takes on a minimum phase characteristic. Specifically, a class of invertible transformations is derived which are suitable to changing the general problem of deconvolution, involving both magnitude and phase estimation of a mixed phase sequence from a quasi-periodic waveform, to a deconvolution problem where only a magnitude estimate is required. This procedure is applied to homomorphic deconvolution[24]. The sensitivities inherent in this deconvolution scheme, due to the requirement of an unwrapped phase [33], are therefore avoided. In addition, it appears that this indirect approach to phase estimation by homomorphic deconvolution is less sensitive to noise disturbances than a direct approach which requires an unwrapped phase. Since most noise reduction systems estimate only the magnitude of a frequency response, this technique is also potentially applicable to signal enhancement[14], where both magnitude and phase estimates are obtained through a magnitude estimate only.

A linear iterative algorithm is also developed for retrieving the phase from the magnitude and a priori information about a desired signal. The algorithm obtains the phase by iteratively imposing the known magnitude function in the frequency domain, and a priori information about the signal in the time domain. The algorithm therefore falls within a class which encompasses, for example, the recently proposed iterative techniques by

Gerchberg[6] and Papoulis[25] for extrapolation of a finite portion of a bandlimited signal.

An analogous iterative procedure is likewise presented for retrieving the magnitude of the frequency response from the phase. When the sequence is minimum phase, applying the Hilbert transform to the unwrapped phase recovers the logarithm of the magnitude (within a scale factor)[22]. Our iterative algorithm, however, when imposing the a priori knowledge of causality, recovers the magnitude from only the principal value of the phase of a minimum phase sequence. The procedure of phase unwrapping is therefore avoided. The algorithm under this particular constraint also serves as the major component within a new phase unwrapping algorithm which does not require the conventional modulo 2π considerations[32]. As we shall see, an unwrapped phase is needed not only in phase estimation by homomorphic deconvolution, but also in other techniques within the thesis.

In parallel with this thesis development, Hayes et al[10] have demonstrated that when imposing a finite length constraint on a sequence, the magnitude of its frequency response in general is uniquely specified by the phase (within a scale factor). As a result, we show that our iterative algorithm can, as well, recover the magnitude from the phase of a finite length mixed phase sequence.

The direct approaches we shall consider do not require an estimate of a magnitude function, but rather require partial knowledge of its phase function which is derived from the phase of the speech waveform. Homomorphic deconvolution is one such approach.

An alternate direct strategy addresses the problem of phase estimation from partial knowledge consisting of harmonic samples of the desired

frequency response or of its principal phase value. Our estimation algorithms attempt to preserve the unwrapped phase at harmonics -- a procedure referred to as "phase tracking". Phase tracking ensures that the unwrapped phase estimate at harmonics equals the unwrapped phase of the desired frequency response. Such a property is desirable since under a band-limited constraint on the unwrapped phase, the entire unwrapped phase may be approximately recovered from only the given harmonic samples. One method of phase tracking invokes an interpolation procedure on harmonic samples, or samples of the principal phase value. In particular, we derive conditions on the desired unwrapped phase under which a simple linear interpolation scheme across two consecutive samples preserves the unwrapped phase at harmonics.

This study leads to a heuristic understanding of the interaction between windowing a speech waveform modeled as exactly periodic over a short time, and the nature of the unwrapped phase of the windowed waveform. In particular, constraints on the duration and position of a specific class of windows are derived for guaranteeing phase tracking by the windowing procedure itself.

Phase estimates derived from the techniques of phase tracking are incorporated within our two speech analysis-synthesis schemes. Constraints on time-domain windowing play a major role in governing the accuracy of the phase estimate within these systems. In the homomorphic scheme tailoring the duration and position of the window is used to improve the phase estimate by homomorphic filtering.

Linear interpolation in the frequency domain can likewise be viewed in the time domain as multiplication by a window. The position of this

window is also important in obtaining an accurate unwrapped phase estimate at harmonics. This consideration will be made in designing a high quality spectral envelope speech analysis-synthesis system with a phase estimate derived from our linear interpolation procedure.

With an alternative interpretation of windowing both analysis-synthesis schemes are shown to be identity systems with respect to reconstruction of a periodic waveform.

Informal listening tests indicate a small, but perceptible improvement in "quality" when in these systems a mixed phase reconstruction replaces its minimum phase counterpart.

1.2 Phase in Speech

The phase of the frequency response of the speech production tract has generally been considered less important than the magnitude function in generating high quality synthetic speech within speech analysis-synthesis systems. Experiments, however, have been reported demonstrating that the envelope of a periodic waveform can be an important factor in audible perception[17]. In particular, it influences sensations of roughness or smoothness. The change from roughness to smoothness may be accomplished by changing the phase or magnitude of a particular component or set of components of a periodic waveform, and the degree of roughness is related to the relative length and depth of the recurrent depressions in the envelope. This "peakiness" factor is thus a determinant in roughness of a periodic waveform.

The ultimate goal in any speech analysis-synthesis system is to extract from the speech waveform in analysis perceptually important information which is used in synthesis to reconstruct the original waveform.

Short-time speech segments are loosely categorized as either voiced or unvoiced. The input-output system model for voiced segments was given in the introduction. Unvoiced speech is likewise modeled over a short-time as the convolution of the combined vocal and nasal tract impulse response with a white noise excitation. The analysis stage usually extracts the magnitude of the Fourier transform of the desired impulse response, a voiced/unvoiced decision, and a pitch measurement for voiced signals. Either a zero phase or a minimum phase function is introduced[2,23].

Both the zero and the minimum phase impulse response estimates are characterized by peakiness. The minimum phase estimate, for example, yields a maximum energy concentration at the signal's origin due to its minimum delay property[22]. Although there does not exist a demonstration of a correlation between peakiness and quality degradation in speech analysis-synthesis systems, the highly quasi-periodic characteristic of speech and perceptual tests on the envelope of a periodic waveform suggests the possibility of "quality" improvement by reduction of peakiness with an accurate phase estimate of the desired response. Two results developed in parallel with this thesis support this conjecture. First, Atal and David [1] have shown that the quality of Linear Prediction Coding (LPC) is improved by introducing an approximation to the LPC error residual. For voiced speech the prediction residual is quasi-periodic with the same period as the speech waveform. A pitch period long segment of the prediction residual can be expressed in terms of a Fourier series expansion as a sum of contributions of the fundamental and the individual harmonics. The contribution of a particular harmonic (e.g. the k th harmonic) is given by a cosine function of a particular frequency, amplitude, and phase:

$A_k \cos(\omega_k n + \theta_k)$. The Fourier series representation allows the variation of the amplitude and phase of different harmonics in any desired fashion. Listening tests indicate a second order, but perceptible change in speech quality for any magnitude condition when the phase changes from zero to the phase derived at each harmonic. Furthermore, there is only a slight difference between the original phase and a fixed frequency-dependent phase. This fixed phase was generated by computing the medium group delay at each harmonic over all pitch periods included in a sentence-like structure.

Gold[8] in preliminary listening tests has also found quality improvement by introducing a phase function within the channel vocoder. A fixed frequency-dependent mixed phase was introduced along with the phase function of a minimum phase vocal tract impulse response estimate derived from LPC.

1.3 Outline of Chapters

The thesis begins in Chapter 2 with developing a framework for the phase estimation problem in the context of speech analysis and discrete-time signals and systems. A review is given of a specific set of results to be used throughout the following chapters.

In Chapter 3 we consider the indirect approach of estimating the phase from the magnitude of a Fourier transform. Constraints and solutions are presented for unambiguous retrieval of the phase from the magnitude function. These constraints are then applied within a linear iterative algorithm for phase retrieval.

In the second portion of this chapter, we investigate methods of converting the phase estimation problem to a magnitude estimation problem. A specific class of invertible transformations is derived for changing an arbitrary mixed phase sequence to a minimum phase sequence, and which are

suitable to preserving the convolutional characteristic of a quasi-periodic waveform.

In Chapter 4 we first review homomorphic deconvolution for directly estimating a phase function from a quasi-periodic waveform. We next describe the heuristics of the sensitivity of this approach due to the need of an unwrapped phase function. Homomorphic deconvolution without an unwrapped phase is then used as an illustration of the indirect approach by transformation of a quasi-periodic waveform given in Chapter 3. Finally in this chapter, we present a new phase unwrapping algorithm which does not require modulo 2π considerations. Toward this objective, we first discuss a class of constraints which guarantees unambiguous magnitude retrieval (within a scale factor) from the phase of a Fourier transform. A linear iterative algorithm, the dual to that in Chapter 3, is then developed to retrieve the magnitude from the phase under such constraints. We illustrate the use of this algorithm by recovering the magnitude of a minimum phase sequence (which yields the same result as would a Hilbert transform), and by recovering also the magnitude of a finite length mixed phase sequence. Finally, this iterative technique is used as the major component in our new phase unwrapping algorithm.

In Chapter 5 we take a direct approach to phase estimation within a different context than found in Chapters 3 and 4. Specifically, we assume only harmonic samples of the desired frequency response are available. Techniques of phase tracking are developed, which involve both linear interpolation in the frequency domain, and windowing in the time domain.

In Chapter 6 techniques of phase tracking derived in Chapter 5 are applied in designing the two previously described high quality speech

analysis-synthesis systems with mixed phase. Finally in this chapter results of informal listening tests are presented, where mixed and minimum phase reconstructions are compared.

Lastly, in Chapter 7 a summary of the main results of the thesis is given. We also suggest a direction and some potential areas of future research.

CHAPTER 2

A FRAMEWORK FOR PHASE ESTIMATION IN SPEECH ANALYSIS

In this chapter a framework is developed for the phase estimation problem in speech analysis in the context of discrete-time signals and systems. We first present the general problem of estimating the magnitude and phase of the Fourier transform of a sequence $h(n)$ which is convolved with a train of equally spaced pulses $p(n)$: $x(n) = h(n) * p(n)$. This formulation is applicable to voiced speech segments modeled over a short duration by a "quasi-periodic" waveform which is produced by exciting the vocal and nasal tract with pulses of air, i.e., the glottal wavelet, caused by vibration of the vocal cords[5]. We shall assume that the desired sequence $h(n)$ consists of the convolution of the impulse response of the vocal and nasal tract and glottal wavelet.

The phase estimation component of this deconvolution problem is often tied to magnitude estimation. In fact, for a certain class of sequences the phase function can be derived from the magnitude which is often more easily and directly measurable than phase. We are therefore led to a review of minimum and maximum phase sequences which fall within this class, and for which the Hilbert transform relations exist between the continuous phase and logmagnitude of the Fourier transform. These relations enable the phase to be reconstructed from the magnitude and, likewise, the magnitude from the phase. In general, without additional constraints neither the magnitude nor the phase is sufficient to completely characterize a mixed phase sequence.

Since the phase of a complex number is a multivalued function, it is generally not unique. With appropriate constraints, however, a unique function can be derived and is termed the unwrapped phase. Besides being a requirement in the Hilbert transform relations for minimum and maximum phase sequences, the unwrapped phase is also an important component in the theory of homomorphic deconvolution, and in other direct phase estimation procedures.

Our primary objective in this chapter is to collect a specific set of results which are useful in the chapters to follow. We will avoid a detailed discussion of subtle mathematical issues, and present only those results necessary in formulating certain techniques of phase estimation.

2.1 Deconvolution of a Quasi-Periodic Waveform

Consider a sequence $x(n)$ which consists of the convolution of a sequence $h(n)$ and $p(n)$ a train of equally spaced pulses:

$$x(n) = h(n) * p(n) \quad (2.1)$$

where $p(n)$ is given by

$$p(n) = \sum_k a(k) \delta(n-kP) \quad (2.2)$$

and where P is the pulse spacing (i.e., the "pitch period" in the context of voiced speech), $\delta(n)$ is the unit-sample sequence[22], and k may range over finite or infinite extent.

From physical considerations $h(n)$ is assumed to be causal, i.e., $h(n)=0$ for $n<0$, and stable, i.e., $\sum_n |h(n)| < \infty$. The z -transform of $h(n)$ is given by

$$H(z) = \sum_{n=0}^{\infty} h(n)z^{-n} \quad (2.3)$$

and has a region of convergence which encompasses the unit circle and the entire z -plane outside the unit circle, including $z = \infty$. $H(z)$ is referred

to as the system function, and when evaluated on the unit circle yields the frequency response of the system: $H(z)|_{z=\exp[j\omega]} = H(\omega)$. In general, $H(\omega)$ is complex and can be expressed in terms of its real and imaginary parts as

$$H(\omega) = H_r(\omega) + j H_i(\omega) \quad (2.4)$$

where r and i denote real and imaginary respectively. In terms of its magnitude and phase (i.e., polar form), (2.4) is expressed as

$$H(\omega) = |H(\omega)| \exp[j\theta_h(\omega)] \quad (2.5)$$

where $\theta_h(\omega) = \arg[H(\omega)]$.

With $X(z)$, $H(z)$, and $P(z)$ denoting the z -transforms of $x(n)$, $h(n)$, and $p(n)$, respectively, it follows from (2.1) that $X(z)$ can be written as

$$X(z) = H(z) P(z) \quad (2.6)$$

$X(z)$ represents the z -transform of a stable sequence when $p(n)$ is absolutely summable, i.e., $\sum_n |p(n)| < \infty$, and in particular when $p(n)$ is bounded and of finite extent so that the range of k is finite.

The Fourier transform of $x(n)$ is found by evaluating $X(z)$ on the unit circle and in polar form is expressed by

$$\begin{aligned} X(\omega) &= X(z)|_{z=\exp[j\omega]} = |X(\omega)| \exp[j\theta_x(\omega)] \\ &= |H(\omega)| |P(\omega)| \exp[j\theta_h(\omega)] \exp[j\theta_p(\omega)] \end{aligned} \quad (2.7)$$

where, $\theta_x(\omega) = \arg[X(\omega)]$ and $\theta_p(\omega) = \arg[P(\omega)]$.

The goal of deconvolution of the quasi-periodic sequence of (2.1) is to extract the magnitude and phase of $H(\omega)$ and $P(\omega)$. Estimation of $|H(\omega)|$ has been extensively investigated and is generally an easier problem than

estimation of $\theta_h(\omega)$. Extraction of $P(\omega)$ is often not required in speech analysis but rather only the period P is sought -- a problem denoted as "pitch detection". Within this dissertation, we shall restrict ourselves to recovering the phase of $H(\omega)$.

2.2 Minimum, Maximum, and Mixed Phase Sequences

In this section, we formulate the definitions and properties of minimum, maximum, and mixed phase sequences. The unwrapped phase function is also defined and its characteristics are discussed.

2.2.1 Definitions

The definitions of minimum and maximum phase sequences are completely analogous to each other and are given below.

Minimum Phase Sequences: A complex function $H(z)$ of a complex variable z is minimum phase if it is analytic and its reciprocal $H^{-1}(z)$ is also analytic for $|z| \geq 1$ in the z -plane. A minimum phase sequence is then defined as a sequence whose z -transform is minimum phase. It follows from this definition that a necessary, but not sufficient, condition for $h(n)$ to be minimum phase is that it be causal, stable, and nonzero at $n = 0$. That is,

$$h(n) = 0 \quad n < 0 \quad (2.8.a)$$

$$h(n) \neq 0 \quad n = 0 \quad (2.8.b)$$

$$\sum_n |h(n)| < \infty \quad (2.8.c)$$

Maximum Phase Sequences: A complex function $H(z)$ of a complex variable z is maximum phase if it is analytic and its reciprocal $H^{-1}(z)$ is also analytic for $|z| \leq 1$ in the z -plane. A maximum phase sequence is defined as a sequence whose z -transform is maximum phase. Such sequences are anti-causal,

stable, and nonzero at $n = 0$. That is,

$$h(n) = 0 \quad n > 0 \quad (2.9.a)$$

$$h(n) \neq 0 \quad n = 0 \quad (2.9.b)$$

$$\sum_n |h(n)| < \infty \quad (2.9.c)$$

Conditions 2.9.a, b, and c, however, are necessary, but not sufficient, for a maximum phase characteristic.

Mixed Phase Sequences: A complex function $H(z)$ of a complex variable z which is neither minimum nor maximum phase is termed mixed phase. A mixed phase sequence has a z -transform which is mixed phase, is stable, and may or may not be causal or anti-causal.

2.2.2 Hilbert Transform Relations for Minimum Phase Sequences

A consequence of the properties of a minimum phase sequence is a useful relationship between the logarithm of $|H(\omega)|$ and $\theta_h(\omega)$. Specifically, the Hilbert transform relations[22] provide a means of retrieving the phase from the magnitude, and the magnitude (within a scale factor) from the phase for such sequences.

To derive these relations consider the complex logarithm of $H(z)$ given by

$$\begin{aligned} \hat{H}(z) &= \log[H(z)] \\ &= \log|H(z)| + j \arg[H(z)] \end{aligned} \quad (2.10)$$

If $\hat{H}(z)$ is viewed as the z -transform of a real sequence $\hat{h}(n)$, then when $\hat{h}(n)$ is causal, $\hat{h}(n)$ can be completely recovered from its even component $\hat{h}_e(n) = (\hat{h}(n) + \hat{h}(-n))/2$, or its odd component $\hat{h}_o(n) = (\hat{h}(n) - \hat{h}(-n))/2$ and $\hat{h}(0)$ [22]. That is, $\hat{h}(n)$ is given by

$$\hat{h}(n) = \begin{cases} \hat{h}_e(0) & n = 0 \\ 2\hat{h}_e(n) & n > 0 \\ 0 & n < 0 \end{cases} \quad (2.11a)$$

or

$$\hat{h}(n) = \begin{cases} \hat{h}(0) & n = 0 \\ 2\hat{h}_o(n) & n > 0 \\ 0 & n < 0 \end{cases} \quad (2.11b)$$

The Fourier transform of $\hat{h}_e(n)$ is $\log |H(\omega)|$, and the Fourier transform of $\hat{h}_o(n)$ is $\arg[H(\omega)]$. Therefore, a consequence of (2.11a) and (2.11b) is that the imaginary component of $\hat{H}(\omega)$, $\arg[H(\omega)]$ can be recovered from its real component, $\log |H(\omega)|$. Likewise, the real component, $\log |H(\omega)|$ can be recovered from its imaginary component, $\arg[H(\omega)]$ within the additive constant $\hat{h}(0)$. In fact, it is possible to obtain direct relations between $\log |H(\omega)|$ and $\arg[H(\omega)]$, i.e., the Hilbert transform relations. Nevertheless, magnitude and phase reconstruction is generally easier to perform indirectly through (2.11a) and (2.11b).

The requirement that $\log |H(\omega)|$ and $\arg[H(\omega)]$ be a Hilbert transform pair is often referred to as the minimum phase condition. It corresponds to the requirement that the sequence $\hat{h}(n)$ be causal and stable and can be shown to be equivalent to the definition of a minimum phase sequence given in the preceding section. The analogous case with $\hat{h}(n)$ anti-causal and stable corresponds to the maximum phase condition.

2.2.3 The Unwrapped Phase Function

For $\hat{h}(n)$ of the preceding section to be causal and stable, $\hat{H}(z)$ must be analytic in the region $|z| \geq 1$. In considering this analyticity, we must appropriately define $\arg[H(z)]$ since any multiple of 2π can be added to the phase without affecting the value of $H(z)$, and thus $\arg[H(z)]$ in general will be discontinuous.

The phase is therefore ambiguous to within a 2π multiple, and can be expressed by

$$\arg[H(z)] = \text{ARG}[H(z)] \pm 2\pi q \quad (2.12a)$$

where $q = 0, 1, 2, \dots$ and

$$-\pi < \text{ARG}[H(z)] \leq \pi \quad (2.12b)$$

$\text{ARG}[H(z)]$ is termed the principal value of $\arg[H(z)]$. This ambiguity is resolved by the fact that analyticity of $\hat{H}(z)$ implies that its real and imaginary parts must be continuous functions of z , and consequently if $\hat{H}(z)$ is to be analytic, we must define $\arg[H(z)]$ to be a continuous function. Furthermore, since $\hat{h}(n)$ is assumed real, $\arg[H(z)]$ will be defined so that for $z = \exp[j\omega]$ it is odd, periodic in ω with period 2π , and a continuous function of ω [32]. A phase function which satisfies these properties is termed the unwrapped phase function.

One approach to computing the unwrapped phase is to assume that a continuous phase function is obtained by integration of the phase derivative. Assuming a differentiable complex logarithm in (2.10), and evaluating the logarithmic derivative on the unit circle, we obtain

$$\begin{aligned} \hat{H}'(\omega) &= H'(\omega)/H(\omega) \\ &= \hat{H}'_r(\omega) + j \hat{H}'_i(\omega) \end{aligned} \quad (2.13)$$

where the prime denotes differentiation with respect to ω . From (2.13) we obtain

$$\begin{aligned}\hat{H}'_i(\omega) &= \arg'[H(\omega)] \\ &= \frac{H_r(\omega)H'_i(\omega) - H_i(\omega)H'_r(\omega)}{H_r^2(\omega) + H_i^2(\omega)}\end{aligned}\quad (2.14)$$

Integrating (2.14) ensures that $\arg[H(\omega)]$ is a continuous function of ω . In addition, to ensure that $\arg[H(\omega)]$ be odd and periodic in ω with period 2π , the following two conditions must be met:

$$\arg[H(\omega)] \Big|_{\omega=0,\pi} = 0 \quad (2.15)$$

Since,

$$H(\omega) \Big|_{\omega=0} = \sum_n h(n) \quad (2.16)$$

and

$$\arg[H(\omega)] \Big|_{\omega=\pi} = \int_0^\pi \arg'[H(\omega)] d\omega \quad (2.17)$$

it follows that only sequences with a positive mean and a zero mean phase derivative are compatible with the above requirements.

We define the linear phase component of $H(z)$ as z^{n_0} , where $n_0 = \arg[H(\pi)]/\pi$ and where $\arg[H(\pi)]$ is given in (2.17). Thus, it is perhaps necessary that $h(n)$ be inverted and shifted (removing the linear phase component corresponds to a shift [22]) in order that (2.15) be satisfied.

Nevertheless, it will be useful to modify our definition of unwrapped phase to encompass a linear phase component. With the presence of z^{n_0} , the

unwrapped phase is defined over the half-open interval $[0, \pi)$ by integration of the phase derivative, and is continuous in this interval. However, to be odd, it must be discontinuous at π [32].

Since many of the techniques developed in this thesis will be implemented on a digital computer, discrete Fourier transform (DFT) implementations are required. In particular, we need to determine samples of the continuous phase function. One technique of obtaining samples of the unwrapped phase first computes the principal value of the unwrapped phase using inverse tangent routines and then "unwraps" by simply adding appropriate multiples of 2π to the principal value until the discontinuities induced by the modulo 2π operation are removed[32]. A more recent technique has been proposed that combines the information in both the phase derivative and the principal value of the phase into an adaptive numerical integration unwrapping scheme[34]. The method adds appropriate multiples of 2π to the principal value of the phase until the phase is "consistent" with the numerically integrated phase derivative.

2.3 Sequences with Rational z-Transforms

In this section we restrict ourselves to sequences whose z-transform is given by a rational function of the form

$$H(z) = Az^{n_0} \frac{\prod_{k=1}^{m_1} (1 - a_k z^{-1}) \prod_{k=1}^{m_0} (1 - b_k z)}{\prod_{k=1}^{p_1} (1 - c_k z^{-1}) \prod_{k=1}^{p_0} (1 - d_k z)} \quad (2.18)$$

where $|a_k|$, $|b_k|$, $|c_k|$, and $|d_k|$ are less than or equal to unity and z^{n_0} is the linear phase component. Factors of the form $(1 - a_k z^{-1})$ and $(1 - c_k z^{-1})$ correspond to zeros and poles on or inside the unit circle, and the factors

$(1-b_k z)$ and $(1-d_k z)$ correspond to zeros and poles on or outside the unit circle.

2.3.1 Minimum and Maximum Phase Rational z-Transforms

We shall interpret the results of the previous sections in the context of this more restrictive class of sequences. Since $\hat{H}(z)$ of (2.10) (or equivalently $H(z)$ and $H^{-1}(z)$) must be analytic in the region $|z| \geq 1$, for $H(z)$ to be a minimum phase function, there can be no poles or zeros of $H(z)$ on or outside the unit circle. This includes poles and zeros at infinity; i.e., if $H(z)$ is minimum phase $\lim_{z \rightarrow \infty} H(z)$ must be a nonzero finite constant. The implication of this restriction is that factors of the form $(1-b_k z)$ and $(1-d_k z)$ do not exist, and a linear phase component is not present.

A in (2.8) may be positive or negative when $H(z)$ is minimum phase. However, for $H(z)$ to be compatible with the requirement that there exists an odd, continuous phase function, A must be positive.

The z-transform of a normalized (i.e., $A=1$) minimum phase sequence can therefore be expressed as

$$H_{\min}(z) = \frac{\prod_{k=1}^{m_i} (1-a_k z^{-1})}{\prod_{k=1}^{p_i} (1-c_k z^{-1})} \quad (2.19)$$

A completely analogous formulation can be made for maximum phase sequences for which poles and zeros lie outside the unit circle. The corresponding normalized z-transform is given by

$$H_{\max}(z) = \frac{\prod_{k=1}^{m_o} (1-b_k z)}{\prod_{k=1}^{p_o} (1-d_k z)} \quad (2.20)$$

Therefore, a mixed phase sequence with no zeros on the unit circle is expressed by

$$h(n) = A h_{\min}(n) * h_{\max}(n) * \delta(n-n_0) \quad (2.21)$$

where $h_{\min}(n)$ and $h_{\max}(n)$ correspond to (2.19) and (2.20), and $\delta(n-n_0)$ accounts for a linear phase component. A mixed phase sequence in general, however, may have zeros on the unit circle.

2.3.2 Magnitude-Phase Relations for Mixed Phase Rational z-Transforms

In this section we consider the problem of reconstructing the magnitude from the phase, and the phase from the magnitude of a Fourier transform, for a mixed phase rational z-transform.

Suppose we are given the magnitude of the Fourier transform of a causal sequence. A phase function can be derived indirectly through (2.11a) to obtain a minimum phase z-transform denoted by $H_{\text{mp}}(z)$. The poles (assumed inside the unit circle) of $H_{\text{mp}}(z)$ are the poles of the original z-transform $H(z)$, and zeros inside the unit circle are also left intact. However, zeros of $H(z)$ outside the unit circle are mapped to their conjugate reciprocal locations[22]. The original z-transform $H(z)$ can therefore be represented by the cascade of a minimum phase system and an all-pass system, $A(z)$: $H(z) = H_{\text{mp}}(z)A(z)$, where an all-pass is defined as a system for which the magnitude of the frequency response is unity for all ω . In particular, an arbitrary all-pass $A(z)$ can be shown to consist of a cascade of factors of the form:

$$\left[\frac{1 - a^*z}{1 - az^{-1}} \right]^{-1} \quad (2.22)$$

where $|a| < 1$.

Consequently, such systems have the property that their poles and zeros occur at conjugate reciprocal locations.

Since the original sequence $h(n)$ is causal, knowledge of the magnitude of $H(\omega)$ uniquely specifies the poles of $H(z)$ which from above are equal to the poles of $H_{mp}(z)$. Therefore, with knowledge of the existence of M zeros of $H(z)$ there exists a maximum of 2^M different phase functions (excluding linear phase) for a given magnitude function. These phase functions can be generated by reflecting zeros about the unit circle through (2.22). Without additional a priori knowledge the exact location of the zeros and thus the original phase cannot be determined.

Consider now the dual problem of recovering the magnitude from the phase of the z -transform on the unit circle. We assume the linear phase component is removed to satisfy the phase continuity condition for the Hilbert transform relations. We shall see that the ambiguity in obtaining the magnitude function is of a different nature from the problem of phase retrieval from magnitude.

As before, the minimum phase counterpart to the original z -transform is found by applying (2.11b). To see the resulting pole-zero pattern consider the all-pass function

$$H_{ap}(z) = H(z)/H(z^{-1}) \quad (2.23)$$

Evaluating $H_{ap}(z)$ on the unit circle, we obtain

$$H_{ap}(z)|_z = \exp[j\omega] = \exp[j2\theta_h(\omega)] \quad (2.24)$$

where $\theta_h(\omega)$ is the known phase function. (2.24) can be written as

$$H_{ap}(\omega) = \left[\frac{H_{\min}(\omega)}{H_{\max}(-\omega)} \right] \cdot \left[\frac{H_{\max}(\omega)}{H_{\min}(-\omega)} \right]$$

$$= \tilde{H}_{\min}(z) \cdot \tilde{H}_{\max}(z) \Big|_{z=\exp[j\omega]} \quad (2.25)$$

where

$$\tilde{H}_{\min}(z) = H_{\min}(z)/H_{\max}(z^{-1}) \quad (2.26a)$$

and,

$$\tilde{H}_{\max}(z) = H_{\max}(z)/H_{\min}(z^{-1}) \quad (2.26b)$$

Since $\tilde{H}_{\min}(z)$ and $\tilde{H}_{\max}(z)$ both have identical phase $\theta_h(\omega)$, on the unit circle, $\tilde{H}_{\min}(z)$ corresponds to the minimum phase counterpart of $H(z)$ with phase $\theta_h(\omega)$. From (2.26a), the poles and zeros inside the unit circle of $H(z)$ remain intact, while the zeros outside the unit circle have been reflected as poles to their conjugate reciprocal locations.

The original z -transform $H(z)$ can therefore be represented by the cascade of a minimum phase system and a zero-phase system $B(z)$ for which the phase of the frequency response is zero for all ω : $H(z) = \tilde{H}_{\min}(z)B(z)$. In particular, an arbitrary zero-phase system $B(z)$ can be shown to consist of a cascade of factors of the form

$$[(1 - az^{-1})(1 - a^*z)]^{\pm 1} \quad (2.27)$$

Given that there exist no poles outside the unit circle and that the original z -transform has a total of L poles inside the unit circle or zeros outside the unit circle, there exists a maximum of 2^L different magnitude functions for a given phase function. These magnitude functions can be

generated by reflecting poles to zeros and zeros to poles through the zero-phase function of (2.27). For example, a maximum phase zero, represented by the factor $(1-bz)$, where $|b| < 1$, is reflected to a minimum phase pole by the operation:

$$(1 - b^*z^{-1})^{-1} = (1 - bz)[(1 - b^*z^{-1})(1 - bz)]^{-1} \quad (2.28)$$

Therefore, without additional a priori knowledge, the magnitude function cannot be uniquely specified from the phase.

CHAPTER 3

PHASE RETRIEVAL FROM MAGNITUDE

We have seen in the preceding chapter that for minimum or maximum phase sequences the phase of a Fourier transform is uniquely recoverable from its magnitude. When a sequence does not fall within this class, it is reasonable to seek alternative a priori knowledge about the sequence which is sufficient to unambiguously retrieve the phase from the magnitude function.

For example, we may know a priori that the zeros of a rational z -transform $H(z)$ lie outside the unit circle and that the corresponding sequence is causal and nonzero at the origin. The zeros of the minimum phase z -transform, with the given magnitude on the unit circle, from (2.22), fall at conjugate reciprocal locations to those of the original z -transform. $H(z)$ and consequently the phase of $H(\omega)$ can therefore be retrieved[10].

In this chapter we shall consider two categories of constraints:

- (i) Constraints on values of points of the sequence
- (ii) Constraints on the values of samples of the phase function

In a number of cases these constraints lead to a set of linear equations for obtaining parameters of a rational z -transform, and thus an indirect means of retrieving the phase function.

An alternative method of phase retrieval invokes a linear iterative procedure which incorporates known values of the sequence and the known magnitude function. This iterative algorithm is particularly useful when a linear solution does not exist for the given constraints.

Recently there has been a great deal of interest in iterative algo-

ithms for signal reconstruction. Gerchberg[6] and Papoulis[25], for example, have developed an iteration for the determination of a bandlimited continuous-time signal $x(t)$ in terms of a finite segment of $x(t)$. Gerchberg and Saxton[7] utilize the magnitude of a complex signal and the magnitude of its Fourier transform in an iterative fashion to obtain the phase functions in both the time and frequency domains. Mersereau and Schafer[20] also have recently performed a comparative study of various iterative deconvolution algorithms. These techniques impose a time-limited, band-limited, and positivity constraint on the output of the deconvolution.

Such algorithms and, likewise, the algorithm proposed within this chapter all fall within a class where information about the sequence and its Fourier transform is iteratively imposed. In the next chapter, we describe an iterative algorithm, also of this class, which recovers the magnitude of a Fourier transform from its phase.

When constraints of the above kind are not specified or accuracy of the magnitude is uncertain, we consider the alternative procedure of converting a phase estimation problem to a magnitude estimation problem by modification of the speech waveform. The second main area of this chapter, in particular, investigates methods to transform a mixed phase sequence to a minimum phase sequence. With an appropriate invertible transformation a magnitude function is then sufficient to completely characterize the modified sequence. Such transformations are useful in changing the general problem of deconvolving a mixed phase sequence from a quasi-periodic waveform to a deconvolution problem where only an estimate of the magnitude of the Fourier transform of the modified sequence is required. The invertibility of the transformation allows approximate recovery of the desired phase

function.

A second area of potential application of this technique is in signal enhancement[14]. We can modify a sequence to take on a minimum phase characteristic before degradation by additive noise. Most noise reduction systems such as spectral subtraction estimate only the magnitude of the Fourier transform[13]. Transformation of a sequence allows both magnitude and phase estimation through only a magnitude estimate.

3.1 Phase Retrieval From Magnitude with Constraints

When a sequence $h(n)$ is causal, stable and has a rational z -transform, $H(z)$ is given by

$$H(z) = \frac{Az^{n_0} \prod_{k=1}^{m_i} (1-a_k z^{-1}) \prod_{k=1}^{m_o} (1-b_k z)}{\prod_{k=1}^{p_i} (1-c_k z^{-1})} \quad (3.1)$$

where $|a_k|$, $|b_k|$, $|c_k|$, and $|d_k|$ are less than unity. We argued in section 2.3.2 that (neglecting the linear phase component z^{n_0}) there exists a maximum of 2^M possible phase functions when $|H(\omega)|$ is given, where $M=m_i+m_o$ is the number of zeros in (3.1). We shall now consider constraints on $h(n)$ and $\theta_h(\omega)$ to resolve this phase ambiguity. These constraints are not exhaustive, but indicative of the nature of requirements for phase recovery.

3.1.1 Constraints on the Sequence

One method of guaranteeing a unique phase function is to constrain values of points of the sequence. In section 3.1.1.1, we specify the first $M+1$ points of $h(n)$ to remove phase ambiguity. Alternatively, in section 3.1.1.2 we demonstrate that for a restricted class of sequences, the ini-

tial value of $h(n)$, $h(0)$ is sufficient for a unique phase determination.

3.1.1.1 Infinite Length Sequences and the Pade Approximation

The first method we consider borrows the philosophy of the Pade approximation of a rational z-transform[18]. In this technique the parameters of the rational model are chosen to exactly match the first $p_i + m_i + m_0 + 1$ points of the sequence, where p_i , m_i , and m_0 are given in (3.1). In a similar manner, we shall show that when the first $m_i + m_0 + 1$ points of the sequence are matched and the magnitude is given, the parameters of the rational model can be uniquely determined, and thus a phase function unambiguously specified.

Assuming that $h(0) \neq 0$, and expanding the numerator and denominator functions of (3.1) as polynomials in z^{-1} , we obtain

$$\begin{aligned} H(z) &= N(z)/D(z) \\ &= \sum_{k=0}^M \hat{a}(k)z^{-k} / \sum_{k=0}^N \hat{b}(k)z^{-k} \end{aligned} \quad (3.2)$$

where $M = m_i + m_0$ and $N = p_i$. When the magnitude of $H(\omega)$ is known, from section 2.3.2, the denominator polynomial $D(z)$ of $H(z)$ can be determined[18]. Cross multiplying by $D(z)$, we can express (3.2) in the time domain as a convolution:

$$\begin{aligned} \hat{a}(n) &= \hat{b}(n) * h(n) \\ &= \sum_{k=0}^N \hat{b}(k)h(n-k) \end{aligned} \quad (3.3)$$

Given the coefficients $\hat{b}(n)$ and $h(n)$ for $n=0,1,2,\dots,M$, the numerator coefficients $\hat{a}(n)$ are easily computed from (3.3).

3.1.1.2 Knowledge of $h(0)$

When $h(n)$ is causal and $h(0) \neq 0$, $H(z)$ in (3.1) is expressed by

$$H(z) = \frac{A \prod_{k=1}^{m_1} (1 - a_k z^{-1}) \prod_{k=1}^{m_0} (z^{-1} - b_k)}{\prod_{k=1}^{p_1} (1 - c_k z^{-1})} \quad (3.4)$$

The zeros of (3.4) may be flipped inside and outside the unit circle by an all-pass function $A(z)$ consisting of factors represented by

$$[(z^{-1} - a^*) / (1 - a z^{-1})]^{+1} \quad (3.5)$$

This particular form maintains $|H(\omega)|$, while preserving the causality of the original sequence $h(n)$. (3.5) also implies a nonzero initial value.

Therefore, from the Initial Value Theorem[22], for each zero geometry, with $|H(\omega)|$ held constant, the initial value $\hat{h}(0)$ is given by

$$\begin{aligned} \hat{h}(0) &= \lim_{z \rightarrow \infty} H(z)A(z) \\ &= A \prod_{k \in I} (-a_k) \prod_{k \in O} (-b_k) \end{aligned} \quad (3.6)$$

where $A(z)$ is an all-pass function, and where I represents the set of integers which correspond to maximum phase factors of the form $(z^{-1} - a_k)$, and O represents the set of integers which correspond to maximum phase factors of the form $(z^{-1} - b_k)$. Thus given arbitrary sets I_1 and O_1 , and I_2 and O_2 , each possible zero geometry corresponds to a distinct $h(0)$ under the constraint[21]

$$\begin{aligned} &\prod_{k \in I_1} (-a_k) \prod_{k \in O_1} (-b_k) \\ &\neq \prod_{k \in I_2} (-a_k) \prod_{k \in O_2} (-b_k) \end{aligned} \quad (3.7)$$

The constraints represented by (3.7) and knowledge of $h(0)$ have neither a simple meaningful interpretation, nor a linear solution, as does matching the first $M+1$ points of $h(n)$. Nevertheless, they provide a flavor for the requirements necessary for phase retrieval.

In summary, when (3.7) is satisfied, $|H(\omega)|$ is given, and $H(z)$ is of the form in (3.4), a unique phase function is guaranteed for each distinct $h(0)$.

3.1.2 Constraints on The Phase Function

Consider the situation where the magnitude is known and samples of the phase function are given. We shall show that when the number of samples of the phase in the frequency interval $[0, \pi]$ is greater than the order of the numerator M in (3.1), the phase is uniquely specified.

Given samples of $\arg(H(z))$ on the unit circle at $\omega = \omega_0, \omega_1, \dots, \omega_M$, samples of $H(z)$ can also be determined on the unit circle since $|H(\omega)|$ is known. Cross-multiplying by $D(z)$ in (3.2) with $z = \exp[j\omega_m]$, we obtain

$$\sum_{k=0}^M \hat{a}(k) \exp[-jk\omega_m] = H(\omega_m) D(\omega_m) \quad (3.8)$$

With $M+1$ samples of the phase function (3.8) represents $M+1$ equations in $M+1$ unknowns. The real component of this set of complex linear equations is given in matrix form by

$$\begin{array}{c} \uparrow \\ M+1 \\ \downarrow \end{array} \begin{array}{c} \leftarrow M+1 \rightarrow \\ \left[\begin{array}{cccc} 1 & \cos(\omega_0) & \cos(2\omega_0) & \dots \cos(M\omega_0) \\ 1 & \cos(\omega_1) & \cos(2\omega_1) & \dots \\ 1 & \cos(\omega_2) & \cos(2\omega_2) & \dots \\ \vdots & \vdots & \vdots & \ddots \\ 1 & \cos(\omega_m) & \dots & \cos(M\omega_m) \end{array} \right] \end{array} \begin{array}{c} \left[\begin{array}{c} \hat{a}(0) \\ \hat{a}(1) \\ \vdots \\ \hat{a}(M) \end{array} \right] \\ \bar{\mathbf{a}} \end{array} = \begin{array}{c} \left[\begin{array}{c} N_r(0) \\ N_r(1) \\ \vdots \\ N_r(M) \end{array} \right] \\ \bar{\mathbf{N}}_r \end{array} \quad (3.9a)$$

$\underbrace{\hspace{15em}}_C$

or,

$$C\bar{\mathbf{a}} = \bar{\mathbf{N}}_r \quad (3.9b)$$

where $N_r(m) = \text{Re}(H(\omega_m)D(\omega_m))$.

McClellan and Parks[19] have demonstrated that the set of functions $\{1, \cos(\omega), \cos(2\omega) \dots \cos(M\omega)\}$ is a Chebyshev set for $\omega \in [0, \pi]$. An implication of this result is that when $0 \leq \omega_m \leq \pi$ and the samples ω_m are distinct, the matrix C is invertible. There exists, therefore, a unique set of numerator coefficients $\hat{a}(k)$ and consequently an unambiguous phase function.

In summary, when $|H(\omega)|$ is known, $h(n)$ is causal, and $M+1$ distinct phase samples are given in the interval $[0, \pi]$, the phase function is unique. Clearly, a consequence of this result is that any finite segment of the phase function is also sufficient to remove ambiguity since $M+1$ phase samples can always be obtained from such a segment.

3.2 An Iterative Procedure to Retrieve Phase From Magnitude

In this section a linear iterative procedure is developed for phase

retrieval. The method invokes values of the sequence and the magnitude of the Fourier transform which guarantee an unambiguous phase function when the z-transform is assumed rational.

3.2.1 Theory

The iterative algorithm will now be described and a proof is given showing that a defined error must decrease monotonically as the algorithm iterates.

The algorithm is illustrated in Figure 3.1. We begin with an initial guess $\theta_0(\omega)$ of the desired phase, and inverse transform the function $M(\omega)\exp[j\theta_0(\omega)]$, where $M(\omega)=|H(\omega)|$ is the given magnitude. This step yields $h_0(n)$, the initial estimate of $h(n)$. Next, the known values of $h(n)$ for $n \in I$ (where I is the set of integers for which $h(n)$ is given) are incorporated in the initial estimate $h_0(n)$ to obtain $\tilde{h}_0(n)$. The magnitude of the Fourier transform of $\tilde{h}_0(n)$ is then replaced by the given magnitude and the procedure is repeated. The steps involved in one iteration are summarized below. $h_k(n)$, $\theta_k(\omega)$, and $M_k(\omega)$ are the signal, phase, and magnitude estimates, respectively on the k th iteration and $\tilde{h}_{k+1}(n)$ is defined by

$$\tilde{h}_{k+1}(n) = \begin{cases} h_k(n) & n \notin I \\ h(n) & n \in I \end{cases} \quad (3.10)$$

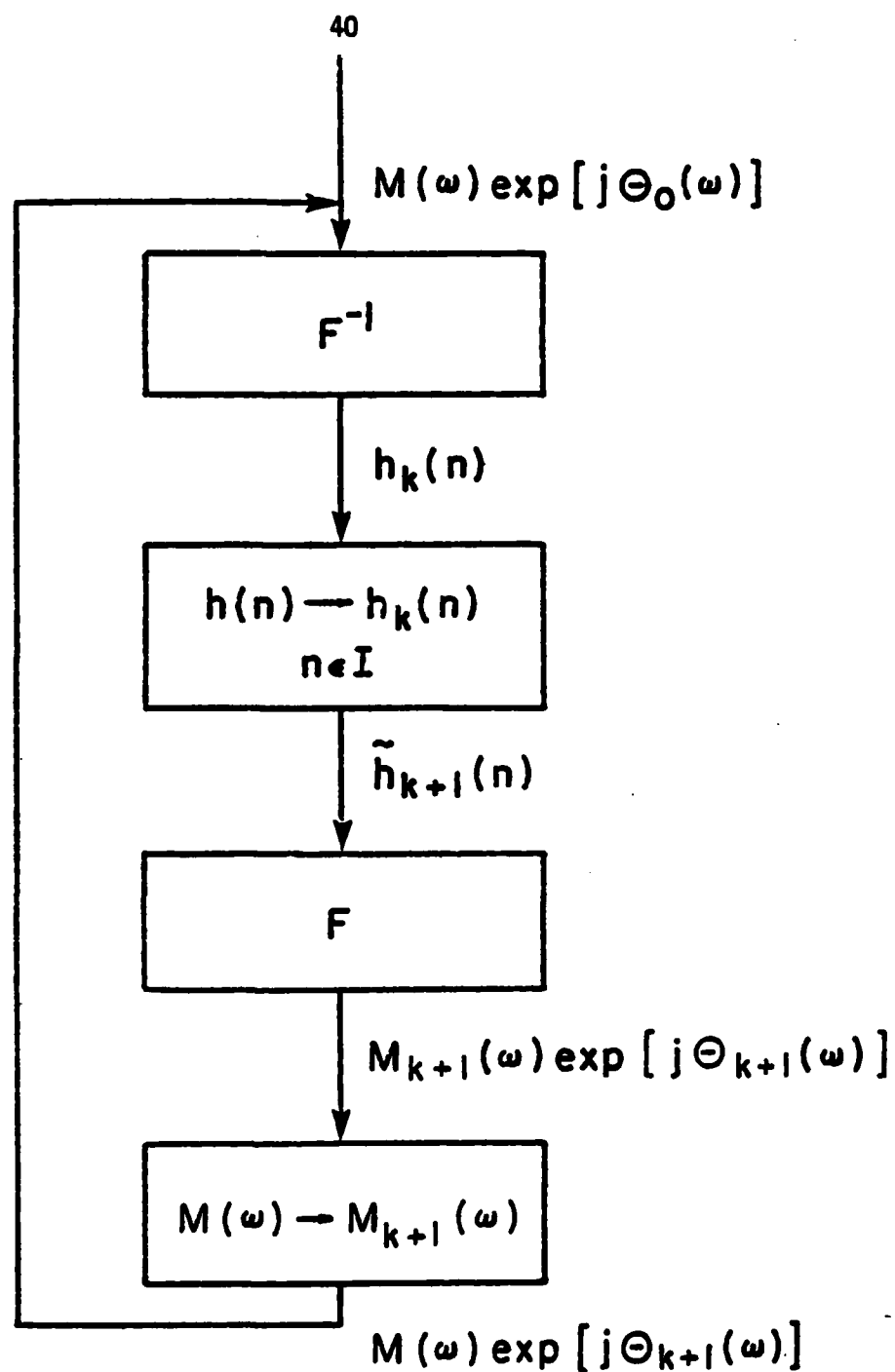


Fig. 3.1 Iterative algorithm to recover phase from magnitude

Iteration to Recover Phase From Magnitude

- (i) Inverse transform $M(\omega)\exp[j\theta_k(\omega)] : h_k(n)$
 - (ii) Replace $h_k(n)$ with $h(n)$ for $n \in I : \tilde{h}_{k+1}(n)$
 - (iii) Forward transform $\tilde{h}_{k+1}(n) : M_{k+1}(\omega)\exp[j\theta_{k+1}(\omega)]$
 - (iv) Replace $M_{k+1}(\omega)$ by $M(\omega) : M(\omega)\exp[j\theta_{k+1}(\omega)]$
 - (v) Repeat
-

The above steps complete one iteration and the algorithm is continued for as many desired number of iterations. In order to demonstrate that the algorithm results in a "reasonable" phase estimate, we wish to choose an error function which is monotone decreasing through each pass of the algorithm. One error function we have considered is the sum of the squared differences between $h(n)$ and the estimate $h_k(n)$ on each iteration:

$$E_k = \sum_n |h(n) - h_k(n)|^2 \quad (3.11)$$

Clearly, from (3.10) the error decreases in the time domain whenever $h_k(n) \neq h(n)$ for $n \in I$, and stays the same whenever $h_k(n) = h(n)$ for $n \in I$. That is,

$$E_k \geq \sum_n |h(n) - \tilde{h}_{k+1}(n)|^2 \quad (3.12)$$

However, in the frequency domain it is not possible to show through a vector argument that $|H(\omega) - \tilde{H}_{k+1}(\omega)|$ monotonically decreases at each frequency ω when the known magnitude is incorporated in $\tilde{H}_{k+1}(\omega)$, the Fourier transform of $\tilde{h}_{k+1}(n)$.

For this reason we choose an alternative error function which can be

easily shown to monotonically decrease upon each pass of the algorithm. The error function is defined as the mean squared difference between the known magnitude and the estimate $M_k(\omega)$ on each iteration:

$$E_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} |M(\omega) - M_k(\omega)|^2 d\omega \quad (3.13)$$

We shall show in two steps that E_k is monotone decreasing: $E_k \geq E_{k+1}$.

Error Reduction in the Time Domain

With the identity $|\exp[j\theta_k(\omega)]|^2 = 1$, the expression for E_k in (3.13) can be written as

$$\begin{aligned} E_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |M(\omega) - M_k(\omega)|^2 |\exp[j\theta_k(\omega)]|^2 d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |M(\omega) \exp[j\theta_k(\omega)] - M_k(\omega) \exp[j\theta_k(\omega)]|^2 d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_k(\omega) - \tilde{H}_k(\omega)|^2 d\omega \end{aligned} \quad (3.14)$$

where $H_k(\omega)$ and $\tilde{H}_k(\omega)$ are the Fourier transforms of $h_k(n)$ and $\tilde{h}_k(n)$, respectively. From Parseval's Theorem (3.14) is given in the time domain by

$$E_k = \sum_n |h_k(n) - \tilde{h}_k(n)|^2 \quad (3.15)$$

From (3.10) it follows that

$$|h_k(n) - \tilde{h}_{k+1}(n)| = \begin{cases} 0 & n \notin I \\ |h_k(n) - h(n)| & n \in I \end{cases} \quad (3.16)$$

and therefore

$$|h_k(n) - \tilde{h}_k(n)| \geq |h_k(n) - \tilde{h}_{k+1}(n)| = 0 \quad n \notin I \quad (3.17a)$$

and

$$|h_k(n) - \tilde{h}_k(n)| = |h_k(n) - \tilde{h}_{k+1}(n)| \quad n \in I \quad (3.17b)$$

Summing (3.17a) and (3.17b) over all n , we obtain

$$E_k = \sum_n |h_k(n) - \tilde{h}_k(n)|^2 \geq \sum_n |h_k(n) - \tilde{h}_{k+1}(n)|^2 \quad (3.18)$$

Error Reduction in the Frequency Domain

From the triangle inequality for vector differences:

$$|H_k(\omega) - \tilde{H}_{k+1}(\omega)| \geq |H_k(\omega)| - |\tilde{H}_{k+1}(\omega)| \quad (3.19)$$

Therefore, we have from Parseval's Theorem and (3.19),

$$\begin{aligned} E_k &\geq \sum_n |h_k(n) - \tilde{h}_{k+1}(n)|^2 \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_k(\omega) - \tilde{H}_{k+1}(\omega)|^2 d\omega \\ &\geq \frac{1}{2\pi} \int_{-\pi}^{\pi} |M(\omega) - M_{k+1}(\omega)|^2 d\omega \\ &= E_{k+1} \end{aligned} \quad (3.20)$$

Since E_k is monotone decreasing and $E_k \geq 0$, it has a lower bound of zero. Consequently, it must converge to a limit point[30].

Although we have shown that the error E_k is monotone decreasing, we have not shown that $\lim_{k \rightarrow \infty} E_k = 0$ nor that $\lim_{k \rightarrow \infty} \theta_k(\omega) = \theta_h(\omega)$, i.e. that $\theta_k(\omega)$

converges to $\theta_h(\omega)$ at each frequency ω . In spite of the possibility that $\lim_{k \rightarrow \infty} \theta_k(\omega) \neq \theta_h(\omega)$ often, in practice, this situation does not occur when the constraints of section 3.1.1 are imposed.

3.2.2 The DFT Realization

Since our iterative algorithm will be implemented on a digital computer, we can compute a Fourier transform at only a finite number of points. In particular, we shall use the discrete Fourier transform (DFT). It is important then to investigate the implications of this implementation.

We first demonstrate that under certain conditions uniformly spaced samples of the magnitude of the Fourier transform (i.e., samples corresponding to the DFT) are sufficient for phase retrieval.

3.2.2.1 Phase Retrieval from Samples of the Magnitude

Suppose we are given the magnitude of the Fourier transform of $h(n)$ and sufficient a priori knowledge about the sequence so that a unique phase function can be specified. We wish to show that when our a priori knowledge includes a finite duration constraint of M points, M or more uniformly spaced samples of $|H(\omega)|$ at $\omega_k = 2\pi k/N$, $\forall \omega_k$ in the interval $[0, \pi]$, are sufficient to recover the phase of the Fourier transform of $h(n)$.

To derive this result, we examine the autocorrelation function $R(n) = h(n) * h(-n)$ whose Fourier transform is given by $|H(\omega)|^2$. Since $h(n)$ is of length M , the symmetric function $R(n)$ is of length $2M-1$. At uniformly spaced samples of the squared magnitude, we have

$$\left. |H(\omega_k)|^2 \right|_{\omega_k = 2\pi k/N} = \sum_{n=-M}^M R(n) \exp[j2\pi kn/N] \quad (3.21)$$

When $N \geq 2M-1$, (3.21) is simply the DFT of $R(n)$. Therefore, $2M-1$ or more samples of $|H(\omega)|^2$ in the interval $[0, 2\pi]$ uniquely characterizes both $R(n)$ and $|H(\omega)|^2$. From symmetry, this condition is equivalent to specifying M or more samples in the interval $[0, \pi]$ [22].

Consider now two M point sequences $h_1(n)$ and $h_2(n)$. Furthermore, suppose that they are recoverable from their respective magnitude functions and values of the sequences over a set I (e.g., the constraints of section 3.1.1), and that these values are identical for each sequence: $h_1(n) = h_2(n) = h(n)$, $n \in I$. Suppose also that their magnitude functions are equal at M or more uniformly spaced points $\omega_k = 2\pi k/N \forall \omega_k$ in the interval $[0, \pi]$. Since both $R_1(n)$ and $R_2(n)$ are $2M-1$ points in extent and are recoverable from the given magnitude samples, it follows from above that $R_1(n) = R_2(n)$ and that $|H_1(\omega)|^2 = |H_2(\omega)|^2$, $\forall \omega$. But we have assumed both $h_1(n)$ and $h_2(n)$ are recoverable from their magnitude functions and the values $h(n)$ for $n \in I$. Therefore, $h_1(n)$ and $h_2(n)$ must be identical.

That is, contrary to our original assumption, there can exist only one M point sequence which satisfies the given a priori knowledge in the time domain, and has a magnitude function with specified values at M or more uniformly spaced points in the interval $[0, \pi]$.

3.2.2.2 Iteration Based on Samples of the Magnitude

From the results of the previous section, we are now in a position to formulate the iterative procedure of section 3.2.1 in terms of a DFT realization. The error function is defined as the mean squared difference between samples of the known magnitude and samples of the estimate:

$$E_k = \frac{1}{N} \sum_{\ell=0}^{N-1} |M(\ell) - M_k(\ell)|^2 \quad (3.22)$$

where N is the DFT length. $h(n)$ is assumed causal and of duration M , so that $N \geq 2M-1$.

Following steps (3.14) through (3.20) and with the application of forward and inverse DFTs we obtain,

$$\begin{aligned}
 E_k &= \frac{1}{N} \sum_{\ell=0}^{N-1} |M(\ell) \exp[j\theta_k(\ell)] - M_k(\ell) \exp[j\theta_k(\ell)]|^2 \\
 &= \frac{1}{N} \sum_{\ell=0}^{N-1} |H_k(\ell) - \tilde{H}_k(\ell)|^2 \\
 &= \sum_{n=0}^{N-1} |h_k(n) - \tilde{h}_k(n)|^2 \\
 &\geq \sum_{n=0}^{N-1} |h_k(n) - \tilde{h}_{k+1}(n)|^2 \\
 &= \frac{1}{N} \sum_{\ell=0}^{N-1} |H_k(\ell) - \tilde{H}_{k+1}(\ell)|^2 \\
 &\geq \frac{1}{N} \sum_{\ell=0}^{N-1} |M(\ell) - M_{k+1}(\ell)|^2 \\
 &= E_{k+1}
 \end{aligned} \tag{3.23}$$

Thus, as shown for the continuous counterpart, $E_k \geq E_{k+1}$.

3.2.3 Examples

In this section we investigate two examples which illustrate the iterative algorithm for recovering phase from magnitude. The initial phase estimate $\theta_0(\omega)$ in both cases is set to zero.

Example 3.1

Consider a causal infinite length mixed phase sequence with $h(0) \neq 0$. The first $M+1$ points of the sequence are constrained where M is the number of zeros of $H(z)$ which consists of two complex pole pairs at 292 Hz and 3500 Hz and one maximum phase complex zero pair at 2000 Hz. Consequently, the first three points of $h(n)$ are necessary.

Since a finite length sequence is required, we assume $h(n)=0$ for $n \geq 256$ and the DFT length is set at 512 points. This approximation does not noticeably alter $H(\omega)$. Nevertheless, the magnitude of the Fourier transform of the truncated sequence was incorporated in the iteration.

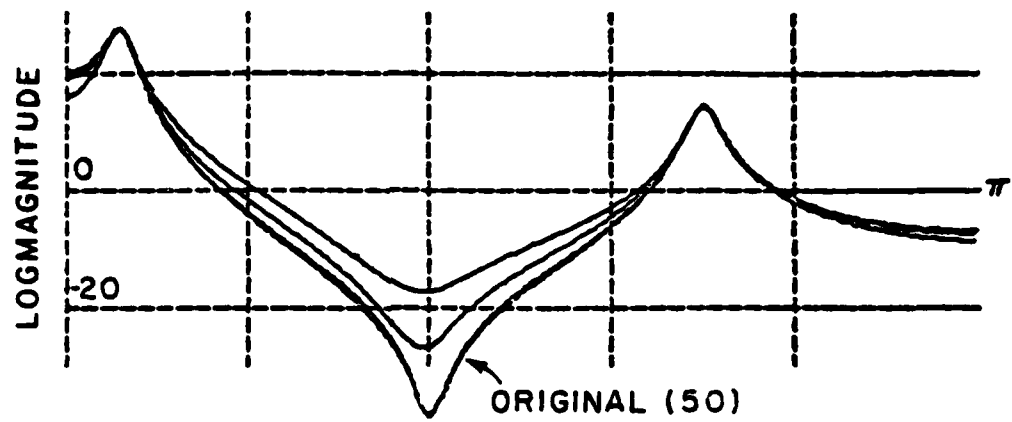
The sequence of functions $\log[M_k(\omega)]$ and $\theta_k(\omega)$ are depicted in Fig. 3.2 superimposed on the originals for 5, 10, 20 and 50 iterations. $\log[M_k(\omega)]$ and $\theta_k(\omega)$ are indistinguishable from the originals after 50 iterations.

Example 3.2

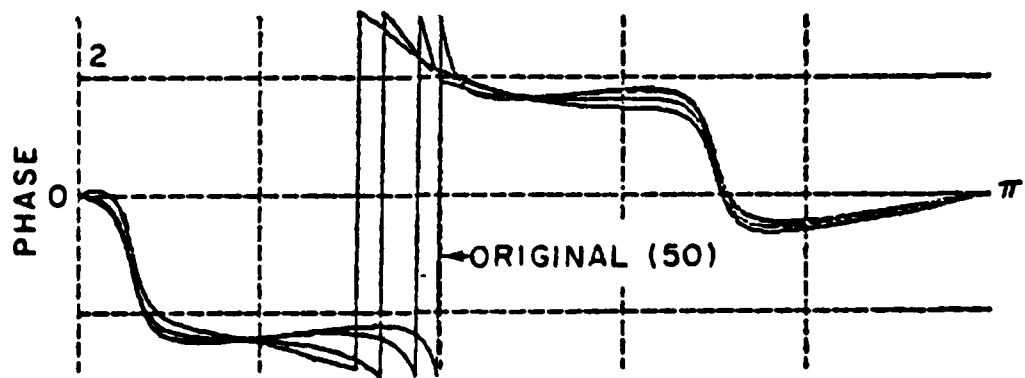
Consider a causal mixed phase finite length sequence where only the first point of $h(n)$ is constrained and the zero geometry of $H(z)$ satisfies the condition of (3.7). The original sequence is eight points in duration, and a 512 point DFT is used. The sequence of functions $\log[M_k(\omega)]$ and $\theta_k(\omega)$ are depicted in Fig. 3.3, superimposed on the corresponding originals for 1, 2, 4, 10, 80 and 200 iterations. $\log[M_k(\omega)]$ and $\theta_k(\omega)$ are indistinguishable from the originals after 200 iterations.

3.3 Phase Estimation From Magnitude by Transformation

In the previous sections we presented methods of retrieving the phase from the magnitude function with sufficient a priori information. An alternative approach which requires only a magnitude function is to convert a



(a)



(b)

*Fig. 3.2 (a) Convergence of $\log[M_k(\omega)]$ in example 3.1, (b) Convergence of $\theta_k(\omega)$ in example 3.1.

* Logmagnitude and phase functions are measured in decibels and radians, respectively.

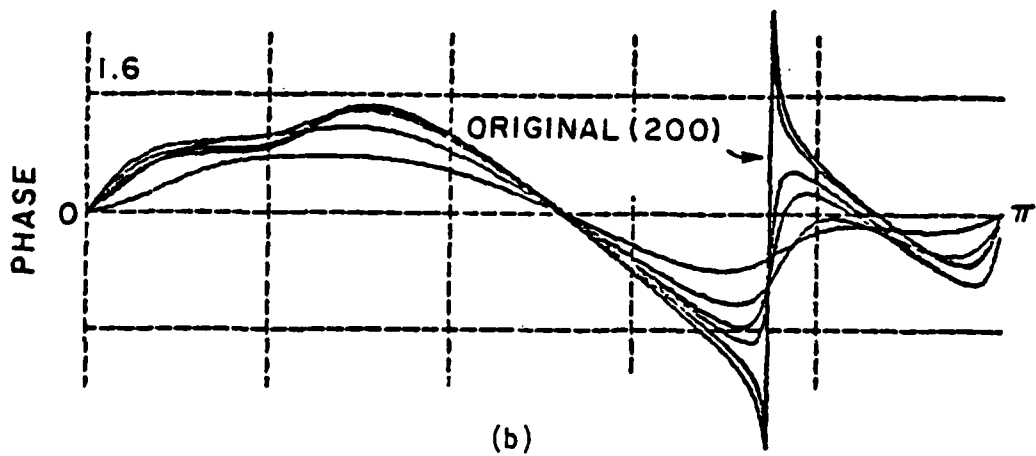
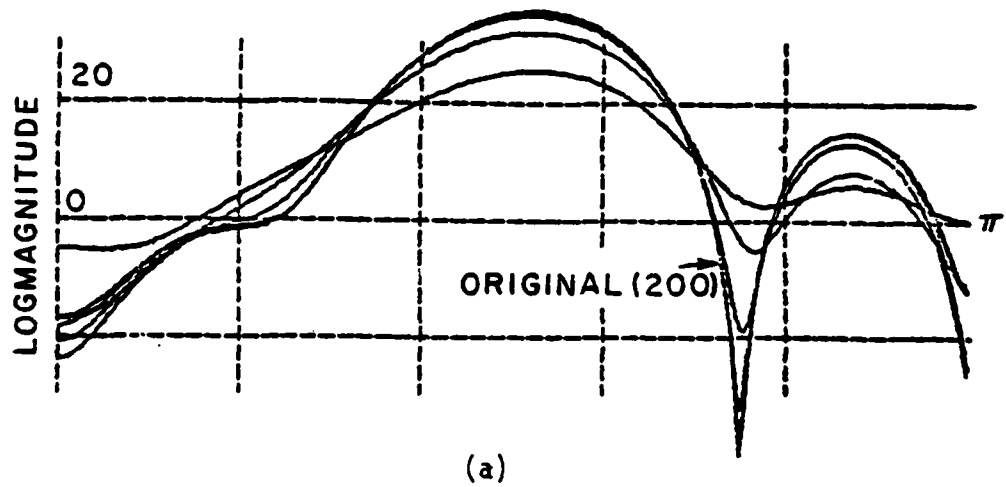


Fig. 3.3 (a) Convergence of $\log[M_k(\omega)]$ in example 3.2, (b) Convergence of $\theta_k(\omega)$ in example 3.2.

phase estimation problem to a magnitude estimation problem by transforming the desired sequence to take on a minimum phase characteristic. This approach does not require the strict constraints given in section 3.1, but does require some general knowledge about the sequence or its Fourier transform.

In the context of deconvolution of a quasi-periodic waveform, we shall see that our class of transformations can be applied to modify a desired mixed phase sequence while preserving the convolutional characteristic of the quasi-periodic waveform. After obtaining an estimate of the magnitude of the modified (minimum phase) sequence, we can obtain an estimate of its phase by applying a Hilbert transform. Performing, finally, the inverse operation to the original transformation yields an estimate of the desired sequence, and hence an estimate of its phase.

3.3.1 Conditions for a Minimum Phase Property

It is desirable to detect a minimum phase characteristic by some simple operation on the sequence or its Fourier or z-transform so that appropriate modifications for minimum phase conversion are easily derived. One method of testing whether a sequence with a rational z-transform is minimum phase is to find the locations of its poles and zeros. Fortunately, a number of tests are available that enable us to determine whether or not the roots of the numerator and denominator polynomials of the z-transform lie within the unit circle. For example, Jury's criterion[29], a counterpart of the Routh-Hurwitz criterion[4] for stability of continuous-time systems, provides information on the whereabouts of the roots of a polynomial with respect to the unit circle in the z-plane. Such tests enable one to make the denominator polynomial of a system function minimum phase

(i.e., stable) by appropriately applying feedback to the given system.

For our purposes, however, the discrete-time version of the Nyquist criterion for stability will be more useful since it provides a graphical interpretation of a necessary and sufficient condition for a minimum phase property. The Nyquist criterion will also be useful in deriving a sufficient condition for a minimum phase z-transform denoted as the "positive real" constraint. The positive real constraint is then used to derive other sufficient conditions, and also transformations to ensure the minimum phase property.

3.3.1.1 The Nyquist Criterion

The Nyquist criterion is based on a mapping theorem by Cauchy[29]. If a complex variable z in the z -plane describes a closed contour C_1 in a positive sense, then $H(z)$ will describe a closed contour C_2 in the $H(z)$ plane. The contour C_2 , the polar plot, will encircle the origin M times in the positive direction, where M is the difference between the number of zeros and poles of a rational z-transform $H(z)$ enclosed by C_1 . C_1 is taken to be the closed contour depicted in Fig. 3.4a where the inner radius is the unit circle and the outer radius R is made to approach infinity. The only contribution to the contour C_2 , i.e. the polar plot, results from the component of C_1 along the unit circle[29]. The criterion requires that $\lim_{z \rightarrow \infty} H(z) =$ constant which holds in our case since $h(n)$ is assumed to be causal. When $h(n)$ is also stable all poles of $H(z)$ lie within the unit circle and thus the polar plot will encircle the origin M times, where M is the number of zeros outside the unit circle.

A simple example will be used to illustrate the Nyquist criterion. Consider a three point sequence with z-transform given by

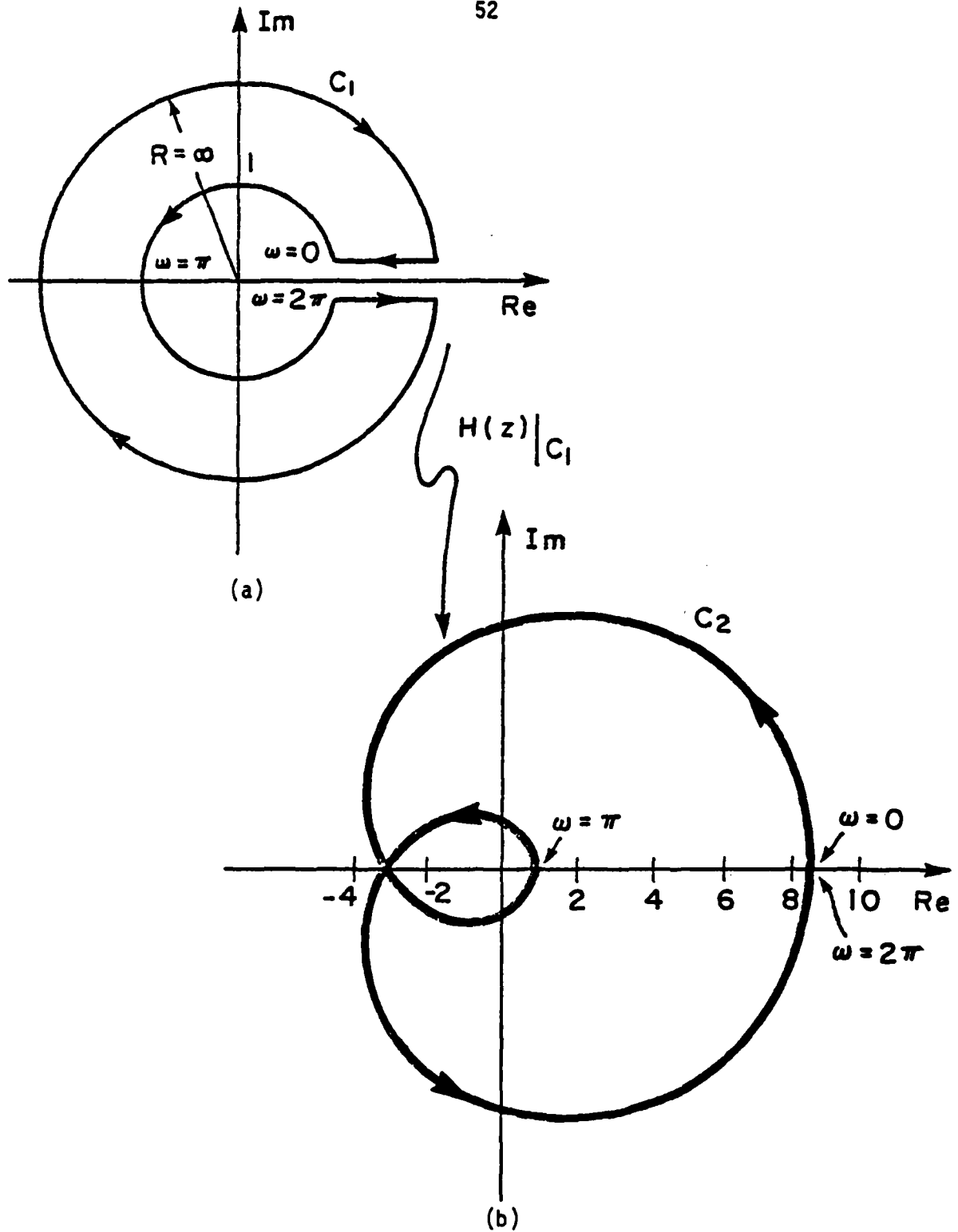


Fig. 3.4 (a) Contour within the z -plane required by the Nyquist criterion, (b) Polar plot of $H(z)$ in (3.24).

$$H(z) = h(0) + 3.9z^{-1} + 4.0z^{-2} \quad (3.24)$$

where $h(0)$ is variable. For $h(0)=1$, the z -transform consists of two zeros outside the unit circle and two poles at the origin. The polar plot of $H(z)$ for $z=\exp[j\omega]$ with $0 \leq \omega \leq 2\pi$ is given in Fig. 3.4b. There exist two clockwise encirclements of the origin, indicating the presence of two zeros outside the unit circle.

One method of modifying this sequence so that it is transformed to a minimum phase sequence is to increase $h(0)$ so that there occur no encirclements of the origin. This is equivalent to adding a positive constant to $H(z)$ or the real component of $H(\omega)$, $H_r(\omega)$; thus the entire polar plot is shifted to the right in the z -plane. From Fig. 3.4b we observe that the leftmost real axis crossing (i.e. $H_i(\omega)=0$) occurs when $H_r(\omega) \approx -3.0$. In this particular example, then, augmenting $h(0)$ so that $h(0) \geq 4$ will ensure no encirclements of the origin. In general, however, constraining the leftmost real axis crossing of the polar plot to fall to the right of the origin is only sufficient for the minimum phase condition[4].

Suppose now that $h(0)$ is augmented further so that $H_r(\omega) > 0 \forall \omega$. Clearly from Fig. 3.4b this constraint represents a sufficient condition on $H_r(\omega)$ to guarantee no encirclements of the origin; that is $H_r(\omega)$ must take on a negative value for some ω for an encirclement of the origin to occur. In our example, $h(0) \geq 5$ guarantees $H_r(\omega) > 0$.

In summary, we can enumerate three increasingly restrictive constraints on $H(\omega)$ for $H(z)$ to be a minimum phase z -transform when $h(n)$ is assumed causal and stable. We shall show that these constraints imply increasingly restrictive limitations as well on the unwrapped phase function of $H(\omega)$. The first is the Nyquist criterion which is a necessary and sufficient con-

dition on the polar plot of $H(\omega)$. The polar plot can be interpreted as a tracing of the path of $H(\omega)$ as a function of magnitude and unwrapped phase. Clearly, the unwrapped phase function which corresponds to this condition generally is unbounded. For example, the polar plot may spiral outward and then by necessity spiral inward without a net encirclement of the origin. The return through an inward spiral must occur since the final value of the unwrapped phase equals its initial value over the interval $[0, \pi]$ when the zeros of $H(z)$ lie within the unit circle.

The second condition is more restrictive, follows directly from the Nyquist criterion, and is only a sufficient condition. This condition requires that the leftmost zero crossing of the polar plot must fall to the right of the origin. As implied by Fig. 3.4b this second condition guarantees that the unwrapped phase of $H(\omega)$ (assuming $H(0) > 0$ and thus $\theta_h(0)=0$) cannot exceed π in absolute value: $|\theta_h(\omega)| < \pi$.

The third condition is somewhat more restrictive than the second, again follows directly from the Nyquist criterion, and requires that the real component of $H(\omega)$ be positive. As implied by Fig. 3.4b this positivity constraint guarantees that the unwrapped phase cannot exceed $\pi/2$: $|\theta_h(\omega)| < \pi/2$. We shall see in the next section that the implications of this positivity constraint are even greater since $H(z)$ under this condition falls into a class denoted as "positive real z-transforms".

The phase restrictions due to the latter two conditions guarantee that the unwrapped phase equals the principal value of the phase. Consequently, the unwrapped phase may be computed at samples by simply an arctangent routine.

3.3.1.2 The Positive Real Constraint

A rational function of a complex variable s , $F(s)$ which is real for real values of s , and whose real part is positive for all values of s with a positive real part, is called a positive real function. Functions of this sort play an important part in electrical network theory and have been studied extensively[9]. In this section we first briefly review the properties of a positive real function. The minimum phase characteristic of such a function is proven in a new way through the continuous-time version of the Nyquist criterion. We then proceed to develop an analogous theory for complex functions with respect to the unit circle. Specifically, a positive real rational z -transform for a causal and stable sequence is defined by the following two properties:

- (i) $H(z)$ is real for real values of z
- (ii) $H_r(z) > 0$ for $|z| \geq 1$

The first property implies simply that the sequence is real. We saw in the previous section that when $h(n)$ is causal and stable and $H_r(\omega) > 0$, $h(n)$ is a minimum phase sequence. We shall show that these conditions are necessary and sufficient for $H(z)$ to be a positive real z -transform. Thus investigation of the positivity of $H_r(z)$ over the entire region $|z| \geq 1$ is not necessary in determining whether $H(z)$ is positive real.

Consider now the positive real function $F(s)$ whose properties were described above. Guillemin has shown that a positive real function cannot have poles in the right half s -plane (RHP) since the presence of a pole implies $F_r(s) < 0$ for some s in the region of the pole[9]. Clearly $F(s)$ cannot have zeros in the RHP and therefore $F(s)$ represents a minimum phase function. Through appropriate use of the theory of functions of a complex

variable, it is possible to show that a necessary and sufficient condition for $F(s)$ to be positive real is that $F(s)$ is positive on the $j\omega$ axis and analytic for $\text{Re}(s) > 0$ [9]. Therefore, an alternative way of proving the minimum phase characteristic of positive real functions is to apply the continuous-time version of the Nyquist criterion to the polar plot of $F(s)$ along the $j\omega$ axis. The analyticity constraint for $\text{Re}(s) > 0$ guarantees no poles in the RHP. The positivity constraint guarantees that there can exist no encirclements of the origin by the polar plot, C_2 ,

where C_1 is taken to be the boundary of the RHP, including the $j\omega$ axis.

Therefore, the RHP is also zero free and $F(s)$ is a minimum phase function.

The discrete-time counterpart to this theory easily follows:

A z-transform $H(z)$ is positive real with respect to the unit circle if and only if $H(z)$ is analytic for $|z| \geq 1$ and $H_r(z) > 0$ on the unit circle.

The proof of the "only if" component of this proposition is identical to that by Guillemin for a positive real function[9], and will not be presented because of the little insight gained. The proof of the "if" component, on the other hand, provides a rather interesting view into the nature of a positive real z-transform. The proof is also similar to that by Guillemin, but is modified to address the discrete-time nature of the problem.

The proof calls upon the Extremum Theorem[9] of complex analysis which concerns the real part of a function, analytic over a given region,

and on the boundary of that region. The theorem states that the largest and smallest values which the real part (or imaginary part) assumes throughout the given region including its boundary must lie on the boundary.

Therefore the smallest values which the real part assumes throughout this region including its boundary must lie on the boundary. In our particular problem the smallest value assumed by the real part of $H(z)$ in the infinitely large annulus whose boundary includes the unit circle and a circle at infinity must occur then on this boundary. Since $H_r(z) > 0$ on the unit circle, the value of $H(z)$ at infinity is also positive:

$$\lim_{z \rightarrow \infty} H(z) = h(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_r(\omega) d\omega > 0 \quad (3.25)$$

Thus our entire boundary is positive and from the Extremum Theorem no point within this boundary can be zero or negative.

Clearly, we can use the positivity of $H_r(\omega)$ and analyticity of $H(z)$ for $|z| \geq 1$ to prove that a positive real z -transform is minimum phase. We have in fact already shown this property in the preceding section through the Nyquist criterion.

3.3.1.3 A Sufficient Condition on $h(0)$

The positive real constraint is useful in testing for a minimum phase characteristic. It is also useful in deriving other sufficient conditions for ensuring this property. In particular, we shall capitalize on the positive real constraint in proving the following sufficient condition:

A causal sequence $h(n)$ is minimum phase if the absolute value of $h(0)$ is greater than the sum of the absolute value of the remaining terms:

$$|h(0)| > \sum_{n \geq 1} |h(n)| \quad (3.26)$$

We shall show by example that this condition is more restrictive than the positive real constraint.

Since $h(n)$ is causal and stable, $H(z)$ is analytic for $|z| \geq 1$. Without loss of generality we also assume that $h(0)$ is positive. Suppose now that for some ω , the positive real condition doesn't hold:

$$H_r(\omega) = h(0) + \sum_{n \geq 1} h(n) \cos \omega n < 0 \quad (3.27)$$

Therefore,

$$\sum_{n \geq 1} h(n) \cos \omega n < -h(0) \quad (3.28)$$

Since $h(0) > 0$, we have

$$\left| \sum_{n \geq 1} h(n) \cos \omega n \right| > |h(0)| \quad (3.29)$$

and so, since $|\cos \omega n| \leq 1$

$$\begin{aligned} \sum_{n \geq 1} |h(n)| &\geq \sum_{n \geq 1} |h(n) \cos \omega n| \\ &\geq \left| \sum_{n \geq 1} h(n) \cos \omega n \right| > |h(0)| \end{aligned} \quad (3.30)$$

In conclusion,

$$|h(0)| < \sum_{n \geq 1} |h(n)| \quad (3.31)$$

which is a contradiction to our original assumption. Therefore, $H_r(\omega) > 0$, $\forall \omega$ and since $H(z)$ is analytic for $|z| \geq 1$, $H(z)$ is positive real, and consequently minimum phase.

The positive real constraint, however, in general does not imply con-

dition (3.26). A counterexample is the sequence $h(n) = \alpha^n u(n)$ where $.5 < \alpha < 1$ and where $u(n)$ is the unit-step sequence. Summing, we obtain

$$\begin{aligned} \sum_{n \geq 1} |h(n)| &= \sum_{n=1}^{\infty} \alpha^n \\ &= \alpha / (1 - \alpha) \\ &> |h(0)| \end{aligned} \tag{3.32}$$

for $.5 < \alpha < 1$. But, $H_r(\omega) > 0 \forall \omega$

3.3.2 Transformations for a Minimum Phase Sequence

In this section we present a number of techniques for transforming an arbitrary, causal mixed phase sequence to a minimum phase sequence. These transformations are based on the conditions discussed in the previous section, and are restricted to a class which has two specific properties:

- (i) A transformation T must be invertible
- (ii) In the context of deconvolution of a quasi-periodic waveform, T must modify the desired sequence, while preserving the convolutional characteristic of the original waveform. That is, a transformation T must be such that

$$\begin{aligned} \tilde{x}(n) &= T[x(n)] \\ &= T[h(n) * p(n)] \\ &= T[h(n)] * \hat{p}(n) \end{aligned} \tag{3.33}$$

where T^{-1} exists, i.e., $T^{-1}[T[h(n)]] = h(n)$, and $p(n)$ may or may not equal $\hat{p}(n)$, but consists of a train of equally spaced samples with spacing P .

We shall consider two such transformations:

- (i) Exponential weighting
- (ii) Addition of a "reference" signal

3.3.2.1 Exponential Weighting

Consider weighting a mixed phase sequence with a decaying exponential [32]:

$$\tilde{h}(n) = \alpha^n h(n) \quad (3.34)$$

This operation results in preservation of the convolutional property and in individual weighting of both components of a quasi-periodic waveform:

$$\begin{aligned} \tilde{x}(n) &= \alpha^n (h(n) * p(n)) \\ &= \alpha^n h(n) * \alpha^n p(n) \end{aligned} \quad (3.35)$$

The z-transform of $\tilde{h}(n) = \alpha^n h(n)$ is given by

$$\tilde{H}(z) = \sum_n \alpha^n h(n) z^{-n} = H(\alpha^{-1} z) \quad (3.36)$$

Therefore if $H(z)$ has a pole or zero at $z=z_0$, $\tilde{H}(z)$ has a pole or zero at αz_0 . $h(n)$ can then be made minimum phase by multiplying $x(n)$ by α^n where α is small enough to move the pole or zero of $h(n)$ with the greatest magnitude inside the unit circle: $h_{mp}(n) = \alpha^n h(n)$. $h(n)$ can always be recovered from $h_{mp}(n)$ by inverse transformation: $h(n) = \alpha^{-n} h_{mp}(n)$.

One drawback to this technique is that often the required α is small enough to cause trouble with rounding error when implemented with a digital computer[32]. A second drawback is that α depends on knowledge of zero locations.

3.3.2.2 Addition of a Reference Signal

A different approach is to add a causal "reference" signal $g(n)$ to $h(n)$ to create a minimum phase sum $\tilde{h}(n) = g(n) + h(n)$. Such transformations are also suitable for preserving the features of a quasi-periodic waveform since with a priori knowledge of $p(n)$, we can form the sequence

$$\begin{aligned}\tilde{x}(n) &= h(n)*p(n) + g(n)*p(n) \\ &= (h(n) + g(n))*p(n)\end{aligned}\tag{3.37}$$

One simple way of guaranteeing that $\tilde{h}(n)$ is minimum phase is to ensure that the polar plot of $\tilde{H}(\omega)$ has no encirclements of the origin by imposing the constraint $\tilde{H}_r(\omega) > 0$. An alternative is to force the leftmost real axis crossing of the polar plot to lie to the right of the origin in the z -plane. The latter condition can be checked by investigation of the gain/phase margin for stability and thus directly from a logmagnitude and unwrapped phase function[4].

A third approach is to add a minimum phase signal. We shall show with the aid of the positive real condition that when $G(z)$ is minimum phase and $|G(\omega)| > |H(\omega)|$, $\forall \omega$, $\tilde{H}(z) = G(z) + H(z)$ is minimum phase. To see this, we express $\tilde{H}(z)$ by

$$\tilde{H}(z) = G(z)(1 + H(z)/G(z))\tag{3.38}$$

$G(z)$ is minimum phase, so the remaining problem is to show that $(1+H(z)/G(z))$ is minimum phase. Since $|H(\omega)/G(\omega)| < 1$, the polar plot of $H(\omega)/G(\omega)$ must lie within the unit circle $|z| < 1$. Therefore, $1+H(\omega)/G(\omega)$ lies to the right of the imaginary axis in the z -plane, so that $\text{Re}(1+H(\omega)/G(\omega)) > 0$. Now, both $H(z)$ and $G(z)$ represent causal sequences and since $G(z)$ is minimum phase, $G^{-1}(z)$ also represents a causal sequence (see section 2.2.1).

Therefore $1+H(z)/G(z)$ corresponds to a causal sequence, and must be analytic for $|z| \geq 1$. Consequently, $1+H(z)/G(z)$ is positive real, and thus minimum phase.

We have already seen an example in section 3.3.1.1 of transforming a mixed phase sequence to a minimum phase sequence by augmenting $h(0)$ with a positive constant: $g(n) = A\delta(n)$. The conditions we have been discussing together with that of section 3.3.1.3 represent restrictions on the values of A of different degrees.

As a point of interest, we can show that modifying the value of $h(0)$ to form a minimum phase sequence is analogous to changing the gain of a feedback control system for stability. Factoring out A in $\tilde{h}(n)=A\delta(n)+h(n)$, we obtain

$$\tilde{H}(z) = A(1 + A^{-1}H(z)) \quad (3.39)$$

$A^{-1}H(z)$ is reminiscent of an "open loop transfer function" where A^{-1} is the "open loop gain"[4].

CHAPTER 4

MIXED PHASE DECONVOLUTION

The problem of deconvolving a sequence from a quasi-periodic waveform was described in section 2.1. In the frequency domain the problem of deconvolution is twofold: to estimate the magnitude and phase of a Fourier transform. In Chapter 3 we assumed that the desired magnitude is known or measurable and that, with a priori information about the sequence or phase function, we found that unambiguous phase recovery is possible. We also investigated methods of creating a situation where the phase can be recovered from the magnitude. That is, we derived transformations which convert a mixed phase sequence to a minimum phase sequence, and which are applicable to magnitude-only deconvolution.

One technique which is potentially capable of directly and accurately estimating both the magnitude and phase of a Fourier transform from a quasi-periodic waveform is homomorphic deconvolution[24,32]. A drawback to this direct approach, however, is the requirement of an unwrapped phase function. The unwrapped phase is generally difficult to compute at samples due to modulo 2π considerations. Available unwrapping algorithms are prone to error when frequency sampling is not "sufficiently" dense[32,34] or the Fourier transform contains regions of low energy which are particularly susceptible to degradation by quantization noise. Moreover, as we shall see, the envelope (i.e., slowly varying component) of the unwrapped phase is quite sensitive to small changes of a sequence in the time-domain. This sensitivity leads to an inherently ill-conditioned problem.

As a result, we wish to either avoid the use of an unwrapped phase function, or compute this function from only frequency regions with a high

signal-to-noise ratio. The latter approach was first taken by Tribolet[33] in confronting the problem of low-pass and high-pass filtered seismic data. His approach entailed shifting and stretching the signal's passband to occupy the entire frequency interval $[0, \pi]$, while preserving the convolutional characteristic of the original sequence. Our problem, however, is not suitable to this approach since the periodic-like nature of our waveforms corresponds in the frequency domain, to multiple bands of low signal-to-noise. In Chapter 5 we develop a technique within this same philosophy, which addresses the harmonic structure of our spectra.

The alternative approach is to avoid the issue of phase unwrapping by conversion to a magnitude-only deconvolution problem. Applying the transformations of section 3.3.2.2, we can perform homomorphic deconvolution without an unwrapped phase. Note that although we shall illustrate magnitude-only deconvolution through homomorphic filtering, any deconvolution technique yielding a minimum or zero phase solution is applicable.

When such an approach is difficult to take, as for example with inadequate knowledge of $p(n)$, we must compute DFT samples of an unwrapped phase in directly estimating phase by homomorphic deconvolution. The need of an unwrapped phase arises also in the following chapter where a phase estimate is deduced from only harmonic samples of a Fourier transform.

The second major area of this chapter presents a new technique for determining the unwrapped phase which does not require conventional modulo 2π considerations. In developing this technique, we first investigate magnitude retrieval from the phase of a Fourier transform. In a manner similar to our development in sections 3.1 and 3.2 we formulate a linear iterative procedure which retrieves the magnitude from the phase, and which

constrains the values of a sequence over a specified region.

Under a causality constraint, the algorithm provides an alternative to the Hilbert transform in obtaining the magnitude from the phase for a minimum phase sequence. However, it has the advantage of not requiring an unwrapped phase function, but only the principal value of the phase. Under a finite length constraint, the algorithm is capable of unambiguously recovering a magnitude from an arbitrary phase.

Finally, in the context of phase unwrapping, the algorithm serves as a major component within our new phase unwrapping algorithm.

4.1 Phase Estimation by Homomorphic Deconvolution

When we confront the problem of filtering signals which have been added, we often use a linear filter. Extracting the phase or the magnitude of $H(\omega)$, on the other hand, is a nonlinear problem since $H(\omega)$ and $P(\omega)$ are multiplicatively combined. The approach of homomorphic system analysis proposed by Oppenheim[24] transforms this nonlinear problem to a linear filtering problem. In this section we first review the theory of multiplicative homomorphic systems, and "direct" phase estimation by homomorphic filtering. We then proceed to illustrate the sensitivity inherent in this direct approach due to the requirement of an unwrapped phase function. Finally, a specific transformation of section 3.3.2.2 is applied in demonstrating the indirect approach of phase estimation by magnitude-only homomorphic deconvolution.

4.1.1 Theory

Much of the groundwork for multiplicative homomorphic systems has been laid in Chapter 2. The purpose of such systems is to transform the problem of separating the multiplicative components of $\chi(\omega) = H(\omega)P(\omega)$ to a problem of linearly filtering additive components. In particular, the complex log-

arithm provides the means of obtaining this additivity:

$$\begin{aligned}\log[X(z)] &= \log[H(z)] + \log[P(z)] \\ &= \log|H(z)| + \log|P(z)| + j(\theta_h(z) + \theta_p(z))\end{aligned}\quad (4.1)$$

The additivity of the imaginary components (i.e., the phase components) of the complex logarithm of $X(z)$ holds when the unwrapped phase is used in (4.1). A linear phase component z^{n_0} may also be included when its unwrapped phase is defined to be odd, but discontinuous at π . This term, nevertheless, should be removed since it interferes with the estimation of $\theta_h(\omega)$ [32].

Under these conditions, the complex logarithmic operation represents a homomorphic system which maps multiplication to addition. Taking the inverse transform of (4.1) we obtain the real sequence

$$\hat{x}(n) = \hat{h}(n) + \hat{p}(n) \quad (4.2)$$

$\hat{x}(n)$ is termed the "complex cepstrum" to emphasize the use of the complex logarithm. We use the term "real cepstrum" when the real logarithm is used and thus only the magnitude of the Fourier transform of $x(n)$ is retained. Likewise, "phase cepstrum" refers to the case where only the phase is retained.

For a rational $H(z)$, (4.2) can be expressed in terms of the complex cepstrum of the minimum and maximum phase components of the desired sequence and the pulse train:

$$\hat{x}(n) = (\log[A])\delta(n) + \hat{h}_{\min}(n) + \hat{h}_{\max}(n) + \hat{p}(n) \quad (4.3)$$

where $\hat{h}_{\min}(n)$ and $\hat{h}_{\max}(n)$ are normalized so that $\hat{h}_{\min}(0) = \hat{h}_{\max}(0) = 1$.

Using the well-known power series expansions:

$$\log[1 - \alpha z^{-1}] = - \sum_{n=1}^{\infty} \frac{\alpha^n}{n} z^{-n} \quad |z| > |\alpha| \quad (4.4a)$$

and

$$\log[1 - \alpha z^{+1}] = - \sum_{n=1}^{\infty} \frac{\alpha^n}{n} z^{-n} \quad |z| < |\alpha| \quad (4.4b)$$

From (2.19) and (2.20), it follows that

$$\hat{h}_{\min}(n) = \left[- \sum_{k=1}^{m_1} \frac{a_k^n}{n} + \sum_{k=1}^{p_1} \frac{c_k^n}{n} \right] u(n-1) \quad (4.5a)$$

and

$$\hat{h}_{\max}(n) = \left[+ \sum_{k=1}^{m_0} \frac{b_k^n}{n} - \sum_{k=1}^{p_0} \frac{d_k^n}{n} \right] u(-n-1) \quad (4.5b)$$

where $u(n)$ is the unit-step function and $|a_k|$, $|b_k|$, $|c_k|$, and $|d_k|$ are less than unity.

The properties of the complex cepstrum that are of importance in the sequel can be summarized as follows:

P1 $\hat{h}(n)$ decays at least as fast as $1/n$. Specifically,

$$|\hat{h}(n)| < C \left| \frac{\beta^n}{n} \right| \quad (4.6)$$

where C is a constant and β equals the maximum of $|a_k|$, $|b_k|$, $|c_k|$, and $|d_k|$.

P2 The complex cepstrum of a minimum phase sequence is stable and causal, and likewise the complex cepstrum of a maximum phase sequence is stable and anti-causal.

P3 The complex cepstrum $\hat{p}(n)$ of a train of equally spaced

pulses $p(n)$, is also a train of equally spaced pulses with the same spacing.

4.1.2 The Direct Approach

Suppose that $\hat{h}(n)$ decays rather quickly and that the sample spacing of $\hat{p}(n)$ is sufficiently large so that $\hat{h}(n)$ and $\hat{p}(n)$ do not considerably overlap. Under these conditions we can filter $\hat{h}(n)$ from $\hat{p}(n)$ by applying a "low-time gate" to $\hat{x}(n)$. An estimate of $h(n)$ is computed by transforming the estimate of $\hat{h}(n)$, exponentiating, and inverse transforming.

Estimation of $\hat{h}(n)$ is equivalent to obtaining an estimate of the even and odd components of $\hat{h}(n)$. These components correspond to $\log|H(\omega)|$ and $\theta_h(\omega)$ respectively. Therefore, we might estimate the two components independently with low-time gates tailored to each separately.

4.1.3 The Heuristics of Unwrapped Phase Sensitivity

Suppose that the sequence $x(n) = h(n)*p(n)$ is modified by an additive disturbance $d(n)$:

$$x(n) = h(n)*p(n) + d(n) \quad (4.7)$$

In the frequency domain, (4.7) is given by

$$\begin{aligned} X(\omega) &= H(\omega)P(\omega) + D(\omega) \\ &= H(\omega)P(\omega)(1 + D(\omega)/H(\omega)P(\omega)) \\ &= H(\omega)P(\omega)E(\omega) \end{aligned} \quad (4.8)$$

where,

$$E(\omega) = (1 + D(\omega)/H(\omega)P(\omega)) \quad (4.9)$$

Therefore,

$$\log[X(\omega)] = \log[H(\omega)P(\omega)] + \log[E(\omega)] \quad (4.10)$$

Thus, the effect of $D(\omega)$ on the complex cepstrum is strongly dependent on the ratio $D(\omega)/H(\omega)P(\omega)$.

For homomorphic deconvolution to be robust in the presence of a disturbance $d(n)$, the slowly varying components (i.e., the envelope) of the complex logarithm should not be susceptible to a large change with a small change in $x(n)$ due to $d(n)$. We wish to preserve the slowly varying components of $\log|X(\omega)|$ and $\theta_x(\omega)$ since they map to the low-time region of the complex cepstrum. In the context of quasi-periodic waveforms the following observations were made:

- (i) "Small" disturbances in the sequence $x(n)$ tend to propagate small changes in the envelope of $\log|X(\omega)|$ and thus small changes in the low-time region of the real cepstrum.
- (ii) Small disturbances in $x(n)$ often propagate "large" changes in the envelope of the unwrapped phase of $\theta_x(\omega)$ and thus large changes in the low-time region of the phase cepstrum.
- (iii) Large changes in the unwrapped phase envelope of $\theta_x(\omega)$ may or may not be mapped through low-time cepstral gating to large changes in the impulse response estimate[28,35].

In summary, an estimate of $h(n)$ obtained by homomorphic deconvolution which requires a magnitude function only is less susceptible to degradation by time-domain disturbances than an estimate based on magnitude and phase.

The sensitivity of the unwrapped phase function is understood by the following heuristic argument. Consider a region where $|D(\omega)/H(\omega)P(\omega)| < 1$. Under this constraint the unwrapped phase of $E(\omega)$ cannot exceed $\pi/2$ in

absolute value, as depicted in Fig. 4.1a, where the "noise vector", $D(\omega)/H(\omega)P(\omega)$ remains smaller in magnitude than the "signal vector", $1+j0$.

When the noise vector exceeds the signal vector in magnitude, as for example in a low-energy region of $H(\omega)P(\omega)$, the polar plot of $E(\omega)$ may encircle the origin resulting in a 2π jump in the unwrapped phase of $E(\omega)$.

Such 2π jumps can accumulate and significantly distort the envelope of the unwrapped phase of $\theta_x(\omega)$, and thus the low-time region of the phase cepstrum. Gating the phase cepstrum may therefore not extract an accurate estimate of the desired phase, $\theta_h(\omega)$.

4.1.4 Magnitude-Only Deconvolution by Transformation: An Indirect Approach

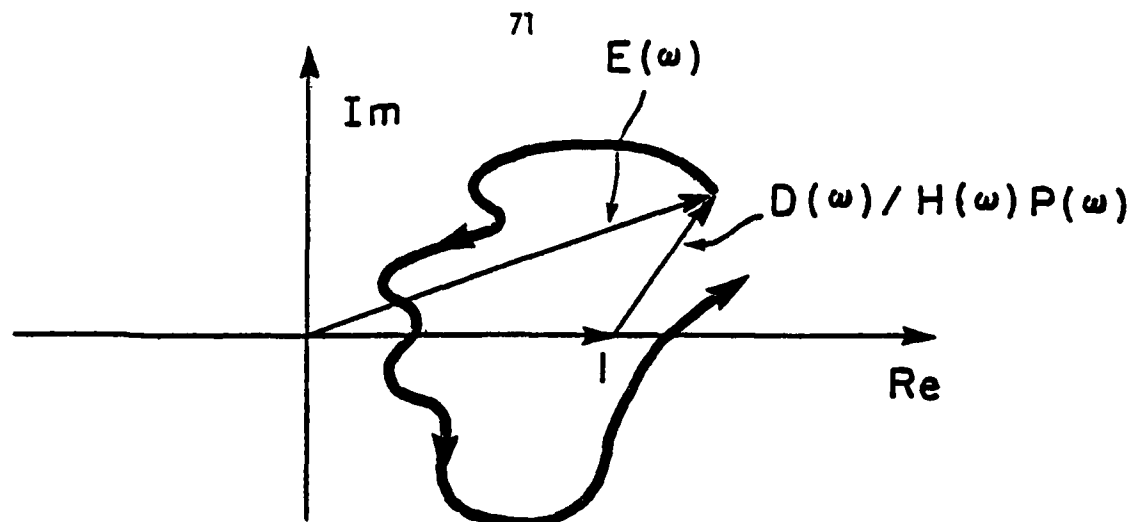
Consider modifying a causal sequence $h(n)$ by adding to it a sequence $g(n)$ so that the resulting sum $v(n) = h(n) + g(n)$ is minimum phase. In particular, we choose $g(n) = A\delta(n)$. From (3.36) to preserve the convolutional property of $x(n)$, and to appropriately modify $h(n)$, we add to $x(n)$ the sequence $A\delta(n)*p(n)$ to obtain:

$$\begin{aligned} y(n) &= (h(n) + A\delta(n)) * p(n) \\ &= v(n) * p(n) \end{aligned} \quad (4.11)$$

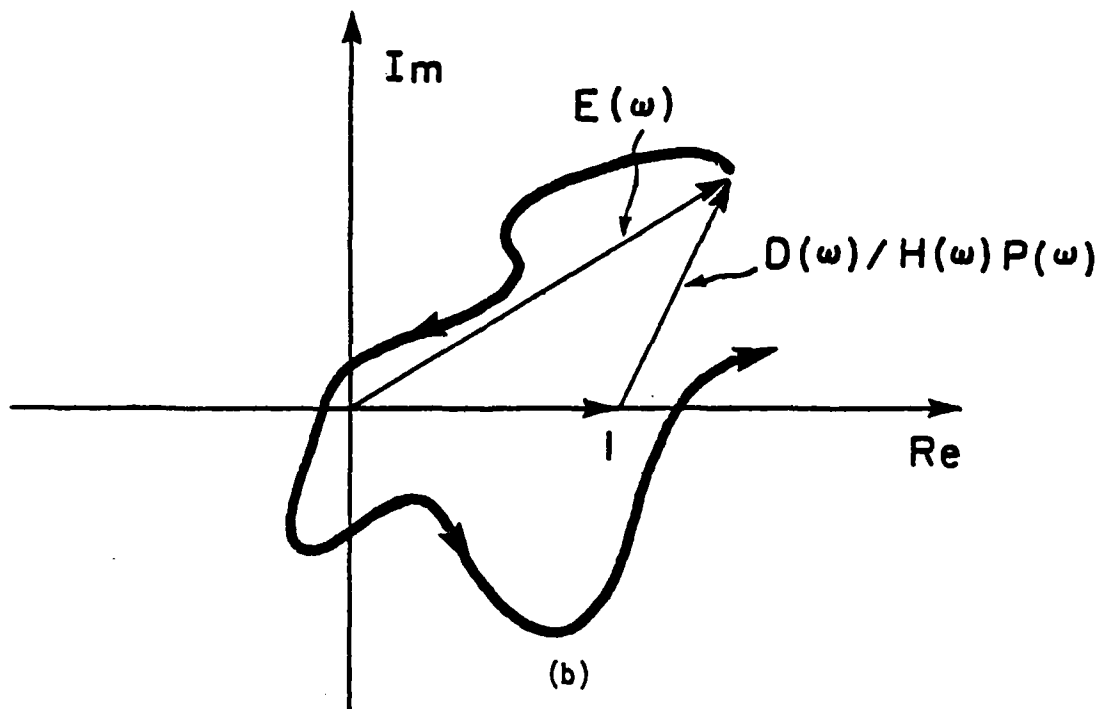
Since $A\delta(n)$ is such that $v(n)$ is minimum phase, from property P2 of section 4.1.1, the complex cepstrum of $v(n)$, $\hat{v}(n)$ is causal and thus completely characterized by its even component or equivalently by $\log|V(\omega)|$. From (2.11a) the complex cepstrum of $v(n)$ is therefore expressed by

$$\hat{v}(n) = L(n)\hat{v}_r(n) \quad (4.12)$$

where $\hat{v}_r(n)$ is the real cepstrum of $v(n)$ and $L(n)$ is the causal sequence given by



(a)



(b)

Fig. 4.1 (a) Polar plot of $E(\omega)$ in (4.9) with a small disturbance (no origin encirclement)
 (b) Same as (a) with a large disturbance (origin encirclement)

$$L(n) = \begin{cases} 1 & n=0 \\ 2 & n>0 \\ 0 & n<0 \end{cases} \quad (4.13)$$

Since $H(z)$ is assumed rational, $V(z)$ is also rational and therefore $\hat{v}(n)$ satisfies properly P1 of section 4.1.1.

The real cepstrum of $p(n)$ is the even component of the complex cepstrum of $p(n)$, $\hat{p}(n)$ and from property P3 of section 4.1.1 consists of equally spaced samples with spacing P . Consequently, when $\hat{v}(n)$ decays rather quickly, the real cepstra $\hat{v}_r(n)$ and $\hat{p}_r(n)$ do not significantly overlap. From this property and (4.12) we can obtain an estimate of $\hat{v}(n)$ by gating the real cepstrum of $y(n)$:

$$\begin{aligned} \hat{v}(n) &= g(n) \hat{y}_r(n) \\ &= g(n) \hat{v}_r(n) + g(n) \hat{p}_r(n) \\ &= \begin{cases} \hat{v}(n) & 0 < n < P \\ \hat{v}(0) + \hat{p}(0) & n=0 \\ 0 & n<0, \quad n \geq P \end{cases} \end{aligned} \quad (4.14)$$

where

$$g(n) = \begin{cases} 1 & n=0 \\ 2 & 0 < n < P \\ 0 & n<0, \quad n>P \end{cases} \quad (4.15)$$

With $\hat{p}(0)=0$, we then obtain an estimate of $v(n)$ by transforming $g(n)\hat{y}_r(n)$, exponentiating and inverse transforming. Finally, an estimate of $h(n)$ is obtained through the inverse operation of subtracting $A\delta(n)$ from the estimate of $v(n)$.

With respect to the choice of the value A , a number of comments are in order. We saw in section 3.3.2.2 that the value of A may be chosen to satisfy any one of a number of sufficient conditions for ensuring a minimum phase characteristic. The extreme case of choosing A so that the Nyquist criterion is just met should be avoided. This is because such a choice places a zero of $V(z)$ "just on" the unit circle. From property P1 of section 4.1.1, the complex cepstrum of $v(n)$ decays as approximately $C|n|^{-1}$, and so the requirement that $\hat{v}(n)$ decay rather quickly is not satisfied.

The alternative extreme is to allow the value of A to become very large. With this choice, the complex cepstrum of $y(n)$ has an interesting and rather useful property. Replacing αz^{-1} by $H(z)/A$ in the logarithmic expansion of (4.4a), and assuming $|H(z)/A| \ll 1$, we obtain

$$\begin{aligned}\log[Y(z)] &= \log[A + H(z)] + \log[P(z)] \\ &= \log[A] + \log[1 + H(z)/A] + \log[P(z)] \\ &\approx \log[A] + H(z)/A + \log[P(z)]\end{aligned}\quad (4.16)$$

$\hat{y}(n)$ is then written approximately as

$$\hat{y}(n) = \begin{cases} \log[A] + h(0)/A + \hat{p}(0) & n=0 \\ h(n)/A + \hat{p}(n) & n>0 \\ \hat{p}(n) & n<0 \end{cases} \quad (4.17)$$

Therefore, $h(n)$ can be recovered "almost exactly" and directly from the complex cepstrum by setting $\hat{y}(n)=0$ for $n=kP$, where $k=\pm 1, \pm 2, \dots$, subtracting $\log[A] + \hat{p}(0)$ from $\hat{y}(0)$, and scaling the result by A . An equivalent operation can also be carried out on the real cepstrum.

In order to eliminate the need of a priori knowledge of $\hat{p}(0)$ in the above algorithms, we apply $g(n) = A\delta(n)$ to a version of $h(n)$ which is shifted to the right:

$$v(n) = A\delta(n) + h(n-n_0) \quad (4.18)$$

For this case, we can show that the estimate obtained from (4.14) for $n \geq n_0$ is given approximately by $\exp[\hat{p}(0)] h(n-n_0)$. Thus, only a scaling degradation occurs so that the desired phase is preserved.

4.1.5 Examples

We now consider two examples of the techniques discussed within this section. We shall compare estimates from the direct and indirect approaches of sections (4.1.2) and (4.1.4) obtained by gating the complex and real cepstrum, respectively, with and without, a white noise disturbance.

Example 4.1

Consider $H(z)$ with two complex pole pairs at 292 Hz and 3500 Hz, and a maximum phase complex zero pair at 2000 Hz. $p(n)$, the pulse train is given by

$$p(n) = .3\delta(n) + \delta(n-64) + .5\delta(n-128) \quad (4.19)$$

where $P=64$. In order to reduce the effect of sidelobe interference of the Fourier transform of a rectangular gate, the complex cepstrum is multiplied by a Hamming window expressed by

$$w(n) = \begin{cases} .54 - .46 \cos \left[\frac{2\pi(n-60)}{120} \right] & -60 \leq n \leq 60 \\ 0 & |n| > 60 \end{cases} \quad (4.20)$$

The original continuous unwrapped phase and its estimate derived from the direct approach are illustrated in Figs. 4.2a and 4.2b, respectively.

Next, the indirect approach is taken where $p(n)$ is appropriately scaled so that $\hat{p}(0)=0$. We set $A=.05$ to create a minimum phase sequence, $v(n)=A\delta(n) + h(n)$. The real cepstrum of $x(n)$ is multiplied by $w'(n)=w(n)L(n)$ where $L(n)$ is given in (4.13) and $w(n)$ is given in (4.20). The resulting continuous phase estimate is depicted in Fig. 4.2c and is almost indistinguishable from that derived from the direct approach.

Example 4.2

This example is identical to example 4.1, but now we add white noise to $x(n)$ with a 40 db S/N. The original unwrapped phase and results from the direct and indirect approach are illustrated in Fig. 4.3. A linear phase component is incorporated so that all three impulse response estimates are causal and nonzero at $n=0$. Note that a jump of approximately 4π occurs near the zero of $H(z)$ at 2000 Hz. This jump indicates that the noise component caused two encirclements of the origin by $E(\omega)$, as depicted in Fig. 4.1b.

4.2 Phase Unwrapping by Phase-Only Signal Reconstruction

When $H(z)$ is rational, we saw in section 2.3.2 that the number of possible magnitude functions is finite when the phase is fixed. To specify a unique magnitude additional constraints must be imposed. One set of constraints of particular interest is causality and phase continuity. These constraints are compatible with only a minimum phase sequence and are easily incorporated within a linear iterative procedure for magnitude retrieval. A consequence of this iterative algorithm and the ultimate objective of

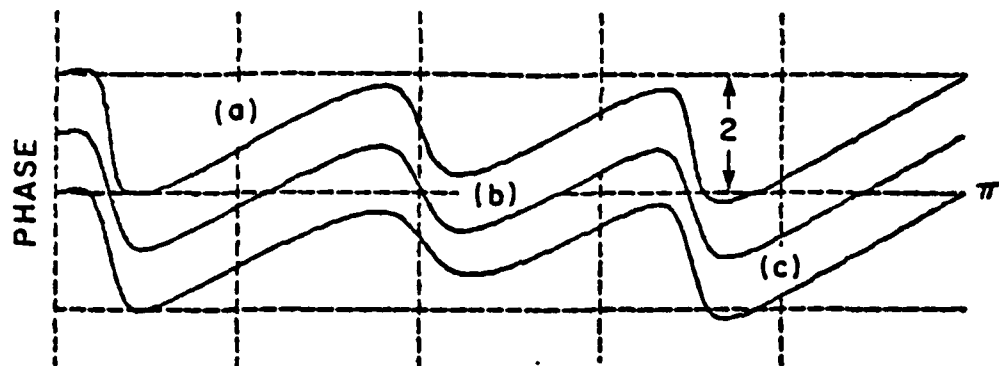


Fig. 4.2 (a) Continuous unwrapped phase of $H(\omega)$ from example 4.1
 (b) Estimate of (a) by the direct approach
 (c) Estimate of (a) by the indirect approach

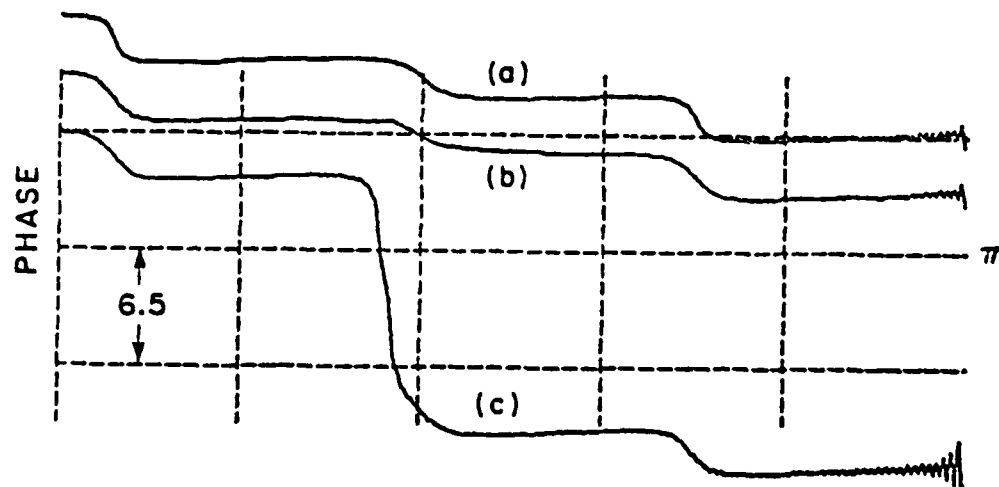


Fig. 4.3 (a) Unwrapped phase (with a linear phase component) of $H(\omega)$ from example 4.2
 (b) Estimate of (a) in the presence of noise by the indirect approach
 (c) Estimate of (a) in the presence of noise by the direct approach

this section is a new means for computing the unwrapped phase without modulo 2π considerations. Another consequence of this iterative procedure, equally important, is a new technique for obtaining the magnitude from the phase of the Fourier transform of a minimum phase sequence. The Hilbert transform, the conventional approach, requires an unwrapped phase. Our procedure, on the other hand, requires only the principal value of the phase, and thus avoids the necessity of phase unwrapping.

Other constraints may also be incorporated within the iterative algorithm. In parallel with this thesis, Hayes et al have demonstrated that under a set of rather loose conditions on $H(z)$ the magnitude function of a finite length sequence is unambiguously determined (within a scale factor) by the phase. As we shall see, this finite length constraint can therefore be imposed within the iterative procedure to retrieve the magnitude. In addition, it is useful in developing a DFT realization of the iteration.

4.2.1 Magnitude Retrieval From Phase With Constraints

In this section, we consider constraints for recovery of the magnitude from the phase of a Fourier transform.

4.2.1.1 Causality and Phase Continuity

Consider a sequence $h(n)$ which is causal and of arbitrary duration. Furthermore, suppose the unwrapped phase of its Fourier transform is continuous and odd, and thus is without a linear phase contribution. We assume that $h(n)$ contains a maximum phase component and then show that this assumption leads to a contradiction.

$h(n)$ can be expressed by

$$h(n) = A h_{\min}(n) * h_{\max}(n) \quad (4.21)$$

where $h_{\min}(n)$ and $h_{\max}(n)$ are normalized so that $h_{\min}(0)=h_{\max}(0)=1$. Consequently, the z-transform of $h(n)$ is given by

$$\begin{aligned} H(z) &= A H_{\min}(z) H_{\max}(z) \\ &= A \left(1 + \sum_{n \geq 1} h_{\min}(n) z^{-n} \right) \\ &\quad \times \left(1 + \sum_{n \leq -1} h_{\max}(n) z^{-n} \right) \end{aligned} \quad (4.22)$$

Since the region of convergence of $H_{\max}(z)$ must be the interior of a circle of finite radius[22], the region of convergence of $H(z)$ cannot include $z=\infty$. However, since $h(n)$ is assumed causal and stable, from (4.22)

$$\lim_{z \rightarrow \infty} H(z) = A \quad (4.23)$$

and so a contradiction arises, and $h(n)$ must be minimum phase.

Therefore, under our constraints of causality and phase continuity, the magnitude of $H(\omega)$ is unambiguously specified (within a scale factor) by the phase.

4.2.1.2 The Finite Length Constraint

Let us suppose that $H(z)$ contains no conjugate reciprocal zeros. Under this condition the magnitude of a Fourier transform is uniquely determined (within a scale factor) by the phase when $h(n)$ is constrained to be of finite duration[10]. Equivalently, suppose $h(n)$ is causal and of finite duration, and $H(z)$ contains no conjugate reciprocal zeros. The phase of $H(\omega)$ is then sufficient to recover a mixed phase $h(n)$, and the corresponding unwrapped phase need not be continuous, i.e. $\theta_h(\omega)$ may include a linear phase component.

This property can be argued heuristically from (2.28). The z-transform

of a finite length sequence contains only zeros. Therefore, creating a magnitude function different from the original and maintaining the phase requires from (2.28) reflecting a zero to a conjugate reciprocal pole. But the presence of a pole implies an infinitely long sequence, and thus only a zero geometry is allowed. This geometry can be shown to be unique as in [10].

Furthermore, under these same conditions when the sequence is M points in duration, M samples of the phase function in the frequency interval $[0, \pi]$ are sufficient to uniquely determine the sequence within a scale factor [10].

4.2.2 An Iterative Procedure to Recover Magnitude From Phase

Solutions to both the above magnitude retrieval problems can be given in closed form. In the former case the Hilbert transform generates the magnitude from the unwrapped phase. In the later case, the magnitude can be obtained indirectly through solution of a set of linear equations [10].

In this section, we discuss an alternative method of solution which invokes a linear iterative algorithm.

4.2.2.1 Theory

The iterative algorithm will now be described and a proof is given showing that a defined error must strictly decrease as the algorithm iterates. We also prove that the sequence of functions $h_k(n)$ derived from the iteration converges to $h(n)$ for $n \in I$ where I is the set of integers for which $h(n)$ is known a priori. The algorithm is illustrated in Fig. 4.4. We begin with an initial guess $M_0(\omega)$ of the desired magnitude and take the inverse Fourier transform of $M_0(\omega) \exp[j\theta(\omega)]$, where $\theta(\omega)$ is the known phase. This step yields $h_0(n)$, the initial estimate of $h(n)$. Next, the known

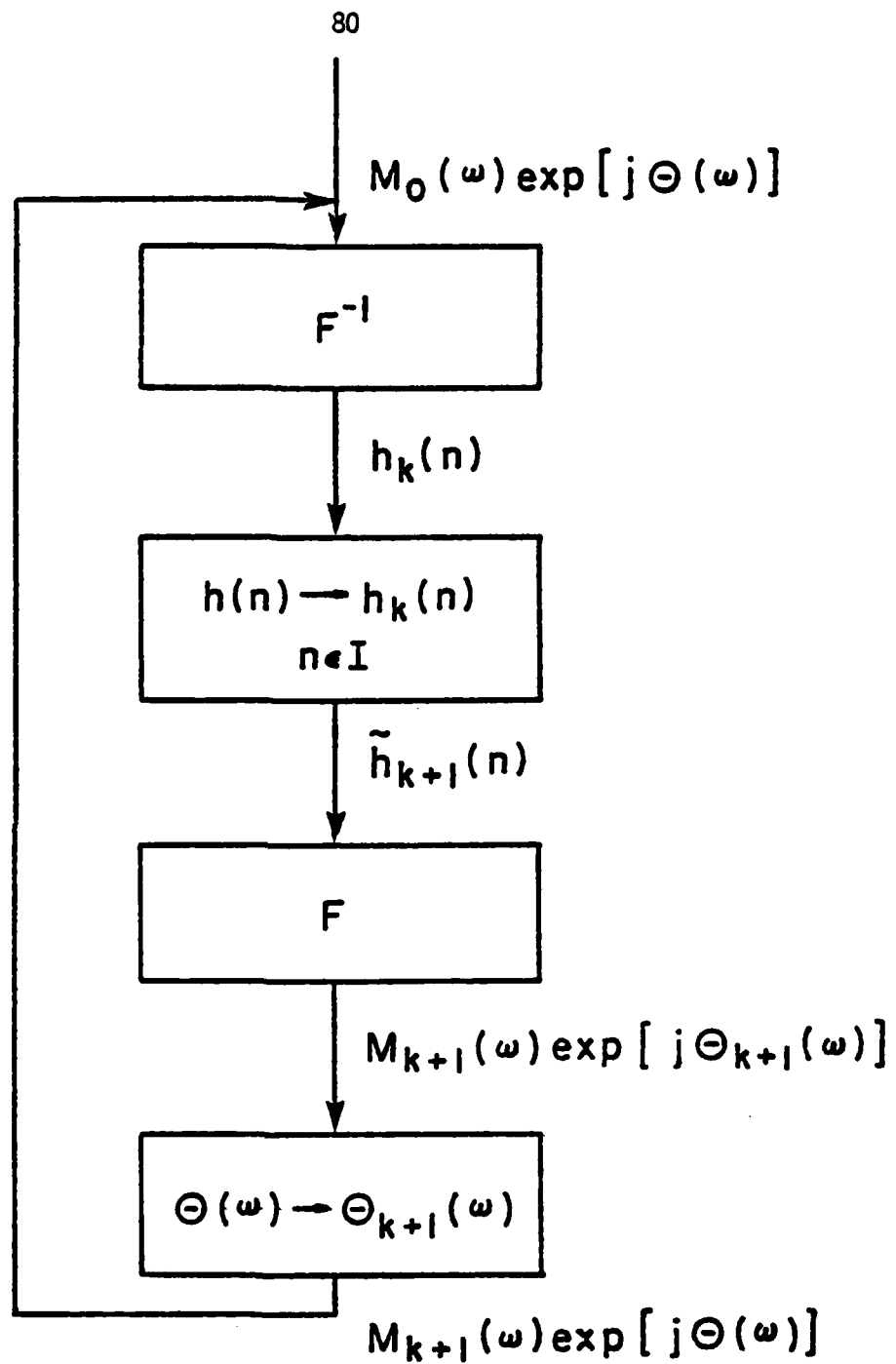


Fig. 4.4 Iterative algorithm to recover magnitude from phase

values of $h(n)$ for $n \in I$ are incorporated in the initial estimate $h_0(n)$ to obtain $\tilde{h}_1(n)$. The phase of the Fourier transform of $\tilde{h}_1(n)$ is then replaced by the given phase and the procedure is repeated. The steps involved in one iteration are summarized below. $h_k(n)$, $\theta_k(\omega)$, and $M_k(\omega)$ are the signal, phase, and magnitude estimates, respectively on the k th iteration, and $\tilde{h}_{k+1}(n)$ is defined in (3.10).

Iteration to Recover Magnitude From Phase

- (i) Inverse transform $M_k(\omega) \exp[j\theta(\omega)]$: $h_k(n)$
 - (ii) Replace $h_k(n)$ with $h(n)$ for $n \in I$: $\tilde{h}_{k+1}(n)$
 - (iii) Forward transform $\tilde{h}_{k+1}(n)$: $M_{k+1}(\omega) \exp[j\theta_{k+1}(\omega)]$
 - (iv) Replace $\theta_{k+1}(\omega)$ with $\theta(\omega)$: $M_{k+1}(\omega) \exp[j\theta(\omega)]$
 - (v) Repeat
-

We shall first show in two steps that the mean squared error between $H(\omega)$ and $H_k(\omega)$ strictly decreases on each iteration.

Error Reduction in the Time Domain

The mean squared error on the k th iteration from Parseval's Theorem can be written as

$$\begin{aligned}
 E_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega) - H_k(\omega)|^2 d\omega \\
 &= \sum_n |h(n) - h_k(n)|^2 \\
 &= \sum_{n \notin I} |h(n) - h_k(n)|^2 + \sum_{n \in I} |h(n) - h_k(n)|^2
 \end{aligned} \tag{4.24}$$

From (3.10), we have

$$|h(n) - h_k(n)| = |h(n) - \tilde{h}_{k+1}(n)|, \quad n \in I \quad (4.25a)$$

and

$$|h(n) - h_k(n)| \geq |h(n) - \tilde{h}_{k+1}(n)| = 0, \quad n \in I \quad (4.25b)$$

Summing (4.25a) and (4.25b) over all n , we obtain

$$\begin{aligned} E_k &= \sum_n |h(n) - h_k(n)|^2 \\ &\geq \sum_n |h(n) - \tilde{h}_{k+1}(n)|^2 \end{aligned} \quad (4.26)$$

Error Reduction in the Frequency Domain

From Parseval's Theorem, we write (4.26) as

$$\begin{aligned} E_k &\geq \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega) - \tilde{H}_{k+1}(\omega)|^2 d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |M(\omega)\exp[j\theta(\omega)] - M_{k+1}(\omega)\exp[j\theta_{k+1}(\omega)]|^2 d\omega \end{aligned} \quad (4.27)$$

With the triangle inequality for vector differences, we have at each frequency ω :

$$\begin{aligned} &|M(\omega)\exp[j\theta(\omega)] - M_{k+1}(\omega)\exp[j\theta_{k+1}(\omega)]| \\ &\geq |M(\omega)\exp[j\theta(\omega)]| - |M_{k+1}(\omega)\exp[j\theta_{k+1}(\omega)]| \\ &= |M(\omega) - M_{k+1}(\omega)| \end{aligned} \quad (4.28)$$

Therefore, from (4.27), (4.28), and the identity $|\exp[j\theta(\omega)]|^2 = 1$:

$$E_k \geq \frac{1}{2\pi} \int_{-\pi}^{\pi} |M(\omega) - M_{k+1}(\omega)|^2 d\omega$$

$$\begin{aligned}
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |M(\omega)\exp[j\theta(\omega)] - M_{k+1}(\omega)\exp[j\theta(\omega)]|^2 d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega) - H_{k+1}(\omega)|^2 d\omega \\
&= E_{k+1} \tag{4.29}
\end{aligned}$$

Therefore, E_k is at least monotone decreasing upon each iteration. Since $E_k \geq 0$, it has a lower bound of zero. Therefore, E_k must converge to a unique limit[30].

Let us now suppose that on the k th iteration $h_k(n) \neq h(n)$ for $n \in I$. From (4.25b) and (4.26) the mean squared error must decrease on the next iteration. Therefore, the error strictly decreases unless convergence to the limit of E_k has been reached.

Proof of Convergence over I

Since E_k is convergent and real, E_k is a Cauchy sequence[30]. Therefore, given any real number $\epsilon > 0$, there is a positive integer N such that $|E_k - E_\ell| < \epsilon$ whenever $k, \ell > N$. In particular, let $\ell = k+1$ so that for any $\epsilon^2 > 0$, we can find an N so that whenever $k > N$

$$|E_k - E_{k+1}| = E_k - E_{k+1} < \epsilon^2 \tag{4.30}$$

Then, from (4.26) through (4.29), and (4.30)

$$\begin{aligned}
&\sum_n |h(n) - h_k(n)|^2 - \sum_n |h(n) - \tilde{h}_{k+1}(n)|^2 \\
&= \sum_{n \in I} |h(n) - h_k(n)|^2 \\
&< E_k - E_{k+1} < \epsilon^2 \tag{4.31}
\end{aligned}$$

Therefore,

$$|h(n) - h_k(n)|^2 < \epsilon^2 \quad n \in I \quad (4.32a)$$

and

$$|h(n) - h_k(n)| < \epsilon \quad n \in I \quad (4.32b)$$

Finally, (4.32b) implies that

$$\lim_{k \rightarrow \infty} h_k(n) = h(n) \quad n \in I \quad (4.33)$$

From (4.33), and since each sequence $h_k(n)$ has the known phase, $\theta_h(\omega)$, a reasonable conjecture, under the causality and finite length constraints of section 4.2.1 where $h(n)=0$ for $n \in I$, is that

$$\lim_{k \rightarrow \infty} M_k(\omega) = \alpha M(\omega) \quad (4.34)$$

where α is a constant. We shall see in section 4.2.2.4 that in practice the limit given by (4.34) is in fact approached.

4.2.2.2 The DFT Realization

We saw in section 4.2.1.2 that M points of the phase in the frequency interval $[0, \pi]$ are sufficient to characterize an M point sequence within a scale factor. Therefore, we are justified in formulating the iterative algorithm and proofs of error reduction and convergence of section 4.2.1 for a DFT realization with DFT length $2M-1$ or greater in a manner similar to that of section 3.2.

4.2.2.3 Examples

The following two examples illustrate the iterative algorithm with the constraints of section 4.2.1. In both cases the initial magnitude estimate

$M_0(\omega)$ is set to unity.

Example 4.3

Let $h(n)$ be the sequence as defined in example 3.1 whose z-transform contains two minimum phase complex pole pairs and one maximum phase zero pair. Its minimum phase counterpart $h_{mp}(n)$ with the continuous phase of $H(\omega)$ is of infinite duration and so within a DFT realization must be truncated. We assume that $h(n)=0$ for $n>256$ and use a 512-point DFT. Truncation of $h_{mp}(n)$ does not noticeably alter the phase of $H(\omega)$. Nevertheless, the phase of the Fourier transform of the truncated $h_{mp}(n)$ is incorporated within the iteration.

The sequence of functions $M_k(\omega)$ and $\theta_k(\omega)$ are depicted in Fig. 4.5 for 5, 15, 25 and 100 iterations superimposed on the original phase, and magnitude functions of $h_{mp}(n)$. $M_k(\omega)$ and $\theta_k(\omega)$ are indistinguishable from the magnitude and phase of the original minimum phase function after 100 iterations. Note that the maximum phase zeros of the original z-transform now appear as conjugate reciprocal poles as we would predict from (2.28). The negative offset in the logmagnitude estimate occurs since α in (4.34) is less than unity.

Example 4.4

Let $h(n)$ be the mixed phase causal 8-point sequence of example 3.2. A 512-point DFT is used and the linear phase component retained. The sequence of functions $M_k(\omega)$ and $\theta_k(\omega)$ are depicted in Fig. 4.6 superimposed on the original magnitude and phase of $H(\omega)$ for 1, 5, 15, and 100 iterations.

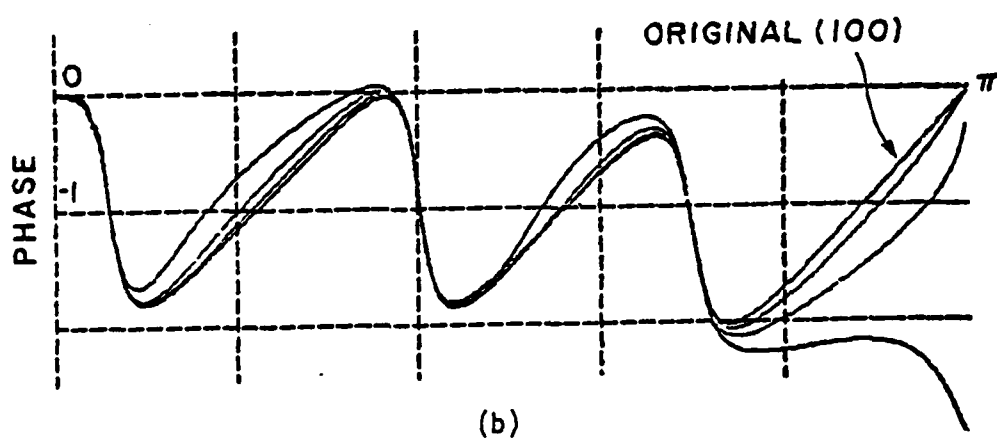
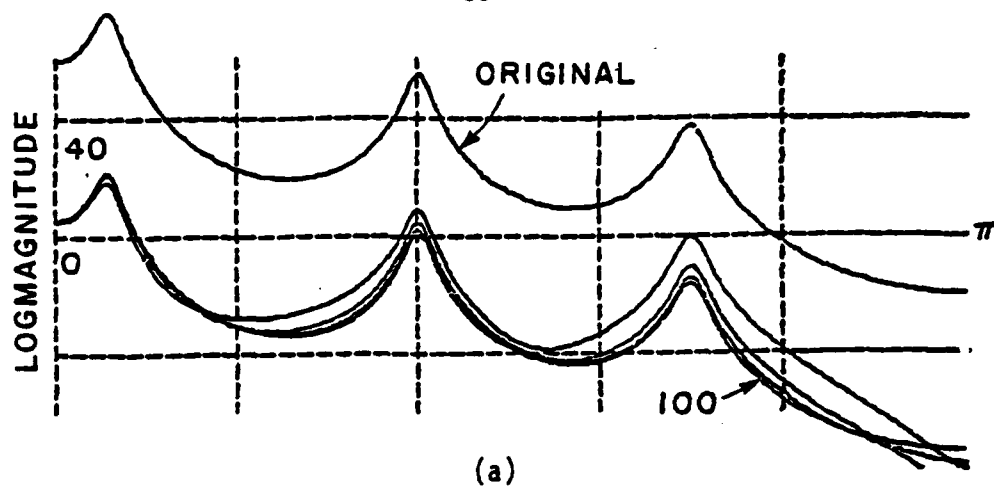


Fig. 4.5 (a) Convergence of $\log[M_k(\omega)]$ in example 4.3
 (b) Convergence of $\theta_k(\omega)$ in example 4.3

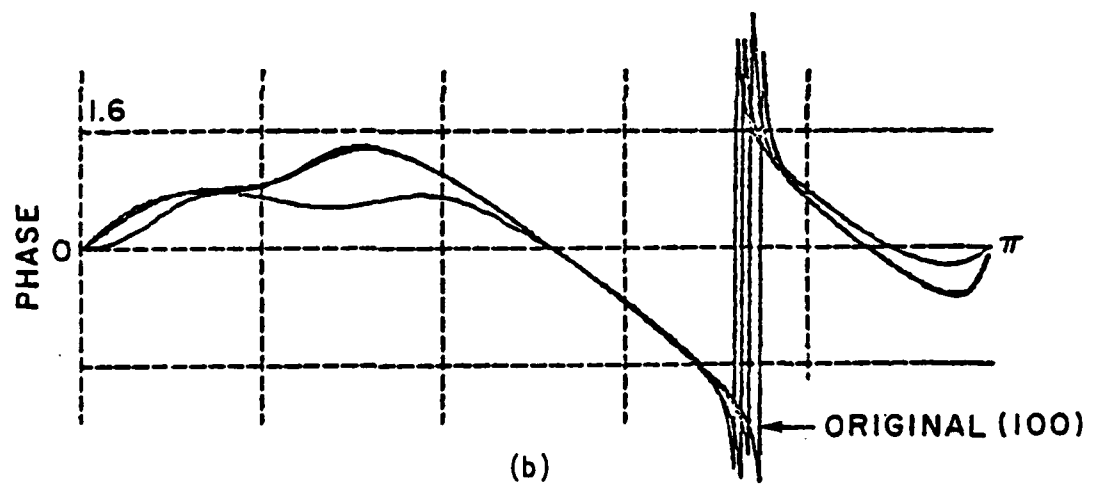
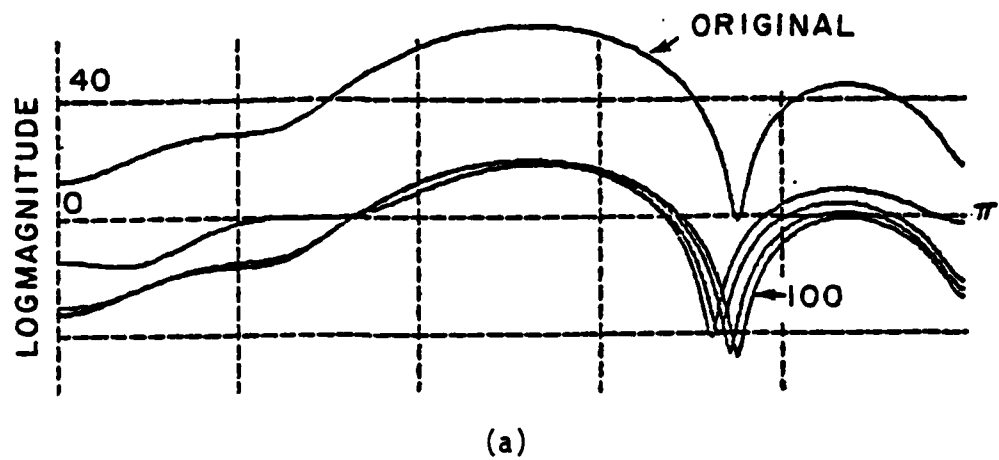


Fig. 4.6 (a) Convergence of $\log[M_k(\omega)]$ in example 4.4
 (b) Convergence of $\theta_k(\omega)$ in example 4.4

4.2.3 A New Phase Unwrapping Algorithm

We are now in a position to state a new phase "unwrapping" algorithm without modulo 2π considerations. Suppose we are given the principal value of the phase of $H(\omega)$ and that the corresponding unwrapped phase is continuous. Our new unwrapping algorithm is outlined below. $M_{mp}(\omega)$ in step (i) denotes the magnitude of the Fourier transform of a minimum phase sequence derived from our iterative procedure under a causality constraint. Step (i) yields the same magnitude (within a scale factor) that would be obtained by applying the Hilbert transform to the unwrapped phase, but bypasses the need of phase unwrapping.

Phase Unwrapping Algorithm

- (i) Apply the iterative algorithm of section 4.2.2.1 with a causality constraint: $M_{mp}(\omega)$
 - (ii) Compute the logarithm of the resulting magnitude function from step (i): $\log[M_{mp}(\omega)]$
 - (iii) Hilbert transform the logmagnitude function derived in step (ii) to obtain the desired unwrapped phase: $\theta_h(\omega)$
-

There are two major considerations in the use of this algorithm. First from our discussion in section 2.3.2, we see that, in general, the minimum phase solution derived from the iteration is of infinite extent regardless of whether the original sequence $h(n)$ is of finite duration. Therefore, a possible problem with aliasing arises. The DFT length must be sufficiently large so that the minimum phase counterpart of $h(n)$, $h_{mp}(n)$ has essentially decayed to zero. In particular, when $h_{mp}(n) \approx 0$ for $n > N$,

the DFT length from the results of section 4.2.1 should be at least $2N-1$.

The second consideration is the linear phase component of $H(\omega)$. The presence of this term represents a potential drawback to the algorithm since a priori knowledge of such a component is often difficult to obtain.

CHAPTER 5

ESTIMATION OF THE UNWRAPPED PHASE FROM HARMONIC SAMPLES

In section 4.1.2 we described a direct approach to phase estimation through homomorphic deconvolution. This approach relies on the phase of the quasi-periodic waveform. The magnitude estimate is found separately and is not used in deriving the phase estimate. The procedure yields an estimate which is bandlimited and is an accurate representation of the desired unwrapped phase.

An implicit assumption in this direct approach is that the pulse train $p(n)$ of (2.6) is not exactly periodic. When $p(n)$ is periodic with period P , $P(\omega)$ is impulsive and periodic with period $2\pi/P$. Under this condition, $P(\omega)$ samples the desired system function $H(\omega)$ at harmonically related frequencies, $\omega_k = 2\pi k/P$, and corresponds to a periodic $x(n)$. Since $X(\omega)$ is zero in frequency bands, the complex cepstrum does not exist [33], and thus our direct approach cannot be applied.

Nevertheless, this situation is useful in representing voiced speech over a short duration in time. Typically, within a voiced speech segment, the vocal tract and vocal cord characteristics are slowly varying. Therefore, voiced speech over a short duration (e.g. 20 msec.) can be modeled as a segment of an infinitely long periodic waveform, $\hat{x}(n) = w(n) x(n)$, where $w(n)$ is a finite length, unity amplitude window over which the vocal tract and vocal cords are time-invariant.

Certainly with only samples of $H(\omega)$, without additional constraints, there exist an infinite number of choices for $H(\omega)$ and likewise for the phase of $H(\omega)$. The number of samples generally may not be sufficient to

uniquely specify the parameters of a rational model through, for example, techniques similar to those in sections 3.2 and 4.2. Alternatively, in this chapter we take a direct nonparametric approach to resolving this ambiguity.

Specifically, we shall view the problem of phase estimation as a problem of polynomial interpolation over harmonic samples of the real and imaginary components of $H(\omega)$, or when only phase samples are given, of the all-pass function $H(\omega)/|H(\omega)|$. In particular we apply linear interpolation over two successive samples. Conditions on $\theta_h(\omega)$ are derived under which its corresponding unwrapped phase at harmonic samples is preserved by this simple procedure - a situation denoted as "phase tracking". Phase tracking preserves the slowly varying component, i.e. the envelope, of the unwrapped phase of $H(\omega)$. With an appropriate bandlimited constraint the entire phase function may then be recovered.

Linear interpolation is also useful in understanding the properties of an arbitrary window $w(n)$ and its relation to the unwrapped phase of the windowed waveform $x(n) = w(n) x(n)$. Windowing can be viewed as an interpolation procedure in the frequency domain, and, in fact, is itself a means of phase estimation. With this viewpoint, constraints on the window duration and alignment (i.e. positioning with respect to $x(n)$) are imposed for phase tracking.

In the final section of this chapter, with appropriate a priori information, a method is derived for phase tracking by windowing without the need of alignment.

5.1 Techniques of Phase Tracking

Consider a periodic train of pulses $p(n)$ with spacing P . With

$p(0) \neq 0$, $p(n)$ in the frequency domain is given by

$$P(\omega) = \frac{2\pi}{P} \sum_k \delta(\omega - \omega_k) \quad (5.1)$$

where $\omega_k = 2\pi k/P$ and $\delta(\omega)$ is the dirac delta function.

From section 2.2.3, we can write the unwrapped phase* of $H(\omega)$ as

$$\theta_h(\omega) = \theta_{\tilde{h}}(\omega) + n_0\omega \quad \omega \in [0, \pi] \quad (5.2)$$

where $\theta_{\tilde{h}}(\omega)$ is continuous and $n_0\omega$ is a linear phase component. For convenience, we express the Fourier transform of $h(n)$ as a function of two variables:

$$H(\omega, n_0) = \tilde{H}(\omega) \exp[jn_0\omega] \quad (5.3)$$

where

$$\tilde{H}(\omega) = |H(\omega, n_0)| \exp[j\theta_{\tilde{h}}(\omega)] \quad (5.4)$$

In the frequency domain we also express the Fourier transform of the periodic waveform $x(n) = p(n)*h(n)$ as a function of two variables:

$$X(\omega, n_0) = H(\omega, n_0)P(\omega) \quad (5.5)$$

Therefore, from (5.1) and (5.5), we have

$$X(\omega, n_0) = \frac{2\pi}{P} \sum_k H(\omega_k, n_0) \delta(\omega - \omega_k) \quad (5.6)$$

and so only samples of $H(\omega, n_0)$ are available.

We have generally assumed throughout this thesis that $h(n)$ is causal, nonzero at the origin, and has a rational z -transform. In particular, when $h(n)$ is a mixed phase sequence, we see from (3.1) and

*Throughout this chapter a phase function $\theta(\omega)$ denotes an unwrapped phase.

AD-A087 670

MASSACHUSETTS INST OF TECH CAMBRIDGE RESEARCH LAB OF--ETC F/G 5/8
PHASE ESTIMATION WITH APPLICATION TO SPEECH ANALYSIS-SYNTHESIS.(U)
NOV 79 T F QUATIERI N00014-75-C-0951
TR-491 NL

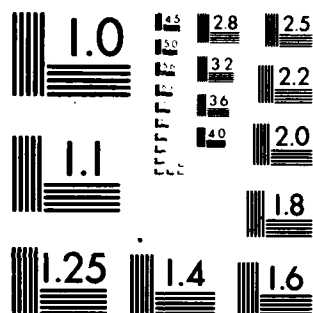
UNCLASSIFIED

2 - 2

2 - 2



END
DATE
FILMED
9 80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(3.4) that n_0 equals the number of zeros of $H(z)$ outside the unit circle. However, given an arbitrary segment of a periodic waveform of the form $x(n) = h(n)*p(n)$, we must define an origin. Consequently, without additional knowledge, the position of $h(n)$ relative to our defined origin is arbitrary, and therefore n_0 in (5.2) is also arbitrary. Because $x(n)$ is periodic, however, we can restrict n_0 to the range $-P/2 \leq n_0 < P/2$.

Since, conceptually $x(n)$ is of infinite extent, we must apply a window, $w(n)$. Throughout this chapter, we assume $w(n)$ is symmetric and that the time origin (i.e. $n=0$) is set at the center of $w(n)$. Therefore, changing n_0 corresponds to sliding $x(n)$ under the window $w(n)$.

In this section, we consider two methods for phase tracking, that is, for retrieving the unwrapped phase of $H(\omega, n_0)$ at harmonic samples. The first method of linear interpolation can be viewed as a special case of the second method of time-domain windowing. These techniques are also applicable to samples of the all-pass function $H(\omega, n_0)/|H(\omega, n_0)|$, and so do not rely necessarily on magnitude information.

5.1.1 Linear Interpolation in the Frequency Domain

One approach to estimating $H(\omega, n_0)$, and thus the phase $\theta_h(\omega)$ is to fit a polynomial of order M to M given samples of $H(\omega_k, n_0)$ over a specified interval. For example, one possibility is to fit a first order polynomial over two successive samples, i.e. perform linear interpolation on $H(\omega_k, n_0)$ and $H(\omega_{k+1}, n_0)$. We choose this particular scheme for three reasons which will become clear in the sequel:

- (i) Linear interpolation illustrates the fundamental problems in preserving the unwrapped phase of $H(\omega, n_0)$ at harmonics by time-domain windowing $x(n)$.

(ii) The unwrapped phase at harmonics derived by linear interpolation equals the unwrapped phase found by a running sum on the principal value of the difference between two successive samples of $\theta_h(\omega)$.

(iii) Phase estimation by linear interpolation is compatible with the spectral envelope speech analysis-synthesis system given in Chapter 6.

A linear interpolation procedure is equivalent to convolving the real and imaginary components of $X(\omega, n_0)$ with a triangular function given by

$$W(\omega) = \begin{cases} 1 - |\omega/(2\pi/P)| & |\omega| \leq 2\pi/P \\ 0 & |\omega| > 2\pi/P \end{cases} \quad (5.7)$$

The estimate of $H(\omega, n_0)$, $\hat{H}(\omega, n_0)$ is therefore expressed as

$$\begin{aligned} \hat{H}(\omega, n_0) &= X(\omega, n_0) * W(\omega) = \frac{2\pi}{P} \sum_k H(\omega_k, n_0) W(\omega - \omega_k) \\ &= \frac{2\pi}{P} \sum_k H_r(\omega_k, n_0) W(\omega - \omega_k) + j \frac{2\pi}{P} \sum_k H_i(\omega_k, n_0) W(\omega - \omega_k) \end{aligned} \quad (5.8)$$

In the time domain, (5.8) is equivalent to multiplying (i.e. windowing) $x(n)$ with a sequence given by

$$w(n) = [\sin(\pi n/P)/\pi n]^2 \quad (5.9)$$

We say that the unwrapped phase of $H(\omega, n_0)$ is tracked by $\hat{H}(\omega, n_0)$ at $\omega = \omega_k$ when the unwrapped phase of $\hat{H}(\omega_k, n_0)$ equals the unwrapped phase of $H(\omega_k, n_0)$:

$$\theta_{\hat{H}}(\omega_k) = \theta_h(\omega_k) \quad (5.10)$$

Alternatively, we refer to (5.10) as the preservation of the unwrapped phase envelope.

From (5.2), the phase difference $\Delta\theta_h(\omega_k)$ between two successive samples of $\theta_h(\omega)$ is written as

$$\begin{aligned}\Delta\theta_h(\omega_k) &= \theta_h(\omega_k) - \theta_h(\omega_{k-1}) \\ &= \theta_h^-(\omega_k) - \theta_h^-(\omega_{k-1}) + n_0 (\omega_k - \omega_{k-1}) \\ &= \Delta\theta_h^-(\omega_k) + 2\pi n_0/P\end{aligned}\quad (5.11)$$

We shall show that a necessary and sufficient condition for $\hat{H}(\omega, n_0)$ to track the unwrapped phase of $H(\omega, n_0)$ at $\omega=\omega_k \forall_k$ is given by

$$\begin{aligned}\hat{\theta}_h(\omega_k) &= \theta_h(\omega_k) \forall_k \quad \text{if and only if} \\ |\Delta\theta_h(\omega_k)| &= |\Delta\theta_h^-(\omega_k) + 2\pi n_0/P| < \pi \forall_k\end{aligned}\quad (5.12)$$

To prove this condition we first note from (5.8) that a linear trajectory is followed by the real and imaginary components of $\hat{H}(\omega, n_0)$ over each interval $[\omega_k, \omega_{k+1}]$. The polar plot of $\hat{H}(\omega, n_0)$ for $\omega \in [\omega_k, \omega_{k+1}]$ therefore also follows a linear trajectory as depicted by trajectories T_1 and T_2 in Figs. 5.1a and 5.1b. Furthermore, we see from Fig. 5.1a that when $|\Delta\theta_h(\omega_k)| < \pi$, the polar plot of $\hat{H}(\omega, n_0)$ is of type T_1 , and therefore $\Delta\hat{\theta}_h(\omega_k) = \Delta\theta_h(\omega_k)$. On the other hand, when $|\Delta\theta_h(\omega_k)| > \pi$, the polar plot of $\hat{H}(\omega, n_0)$ is of type T_2 , and therefore $\Delta\hat{\theta}_h(\omega_k) = \Delta\theta_h(\omega_k) + 2\pi M_k$ where M_k is an integer.

$\hat{\theta}_h(\omega_k)$ can now be expressed by the following running sum:

$$\hat{\theta}_h(\omega_k) = \sum_{l=0}^k \Delta\hat{\theta}_h(\omega_l)$$

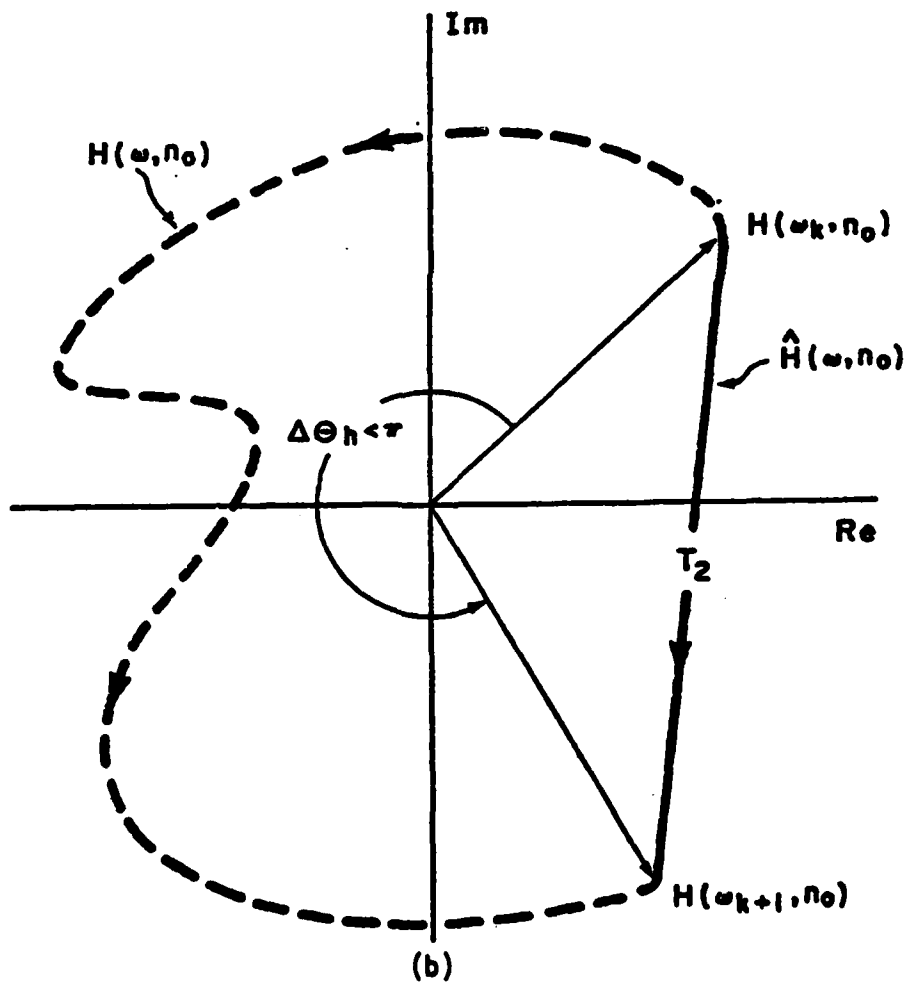
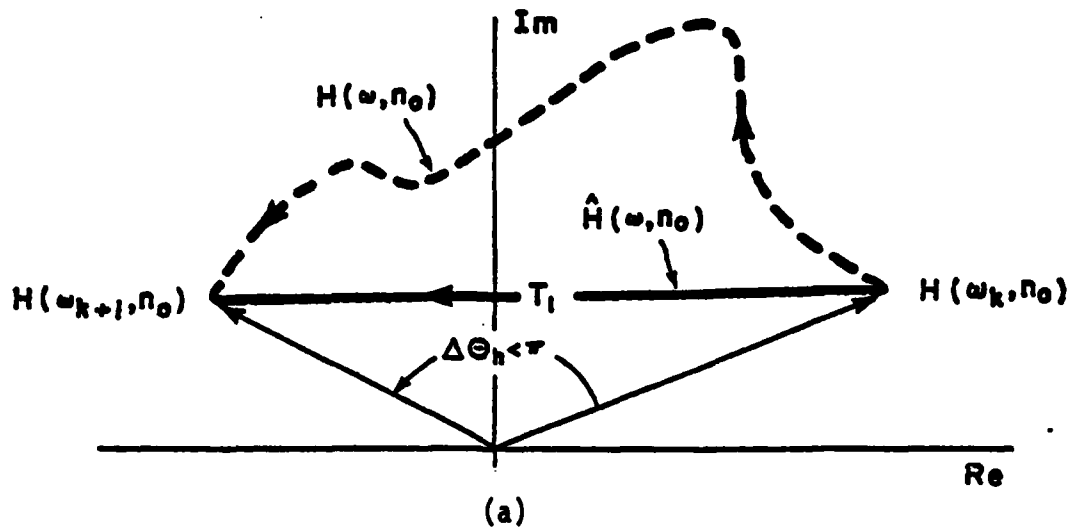


Fig. 5.1 (a) Polar plot of $\hat{H}(\omega, n_0)$ for $\Delta\Theta_h(\omega_k) < \pi$
 (b) Polar plot of $\hat{H}(\omega, n_0)$ for $\Delta\Theta_h(\omega_k) > \pi$

$$\begin{aligned}
 &= \sum_{\ell=0}^k \Delta\theta_h(\omega_\ell) + \sum_{\ell=0}^k 2\pi M_\ell \\
 &= \theta_h(\omega_k) + \theta_{2\pi}(\omega_k)
 \end{aligned} \tag{5.13}$$

where $\theta_{2\pi}(\omega_k)$ termed the " 2π error accumulator function" is given by

$$\theta_{2\pi}(\omega_k) = 2\pi \sum_{\ell=0}^k M_\ell \tag{5.14}$$

Thus if $|\Delta\theta_h(\omega_k)| < \pi \psi_k$, from above $M_k=0 \psi_k$, and so from (5.14) $\theta_{2\pi}(\omega_k)=0 \psi_k$. Therefore, from (5.13) $\theta_h^{\wedge}(\omega_k)=\theta_h(\omega_k) \psi_k$. On the other hand, if $|\Delta\theta_h(\omega_k)| > \pi$ for some particular k , $M_k \neq 0$, and from (5.14) $\theta_{2\pi}(\omega_k) \neq 0 \psi_k$. Finally, from (5.13) $\theta_h^{\wedge}(\omega_k) \neq \theta_h(\omega_k) \psi_k$, and our proposition is proven.

Since, in general, $\Delta\theta_h^{\wedge}(\omega_k)=\Delta\theta_h(\omega_k)+2\pi M_k$ and from Fig. 5.1, $|\Delta\theta_h^{\wedge}(\omega_k)| < \pi$, we conclude from (2.12b) that $\Delta\theta_h^{\wedge}(\omega_k)=PV[\Delta\theta_h(\omega_k)]$ where PV denotes "principal value of". From 5.13 we then have

$$\theta_h^{\wedge}(\omega_k) = \sum_{\ell=0}^k PV[\Delta\theta_h(\omega_\ell)] \tag{5.15}$$

We see from condition (5.12) that the linear phase component $n_0\omega$ plays a major role in phase tracking. That is, the position of $x(n)$ relative to the window's center is significant in preserving the unwrapped phase at harmonics. In Chapter 6, we shall find that the condition $|\Delta\theta_h^{\wedge}(\omega_k)| < \pi$ is "almost always" true for speechlike harmonics. Under this constraint, it directly follows from (5.12) that a necessary and sufficient condition for phase tracking can be stated as

$\hat{\theta}_h(\omega_k) = \theta_h(\omega_k)$ if and only if n_0 falls within the

range given by

(5.16)

$$-P/2 - P\min\Delta\theta_h^-(\omega_k)/2\pi \leq n_0 \leq P/2 - P\max\Delta\theta_h^-(\omega_k)/2\pi$$

where min and max denote "minimum or maximum value of", respectively.

When n_0 falls within the bounds of (5.16) we say that $x(n)$ is "aligned" with our defined origin. If $n_0 \neq 0$, but satisfies (5.16), the unwrapped phase of $\theta_h(\omega_k)$ is tracked and includes the linear phase component $n_0\omega_k$. We can however remove this term to obtain $\theta_h^-(\omega_k)$ since from (5.2)

$$n_0^* = \left. \theta_h(\omega_k)/\pi \right|_{\omega_k=\pi} \quad (5.17)$$

We thus far have investigated the behavior of the unwrapped phase of $\hat{H}(\omega, n_0)$ at harmonics. We now shall demonstrate that $\hat{\theta}_h(\omega)$ follows a monotonic trajectory across each interval $[\omega_k, \omega_{k+1}]$. To see this we consider the phase derivative, $\dot{\hat{\theta}}_h(\omega)$ weighted by the squared magnitude of $\hat{H}(\omega, n_0)$. From (2.14), $|\hat{H}(\omega, n_0)|^2 \dot{\hat{\theta}}_h(\omega)$ is expressed by

$$\begin{aligned} |\hat{H}(\omega, n_0)|^2 \dot{\hat{\theta}}_h(\omega) &= M_1[M_r(\omega - \omega_k) + \hat{H}_r(\omega_k, n_0)] \\ &\quad - M_r[M_1(\omega - \omega_k) + \hat{H}_1(\omega_k, n_0)], \quad \omega \in [\omega_k, \omega_{k+1}] \end{aligned} \quad (5.18)$$

* We assume P is even so that $H(\omega, n_0)$ is sampled at $\omega = \pi$.

where M_r and M_i are slopes of the real and imaginary components of $\hat{H}(\omega, n_0)$ in the interval $[\omega_k, \omega_{k+1}]$ [28]. Expanding (5.18), we obtain

$$\begin{aligned} |\hat{H}(\omega, n_0)|^2 \dot{\theta}_{\hat{H}}(\omega) &= \hat{H}_r(\omega_k, n_0) M_i - \hat{H}_i(\omega_k, n_0) M_r \\ &= c, \quad \omega \in [\omega_k, \omega_{k+1}] \end{aligned} \quad (5.19)$$

where c is a constant. Since $|\hat{H}(\omega, n_0)|^2 > 0$, $\dot{\theta}_{\hat{H}}(\omega)$ from (5.19), is either always positive, negative, or zero in the interval $[\omega_k, \omega_{k+1}]$. Because $\theta_{\hat{H}}(\omega)$ is given by integration of the phase derivative, $\theta_{\hat{H}}(\omega)$ is therefore either monotone increasing or decreasing in our specified interval, and we have proven our proposition.

Throughout this section we have assumed knowledge of samples of $H(\omega, n_0)$. We may however apply linear interpolation directly to samples of the all-pass function $H(\omega)/|H(\omega)|$ and thus discard the magnitude information altogether in our procedure. Clearly, the unwrapped phase at samples derived from both schemes is identical. The difference in the estimate lies in the precise path of the monotonic trajectory from sample to sample.

5.1.1.1 Examples

Figure 5.2a depicts an unwrapped phase function $\theta_{\hat{H}}(\omega) = \theta_{\tilde{H}}(\omega) + n_0 \omega$ of an all-pass system function $H(\omega, n_0)$, with $n_0 = 0$, which consists of factors of the form (2.22). In the following example, we consider estimation of $\theta_{\hat{H}}(\omega)$ by linear interpolation of $H(\omega_k, n_0)$ for various sampling frequencies $2\pi/P$ and linear phase components $n_0 \omega$. Any linear phase in the estimate $\theta_{\hat{H}}(\omega)$ is removed by subtracting $\omega \theta_{\hat{H}}(\pi)/\pi$ to obtain an estimate of $\theta_{\tilde{H}}(\omega)$.

Example 5.1

For a sampling frequency of $2\pi/P \leq 2\pi/130$, $|\Delta \theta_{\tilde{H}}(\omega_k)| > \pi$ at roughly

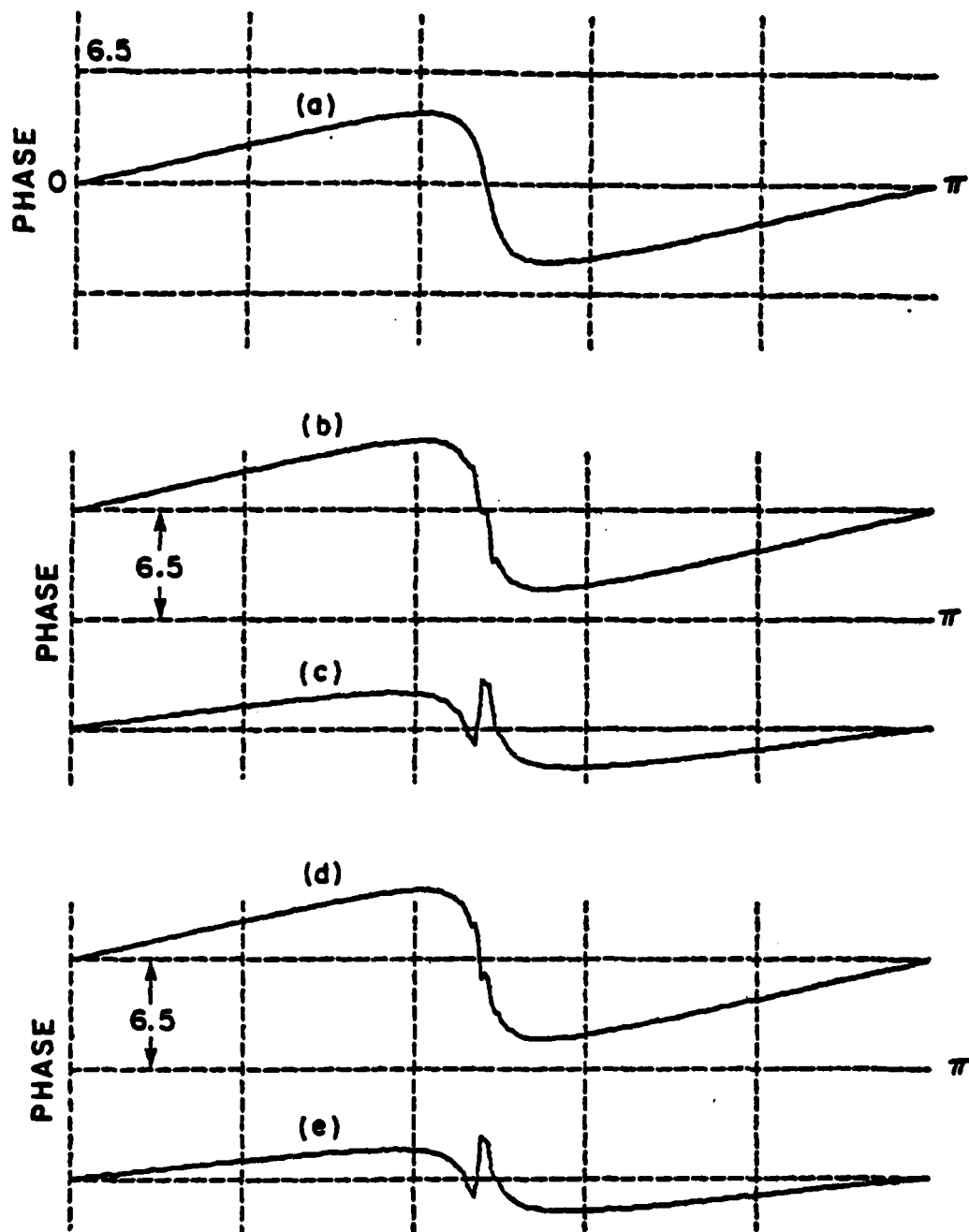


Fig. 5.2 (a) Unwrapped phase of a second order all-pass system function, (b) Estimate of (a) in example 5.1 with $n_0=0$ and $P=130$, (c) same as (b) with $P=129$, (d) same as (b) with $n_0=-6$ and $P=150$, (e) same as (d) with $n_0=-7$.

1.5 radians which occurs in the decreasing region of $\theta_h^-(\omega)$ depicted in Fig. 5.2a. Furthermore, $\theta_h^-(\omega)$ is such that $|\Delta\theta_h^-(\omega_k)| < 2\pi \forall_k$ so that $|\Delta\theta_h^-(\omega_k)| > \pi$ for one and only one value of k within this region. Figures 5.2b and 5.2c illustrate the estimate $\theta_h^-(\omega)$ (with any linear phase component removed) for sampling frequencies $2\pi/130$ and $2\pi/129$, respectively. Note the 2π error at approximately 1.5 radians for sampling frequency $2\pi/129$.

With $P = 150$, $|\Delta\theta_h^-(\omega_k)| < \pi \forall_k$. Changing n_0 then illustrates the effect of the linear phase term $n_0\omega$ in (5.2). We saw from (5.16) that n_0 is constrained within a certain range for phase tracking. In particular, for our example, n_0 can be shown to be restricted roughly to the set $n_0 \in [-6, 72]$ by computing approximate minimum and maximum increments of $\Delta\theta_h^-(\omega_k)$ for $P=150$. Specifically, consider the lower bound. Figures 5.2d and 5.2e illustrate the 2π error which arises when n_0 decreases from -6 to -7.

This example illustrates the extreme sensitivity of the estimate $\theta_h^-(\omega)$ to changing either P or n_0 by as little as one point.

5.1.2 Windowing in the Time Domain

The linear interpolation procedure of the previous section can be interpreted in the time domain as multiplication of $x(n)$ by the window given in (5.9). Consider now the same window but which is not necessarily a function of the pitch period:

$$w_M(n) = [\sin(\pi n/M)/\pi n]^2 \quad (5.20)$$

where the parameter M is variable. The Fourier transform of $w_M(n)$, $W_M(\omega)$ is given by the triangular function of (5.7) where P is replaced by M . Before proceeding to more general windows, we first show that when

$P < M < 2P$, the unwrapped phase at harmonics derived from multiplication by $w_M(n)$ is identical to that derived in the frequency domain from the linear interpolation scheme of the previous section.

Figure 5.3a illustrates the real component of $\hat{H}(\omega, n_0)$ in (5.8) where successive weighted transforms $w_M(\omega - \omega_k)$ are partially overlapping, i.e. $P < M < 2P$. In a region of no overlap, because $w_M(\omega)$ is real and positive the unwrapped phase of $\hat{H}(\omega, n_0)$, as depicted in Fig. 5.3b, is constant:

$$\hat{H}(\omega, n_0) = H(\omega_k, n_0) w_M(\omega - \omega_k) \quad |\omega - \omega_k| < \epsilon/2 \quad (5.21a)$$

and so

$$\theta_{\hat{H}}(\omega) = \theta_{\hat{H}}(\omega_k) \quad |\omega - \omega_k| < \epsilon/2 \quad (5.21b)$$

where ϵ equals the length of the nonoverlapping region. Therefore, in regions of no overlap, $\theta_{\hat{H}}(\omega)$ exhibits what we shall refer to as "harmonic plateaus".

As in section 5.1.1 a linear trajectory is followed by the real and imaginary components of $\hat{H}(\omega, n_0)$ in each region of overlap, as depicted in Fig. 5.3a. Therefore, a linear trajectory is also followed by the polar plot of $\hat{H}(\omega, n_0)$ in these regions. From (5.21a), the magnitude of $\hat{H}(\omega, n_0)$ at the endpoints of the overlapping region is scaled equally. Consequently, the polar plot of $\hat{H}(\omega, n_0)$ is a scaled version with the same slope of that derived by linear interpolation of harmonics (i.e. $M=P$). The unwrapped phase at each harmonic plateau therefore must equal the unwrapped phase derived from our procedure with $M=P$. Thus, all properties given in section 5.1.1 hold also for $P < M < 2P$.

Let us now suppose that we are given an arbitrary finite length symmetric window, with Fourier transform $W(\omega)$. From (5.8) we can view window-

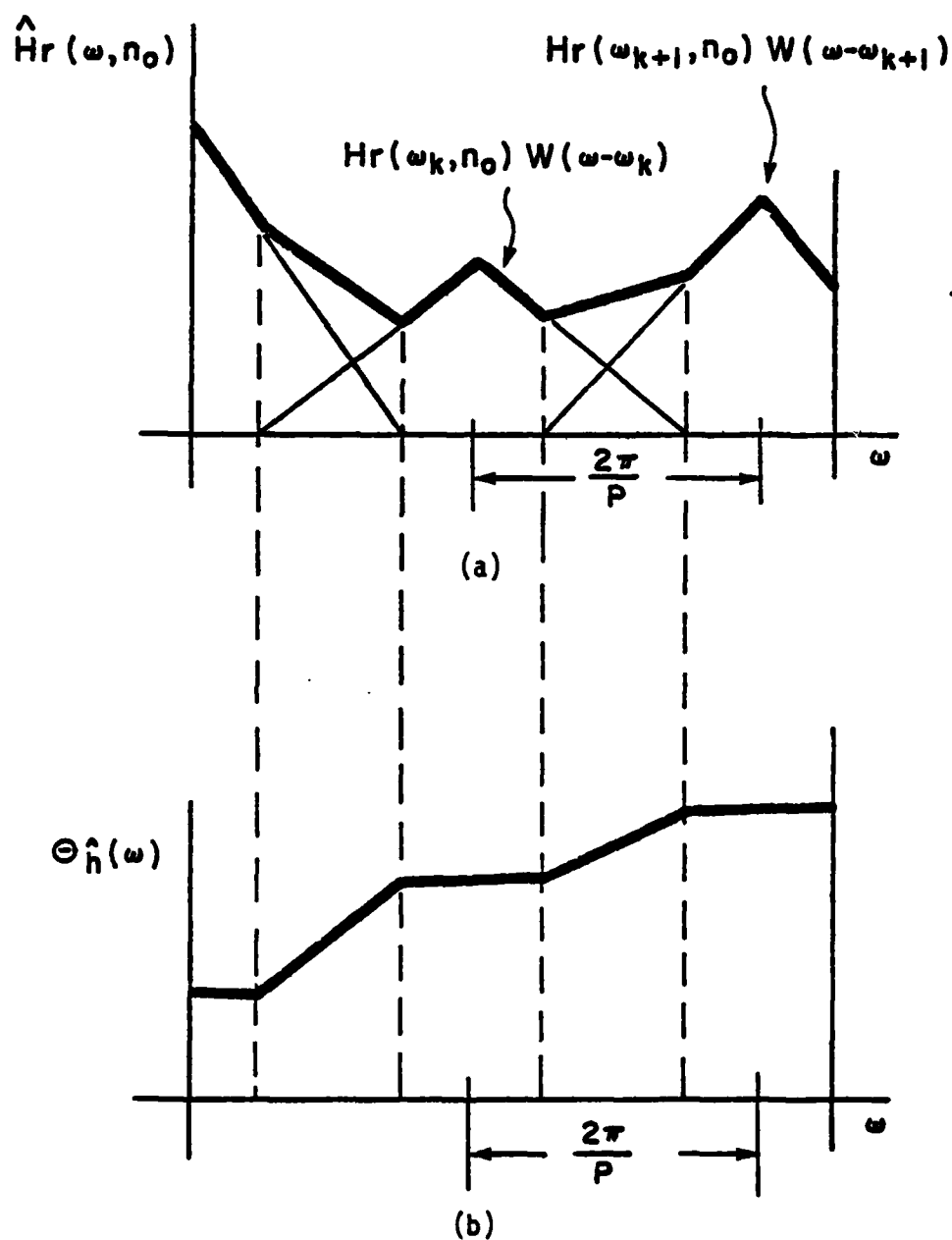


Fig. 5.3 (a) $\hat{H}_r(\omega, n_0)$ derived from the window $w_M(n)$ in (5.20)
 (b) Approximate unwrapped phase corresponding to (a)

ing itself as a phase estimation procedure. The phase trajectory between harmonics is determined by the mainlobe and sidelobe structure of $W(\omega)$ and the degree to which $W(\omega - \omega_k)$ and $W(\omega - \omega_{k+1})$ overlap. If we approximate the trajectory of $\hat{H}(\omega, n_0)$ as linear in a region where mainlobes "significantly" overlap, and if we assume that outside this region the phase is constant as in (5.21b), then the phase trajectory follows that derived with $w_M(n)$.

For example, the Fourier transform of a Hamming window has a mainlobe which can be approximated roughly by the triangular function $w_M(\omega)$. Furthermore, the duration of a Hamming window should be less than about four pitch periods. Otherwise there is no region of mainlobe overlap and the phase trajectory is determined primarily by sidelobe structure, and weights $H(\omega_k, n_0)$ and $H(\omega_{k+1}, n_0)$. We have found such approximations to be useful in describing the empirical behavior of the unwrapped phase of a periodic waveform multiplied by a Hamming window.

5.1.2.1 Examples

In the following example we consider the estimation of the unwrapped phase function $\theta_h(\omega) = \theta_h^*(\omega) + n_0\omega$ of $H(\omega)$ from example 4.1, and which is depicted in Fig. 5.4a with $n_0 = 0$. We apply a Hamming window and demonstrate the effect of modifying its duration and its position with respect to the periodic waveform $x(n) = h(n) * p(n)$ where $p(n)$ has a period of 50 points. $\theta_h^*(\omega)$ is such that $|\Delta\theta_h^*(\omega_k)| < \pi, \forall_k$.

Example 5.2

Figures 5.4b and 5.4c depict the continuous unwrapped phase estimates obtained with application of a Hamming window of length 2 and 3.9 pitch periods, respectively, and where $n_0 = 0$. Note that when P equals two pitch

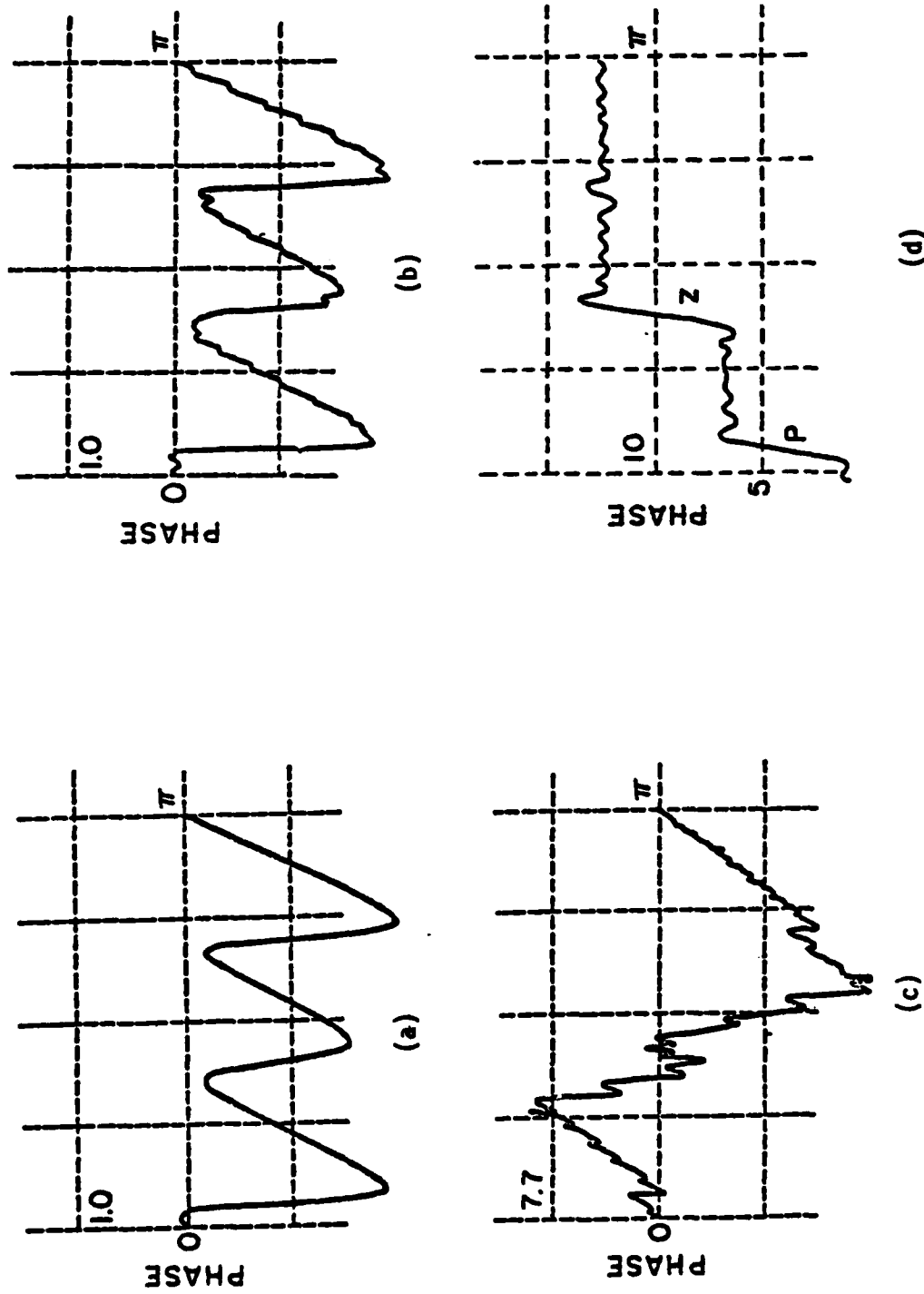


Fig. 5.4 (a) Unwrapped phase of a system function with two complex pole pairs and one complex zero pair, (b) Estimate of (a) in example 5.2 with a 100 point Hamming window where $n_0 = 0$, (c) Same as (b) with a 390 point Hamming window, (d) Error function for estimate of (a) in example 5.2 with a 100 point Hamming window where $n_0 = -20$.

periods, the phase trajectory is generally monotonic between regions where the phase is roughly constant near harmonics.

Figure 5.4d depicts the error function $\theta_e(\omega)$ given by

$$\theta_e(\omega) = \theta_h^-(\omega) - \theta_h(\omega) \quad (5.22)$$

for the unwrapped phase derived with a Hamming window of duration two pitch periods, but where $n_0 = -20$. (5.22) may be viewed as a continuous counterpart to the 2π error accumulator function of (5.14) where $\theta_e(\omega_k) = \theta_{2\pi}(\omega_k)$.

From Fig. 5.4d we deduce that ω_{n_0} is such that $|\Delta\theta_h(\omega_k)| = |\Delta\theta_h^-(\omega_k) + 2\pi n_0/P| > \pi$ in the region of a minimum phase pole (denoted by P) and a maximum phase zero (denoted by Z). Consequently, there arises an error of 2π in both regions.

5.1.3 Comments on Comparing Magnitude and Phase Estimation by Windowing

We have seen in sections 5.1.1 and 5.1.2 that linear interpolation of samples of a Fourier transform through windowing with $w_M(n)$ generates a monotonic phase trajectory in the interval $[\omega_k, \omega_{k+1}]$. Such monotonicity is typical of speechlike phase between harmonics. That is, we do not expect "ripples" in the phase function, but a "smooth" behavior between successive harmonics.

Let us now address magnitude estimation by this interpolation procedure. In particular, let us suppose that we are given two successive samples of $H(\omega, n_0)$, one to the right and one to the left of a pole located close to the unit circle. Furthermore, we assume $n_0=0$ and that $|\Delta\theta_h(\omega_k)| = |\Delta\theta_h^-(\omega_k)| \approx \pi$.

From Fig. 5.1, we see that the magnitude of $\hat{H}(\omega, n_0)$ generates a null where in fact there exists a peak in $H(\omega, n_0)$ due to the presence of a pole.

Such "pole splitting" is therefore potentially a problem with this technique. Linear interpolation applied directly to magnitude samples may thus be preferred in such cases.

5.2 On The Problem of Alignment

In section 5.1.1 we derived the bound (5.16) on n_0 to ensure phase tracking under the constraint $|\Delta\theta_h^-(\omega_k)| < \pi \psi_k$. We referred to this condition as alignment of $x(n)$. A number of examples were given in the previous sections illustrating the importance of this requirement. Our main purpose in this section is to demonstrate that alignment is an inherently ambiguous process with the sole constraint that $|\Delta\theta_h^-(\omega_k)| < \pi$. Thus without additional a priori knowledge, an alignment guess must be made.

In section 5.2.2 a method is described for phase tracking without the need of alignment, under appropriate constraints on the second difference of $\theta_h^-(\omega_k)$.

5.2.1 Alignment Ambiguity

With the constraint $|\Delta\theta_h^-(\omega_k)| < \pi \psi_k$, we would like to determine by inspection of $\theta_h^-(\omega_k)$ whether n_0 falls within the bounds of (5.16) so that phase tracking is guaranteed. From (5.13) our problem is equivalent to detecting the presence of $\theta_{2\pi}(\omega_k)$ which is zero if and only if n_0 satisfies (5.16).

One situation which can arise and lead to an erroneous conclusion about alignment is the case where $\theta_h^-(\pi)=0$ which occurs when $n_0=0$. However, suppose $n_0 = -\theta_{2\pi}(\pi)/\pi$. Then from (5.13) the 2π error accumulator function $\theta_{2\pi}(\omega_k)$ cancels the linear phase component at $\omega=\pi$. Therefore, $\theta_h^-(\pi)=0$ is not a necessary condition for $n_0=0$, nor for phase tracking.

This case is indicative of the problem of alignment ambiguity. More generally, we see from Fig. 5.1 that without additional a priori knowledge, we cannot determine whether the true net phase trajectory between harmonics is clockwise or counterclockwise. This ambiguity arises since when $n_0\omega$ is sufficiently large, $|\Delta\theta_h(\omega_k)| > \pi$.

In conclusion, we cannot detect the presence of $\theta_{2\pi}(\omega_k)$ from only $\theta_h(\omega_k)$.

5.2.2 Phase Tracking Without Alignment

One method of phase tracking which does not depend on alignment capitalizes on the constraint that the second difference of $\theta_h(\omega_k)$ does not exceed π . This constraint is natural for speechlike phase because it removes the possibility of "large" ripples in the unwrapped phase function. Our technique provides a way to undo the effect of misalignment and relies primarily on the fact that second differencing of samples of the unwrapped phase effectively eliminates the linear phase contribution at each sample.

With some algebraic manipulation the second difference of $\theta_h(\omega_k)$ can be expressed by

$$\Delta^2\theta_h(\omega_k) = \Delta\theta_h(\omega_k) - \Delta\theta_h(\omega_{k-1}) \quad (5.23)$$

$$= \Delta^2\theta_h^-(\omega_k) + n_0\omega_1\delta(\omega_k - \omega_1)$$

$$+ \theta_{2\pi}(\omega_k) - 2\theta_{2\pi}(\omega_{k-1}) + \theta_{2\pi}(\omega_{k-2})$$

and so the linear phase component is removed except at $\omega=\omega_1=2\pi/P$.

With the constraints,

$$|\Delta^2\theta_h^-(\omega_k)| < \pi \quad (5.24a)$$

$$|\theta_h^-(\omega_1) + \frac{2\pi n_0}{p}| < \pi \quad (5.24b)$$

and, the initial condition,

$$\Delta^2 \theta_h^-(\omega_1) = \Delta \theta_h^-(\omega_1) = \theta_h^-(\omega_1) \quad (5.25)$$

we can write the principal value of $\Delta^2 \theta_h^-(\omega_k)$ as

$$\Delta^2 \phi_h^-(\omega_k) = PV[\Delta^2 \theta_h^-(\omega_k)] = \Delta^2 \theta_h^-(\omega_k) + \frac{2\pi n_0}{p} \delta(\omega_k - \omega_1) \quad (5.26)$$

The running sum of (5.26) is taken to obtain

$$\Delta \phi_h^-(\omega_k) = \Delta \theta_h^-(\omega_k) + \frac{2\pi n_0}{p} \quad (5.27)$$

Repeating for (5.27), we have the desired unwrapped phase at harmonics

$$\phi_h^-(\omega_k) = \theta_h^-(\omega_k) = \theta_h^-(\omega_k) + n_0 \omega_k \quad (5.28)$$

Note that $PV [\Delta^2 \theta_h^-(\omega_k)]$ need not be computed by unwrapping and differencing, but rather directly through

$$PV [\Delta^2 \theta_h^-(\omega_k)] = \tan^{-1} \left[\frac{\sin[\Delta^2 \theta_h^-(\omega_k)]}{\cos[\Delta^2 \theta_h^-(\omega_k)]} \right] \quad (5.29)$$

since $\sin[\Delta^2 \theta_h^-(\omega_k)]$ and $\cos[\Delta^2 \theta_h^-(\omega_k)]$ can be expressed in terms of $\cos[\theta_h^-(\omega_k)]$ and $\sin[\theta_h^-(\omega_k)]$.

CHAPTER 6

APPLICATION OF PHASE TRACKING TO
SPEECH ANALYSIS-SYNTHESIS

In this chapter, we apply the techniques of phase tracking of the previous chapter in introducing a mixed phase estimate within two speech analysis-synthesis systems:

- (i) The homomorphic system proposed by Oppenheim[23]
- (ii) The spectral envelope system proposed by Paul[26]

With a simple model of the phase of the vocal tract frequency response, we first demonstrate that speechlike bandwidths are such that an unwrapped phase increment across successive harmonics generally does not exceed π , i.e., $|\Delta\theta_h(\omega_k)| < \pi$. Many of the results of the previous chapter are therefore applicable in the analysis of voiced speech when modeled over a short duration as a segment of a periodic waveform.

The concept of short-time homomorphic analysis of a periodic waveform is introduced and is shown to rely strongly on the nature of the window in the time domain. A Hamming window, with a pitch-adapted duration, and appropriately aligned, is applied to improve the phase estimate derived from homomorphic deconvolution. Such a windowing procedure, results in the additional property that our analysis-synthesis system is, potentially, an identity system with respect to a periodic waveform.

Finally, linear interpolation of section (5.1.1) is directly applied in deriving a phase estimate which is incorporated within the spectral envelope system. The original scheme based on magnitude only relies on determining harmonic locations by peak picking, and thus is compatible with our

interpolation procedure for phase tracking.

Informal listening tests indicate a small but perceptible improvement in "quality" within these systems when a mixed phase estimate replaces the minimum phase counterpart.

6.1 Bandwidth-Pitch Period Constraints

The maximum phase increment due to a single pole or zero approaches π as either comes close to the unit circle. The increment will not exceed π , so that in these simple cases $|\Delta\theta_h(\omega_k)| < \pi$. The geometry of a more complicated pole-zero pattern, however, might be such that $|\Delta\theta_h(\omega_k)| > \pi$. It is the purpose of this section to derive bounds on the bandwidth of poles and zeros of a z-transform for which this constraint is satisfied, and to show that typical speech bandwidths fall within these bounds.

Fig. 6.1 depicts two regions, A and B, of an elemental phase function, $\theta(\omega)$ which corresponds to a real maximum phase zero with Fourier transform $H(\omega) = 1 - a\exp[j\omega]$, $0 \leq a \leq 1$. Any complex pole or zero contributes a phase which is a shifted and/or negated version of that in Fig. 6.1. An arbitrary phase increment $\Delta\theta_h(\omega_k)$ can, therefore, be represented by a sum of increments derived from $\theta(\omega)$.

To obtain bounds on this sum, we consider the phase derivative of $\theta(\omega)$:

$$\dot{\theta}(\omega) = (a^2 - a\cos\omega)/(a^2 - 2a\cos\omega + 1) \quad (6.1)$$

The maximum of $\dot{\theta}(\omega)$ in region B occurs at $\omega = \pm \pi$:

$$\max \dot{\theta}(\omega) \Big|_B = \dot{\theta}(\omega) \Big|_{\omega=\pm\pi} = a/(a+1) \quad (6.2)$$

and the minimum in region A occurs at $\omega = 0$:

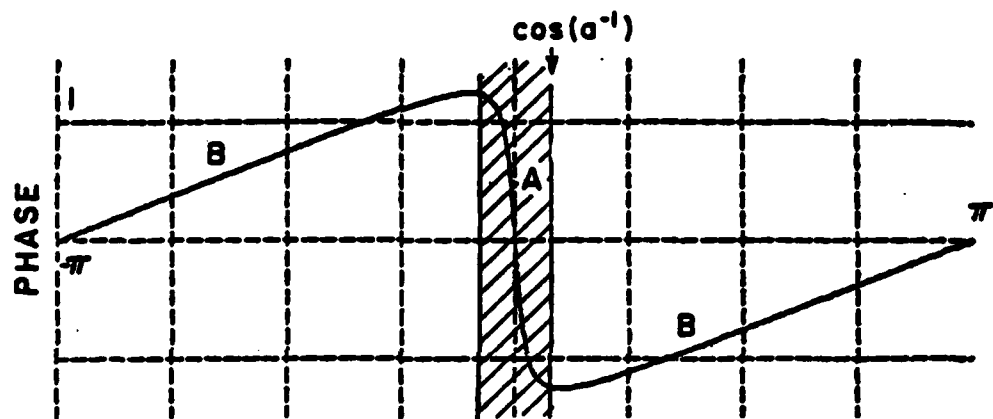


Fig. 6.1 Unwrapped phase of
 $1 - a\exp[j\omega]$

$$\min \theta(\omega) \Big|_A = \dot{\theta}(\omega) \Big|_{\omega=0} = a/(a-1) \quad (6.3)$$

For simplicity and mathematical tractability, we express a phase increment over the interval $[\omega_k, \omega_{k+1}]$ as a sum of increments from region A and from region B. A minimum phase pole or maximum phase zero contributes a phase of the form $\theta(\omega - \omega_0)$. A minimum phase zero contributes a phase of the form $-\theta(\omega - \omega_0)$ so that the contribution to a phase increment from region A or B may therefore be negated.

Since $\theta(\omega)$ is monotonic in region A or B, from (6.2) and (6.3) bounds on an increment $\Delta\theta(\omega_k)$ in the two regions are given by:

$$0 > \Delta\theta(\omega_k) \Big|_A \geq -2\tan^{-1} \left[\frac{a \sin(\pi/P)}{1-a \cos(\pi/P)} \right] > -\pi \quad (6.4)$$

and

$$0 < \Delta\theta(\omega_k) \Big|_B \leq \pi/P < \pi \quad (6.5)$$

We consider two cases which are clearly not exhaustive, but indicative of bandwidth constraints:

Case 1 (sparsely spaced poles and zeros):

When the poles and zeros of $H(z)$ are sparsely spaced, we shall assume that any interval of $\theta_h^-(\omega)$ consists of either the sum of contributions from only B regions, or the sum of contributions from all B regions and at most one A region.

When only B regions overlap, from (6.5) $|\theta_h^-(\omega_k)| < \pi$ because we assume the number of poles and zeros to be less than P. When there exists one A region, we assume the sum of contributions from B regions to be positive and less than π . The positivity constraint holds since for speechlike

spectra, the number of poles is greater than the number of zeros. Therefore, since $0 > \Delta\theta(\omega_k) \Big|_A > -\pi$, we have $|\Delta\theta_h^-(\omega_k)| < \pi$.

Case 2 (closely spaced poles and zeros):

Consider the overlap of two A regions (e.g., two closely spaced poles or a closely spaced pole and maximum phase zero) and an arbitrary number of B regions. Since we assume the net contribution of the B regions is positive, from (6.4) and (6.5), we write the following approximate constraint

$$|\Delta\theta_h^-(\omega_k)| < |2 \tan^{-1} \left[\frac{a \sin(\pi/P)}{1-a \cos(\pi/P)} \right] - K\pi/2P| < \pi \quad (6.6)$$

where $K\pi/2P$ is the "average" contribution from K B regions and

$$2 \tan^{-1} \left[\frac{a \sin(\pi/P)}{1-a \cos(\pi/P)} \right]$$

is the "average" contribution from two A regions, where we have assumed equal variable bandwidths.

With some algebraic manipulation our constraint can be expressed in terms of half power bandwidth α [5] by

$$\begin{aligned} \alpha &= \frac{F_s}{2\pi} \text{Ln} |a^{-1}| \\ &> \frac{F_s}{2\pi} \text{Ln} \left| \cos(\pi/P) + \sin(\pi/P) \tan \left[\left(\frac{\pi}{2} (1+K/2P) \right)^{-1} \right] \right| \end{aligned} \quad (6.7)$$

where F_s is the A/D sampling frequency.

Table 6.1 gives values of α for typical values of P and possible values of K corresponding to four or five real or complex pole pairs, and one or two real or complex zero pairs. Table 6.2 gives average measured bandwidths for the first three formants of vowel utterances for three male subjects

TABLE 6.1

Computed half-power bandwidths required for phase tracking

K P	1	2	3	4	5	6	7	8	9	10
4 msec	83	82	80	79	78	77	76	74	73	72
6 msec	58	57	56	56	55	54	54	53	53	52
8 msec	44	43	43	42	42	42	41	41	41	40
10 msec	36	35	35	34	34	34	34	33	33	33

TABLE 6.2

Mean half-power bandwidths (B1, B2, B3) in cycles per second for the first three resonances of vocal tract configurations. Three male subjects (S1, S2, S3) used eight vowel configurations with two glottal conditions (after House and Stevens[11]).

Band Width	Open Glottis				Closed Glottis			
	S1	S2	S3	Mean	S1	S2	S3	Mean
B1	79	75	67	73	56	50	56	54
B2	88	82	75	81	67	62	61	65
B3	97	98	83	94	76	66	64	70

with open and closed glottis[11]. Since during a vowel utterance the glottis is partly open and partly closed, the actual bandwidths probably lie somewhere between those measured for the open and closed glottis. It is clear that these bandwidths often fall within the derived lower bounds. It is conceivable, however, that spectra of high-pitched speakers with closely spaced formants may yield a phase increment outside of our desired constraint (i.e., $|\Delta\theta_h(\omega_k)| < \pi$).

6.2 Pre- and Post-Alignment

Alignment of $x(n) = h(n)*p(n)$ so that n_0 falls within the constraints of (5.16) will be referred to as pre-alignment. Under this constraint $|\Delta\theta_h(\omega_k)| < \pi$ so that windowing in the time domain or performing linear interpolation in the frequency domain tracks the desired unwrapped phase at harmonics. Since $n_0 = \theta_h(\pi)/\pi$, the linear phase term $n_0\omega$ can be eliminated by subtraction -- a procedure we shall refer to as post-alignment. This operation generates an estimate of $\theta_h(\omega)$ which equals the desired unwrapped phase at harmonics when (5.16) holds, and is an accurate approximation to the desired unwrapped phase between harmonics. Although pre-alignment may guarantee that (5.16) holds, the value of n_0 may not be consistent from frame-to-frame. Consequently, when n_0 is not removed "pitch jitter" can arise at frame boundaries after signal reconstruction, causing degradation in the synthesized speech. That is, the arbitrary displacement of the vocal tract impulse response estimate results in a random change in pitch from frame-to-frame.

In practice, however, estimation and removal of n_0 is complicated by deviations from the assumed harmonic structure of $X(\omega, n_0)$ in (5.6) in high-frequency regions. Two causes for such deviations are: (1) low-pass filter-

ing the speech waveform before A/D conversion, and (ii) a mixed-source excitation at the glottis which is harmonic in a low-frequency region and flat in a high-frequency region[16]. Therefore, deriving n_0 from $\theta_{\hat{h}}(\pi)$ is not recommended.

One simple way of representing high-frequency degradation in the case of low-pass filtering is through a model of $x(n)$ given by

$$x(n) = h(n)*g(n)*p(n) \quad (6.8)$$

where $g(n)$ is a low-pass filter. Ideally, $g(n)$ introduces a linear phase component within the phase of $H(\omega)$. The high-frequency energy of $g(n)$ is, however, quite low and so in practice, the unwrapped phase of $g(n)$ beyond its low-frequency cutoff is erratic due, for example, to quantization noise. Consequently, the value of $\theta_{\hat{h}}(\pi)$ is unpredictable and may have no relationship to the original linear phase of $H(\omega)$. Subtraction of $\omega\theta_{\hat{h}}(\pi)/\pi$, therefore, arbitrarily shifts the estimate of $h(n)$ and results in "pitch jitter" at frame boundaries. There exist then two causes of "pitch jitter": (i) inconsistent pre-alignment, and (ii) erroneous post-alignment.

For the purpose of waveform reconstruction, it is necessary to obtain only relative delays between successive impulse response estimates. An alternative method of post-alignment invokes the cross-correlation function of two successive estimates. Given that two successive estimates $\hat{h}_k(n)$ and $\hat{h}_{k+1}(n)$ are slowly varying except for a delay, their cross-correlation function is given roughly by

$$\begin{aligned} R(n) &= \hat{h}_k(n)*\hat{h}_{k+1}(-n) \\ &\approx [\tilde{h}(n)*\delta(n-n_k)*g(n)]*[\tilde{h}(-n)*\delta(-n-n_{k+1})*g(-n)] \\ &= R_{\tilde{h}}(n)*R_g(n)*\delta(n-(n_k-n_{k+1})) \end{aligned} \quad (6.9)$$

where $R_h(n)$ and $R_g(n)$ are the autocorrelation functions of $\tilde{h}(n)$, our desired response, and $g(n)$, respectively, and n_k and n_{k+1} are the delays in $h(n)$ due to the pre-alignment process. Therefore, the location of the peak in $R(n)$ is an estimate of $n_k - n_{k+1}$, and does not rely on the unwrapped phase at $\omega = \pi$. We can then perform accurate post-alignment by shifting the $(k+1)$ st impulse response estimate by $n_k - n_{k+1}$ points.

6.3 Homomorphic Speech Analysis-Synthesis

In this section we first review the minimum phase homomorphic speech analysis-synthesis system proposed by Oppenheim. With the results of Chapter 5 and sections 6.1 and 6.2 as a foundation, we then develop a high-quality homomorphic system which incorporates a mixed phase estimate.

6.3.1 Minimum Phase Analysis-Synthesis

The analyzer of the minimum phase system consists of Fourier transforming a short-time speech segment, computing the logmagnitude of its Fourier transform, and inverse transforming to generate the real cepstrum. Pitch information is obtained with a cepstral pitch detector[3] which utilizes cepstral peak-picking, and energy and zero-crossing measurements. Three and five point median smoothing[3] was applied to estimate one and two point isolated singularities due to pitch doubling and halving, or voiced/unvoiced errors. Some hand editing was performed at voiced/unvoiced transitions.

The minimum phase impulse response estimate is derived by multiplying the real cepstrum with a 3.2 msec. low-time gate of the form in (4.15). The result is then transformed, exponentiated, and inverse transformed.

The waveform is synthesized by explicitly convolving the impulse response estimate and excitation. During voicing, the excitation consists

of a train of unit impulses with spacing equal to the pitch period. During unvoiced intervals, a noiselike waveform of random polarity with spacing 1 msec. is used. Linear interpolation of the pitch period and impulse response was performed to avoid sudden changes in pitch and spectral information. Such interpolation leads to enhanced quality[23,28]. This system was simulated on a PDP/11-55 computer with floating point arithmetic. 6db/octave pre- and de-emphasis was used, and the input speech was low-pass filtered at 4.8 kHz.

The resulting synthetic speech is of high quality, natural sounding, and provides a reference point for the mixed phase simulations to follow.

6.3.2 Mixed Phase Analysis

The first attempt to introduce a mixed phase estimate within the homomorphic scheme through the complex cepstrum, as defined through (4.1) and (4.2) resulted in large sensitivity of the phase estimate to the position and duration of the time-domain window[31,35]. The system generated synthetic speech characterized by a "hoarse" quality. This sensitivity is clear in light of section 5.1.

The reason for the hoarseness is that neither pre- nor post-alignment was performed and the window duration was in general greater than four times the pitch period. Consequently, both phase degradation and pitch jitter are introduced within the reconstructed waveform. Furthermore, our results illustrate that smoothing the phase through the complex cepstrum is meaningful only when the envelope of the phase of $H(\omega, n_0)$ is preserved; that is, when phase tracking is guaranteed by appropriate windowing. The requirements on the window, however, are such that the windowed speech

waveform may not follow a convolutional model[31,35] -- a phenomenon which can be easily seen from the results of section 5.1. Therefore, pitch and impulse response information in the complex cepstrum are not necessarily additive. Empirical results indicate that for an arbitrary window the breakdown of the additivity assumption occurs severely with respect to phase, but far less so with respect to magnitude.

Applying a low-time gate to the complex cepstrum may be alternatively interpreted as a method of bandlimiting the complex logarithm, and thus smoothing the logmagnitude and phase. When the phase is not tracked at harmonics as, for example, when $x(n)$ does not fall within the constraints of (5.16) or the window is excessively long, large erroneous trends are introduced into the phase function as depicted in Figs. 5.2 and 5.4. In the former case, these trends are due largely to the 2π error accumulator function, $\theta_{2\pi}(\omega)$. It is reasonable to assume that smoothing $\theta_{2\pi}(\omega)$ can alter the principal value of the desired phase in high energy regions or at harmonics. Therefore, with respect to either $h(n)$ or a reconstruction of the original waveform $x(n)$, signal distortion due to phase distortion may be large. Such distortion has, in fact, been observed empirically in both the estimate of $h(n)$ [28,35] and the reconstructed waveform.

6.3.3 Short-Time Reconstruction

The notion of bandlimiting the complex logarithm is an interesting one since it implies an important alternative interpretation of the synthesis procedure. As the cepstral window approaches unity the original windowed waveform is preserved. Therefore, when the windowed waveform has primarily a low-time cepstral composition, gating its complex cepstrum maintains its

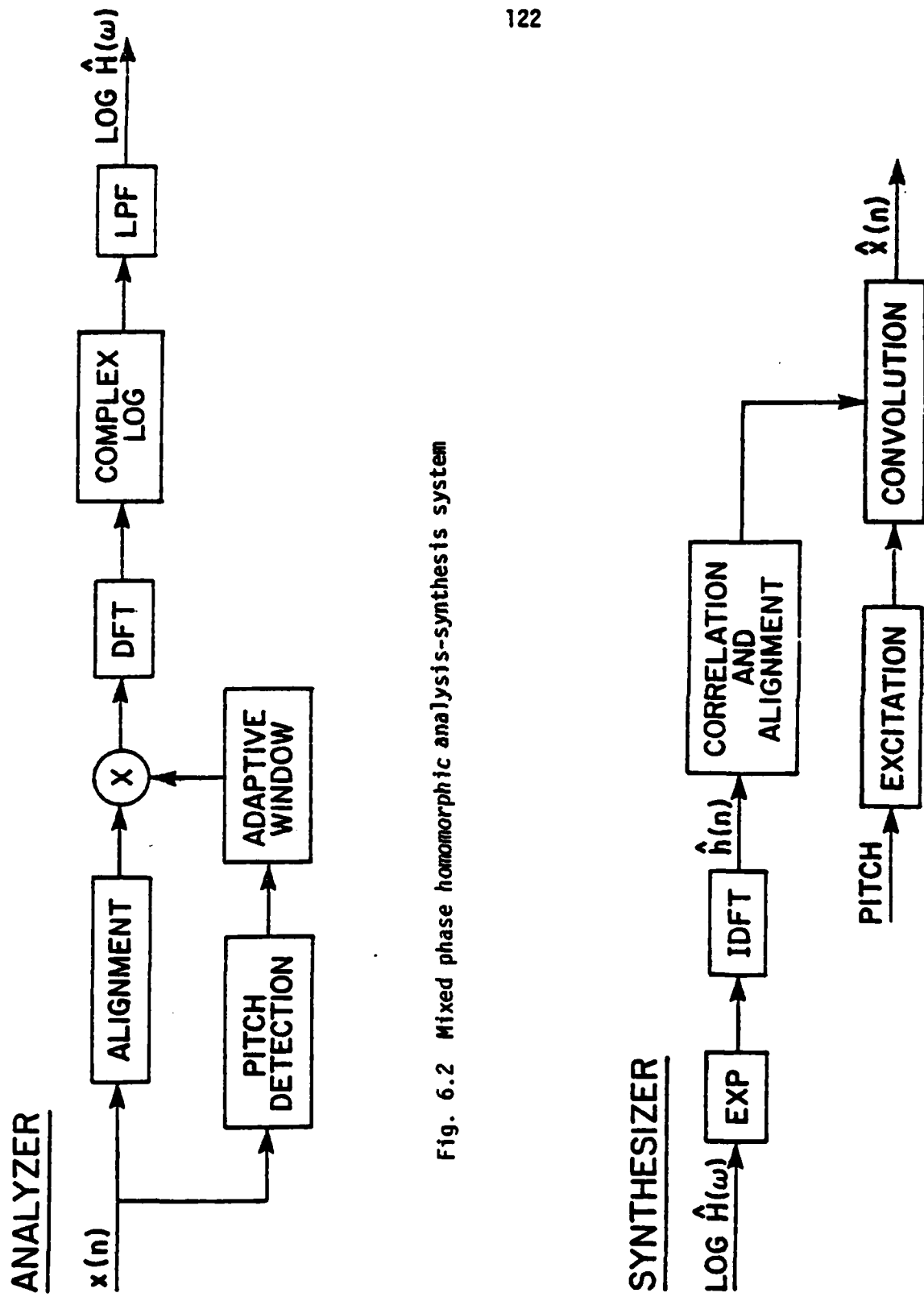
basic characteristics.

Furthermore, when a Hamming window is two pitch periods in extent, and positioned identically with respect to $h(n)$ in successive frames, the synthesis procedure reconstructs a periodic waveform exactly and thus acts as an identity system. This is because a Hamming window of two pitch periods in length has zeros at the spectral harmonics. Equivalently, a Hamming window when repeatedly added to itself delayed by a pitch period results in unity[28]. The identity system, however, does not rely on satisfying the alignment requirement of (5.16) but simply a consistent positioning of the window from frame-to-frame. The overall analysis-synthesis system is depicted in Fig. 6.2.

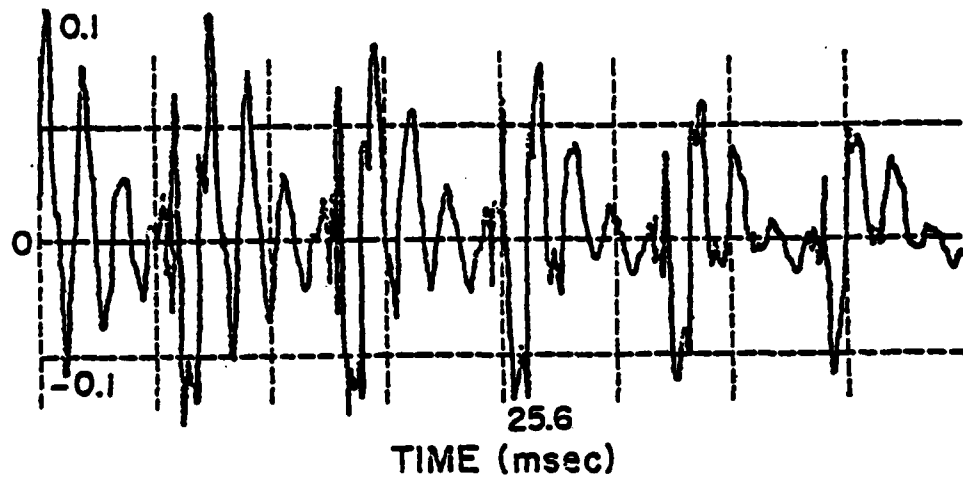
An example of the output of this system for a real speech input is illustrated in Fig. 6.3 for 51.2 msec. of data across three frame boundaries with a 20 msec. frame rate. The cepstral window in this particular case is of duration $.8 P$ (where P equals the pitch period), demonstrating the scheme's potential as an identity system. The pre-alignment process is heuristic, consisting of picking the maximum absolute value within the second pitch period from the current frame number and back tracking 10% of the pitch period. The synthesis includes also post-alignment by cross-correlation of section 6.3.

6.3.4 Informal Listening Tests

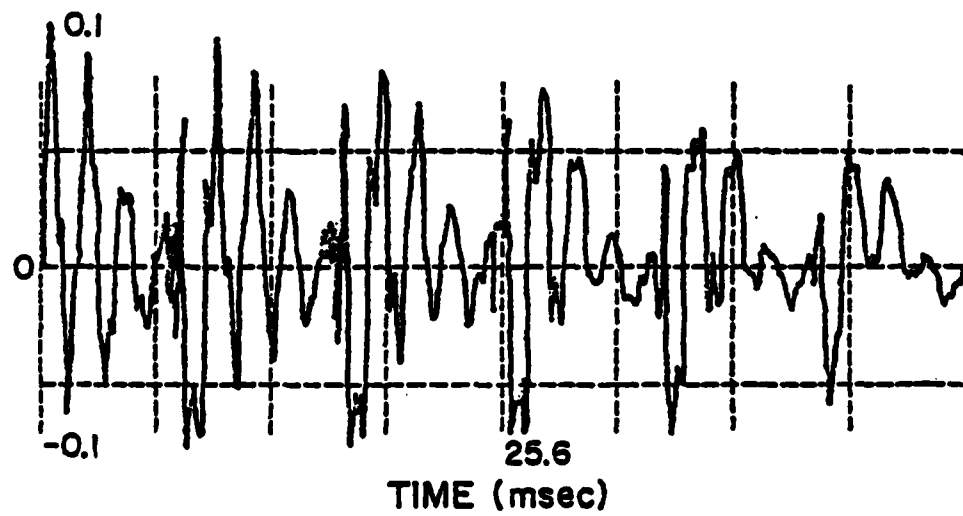
An analysis-synthesis system incorporating a mixed phase estimate was designed and evaluated. A Hamming window of two pitch periods in duration was applied in the analysis. Our requirements on window length, pre-alignment and post-alignment for periodic segments reduced the characteristic hoarseness of this system. Removal of any one of these requirements increased



123



(a)



(b)

Fig. 6.3 (a) Original speech segment
(b) Synthetic speech segment derived from the system in Fig. 6.2

hoarseness. A fixed 20 msec. time window was applied to unvoiced segments, and a symmetric rectangular low-time gate with cutoff 3.2 msec. was applied to the complex cepstrum.

When compared with its minimum phase counterpart, the system with mixed phase produced small, but audible, improvement in "quality". The two systems were identical except for the method of introducing phase and the elimination of post-alignment in the minimum phase version.

An informal A/B listening test was performed where each listener was asked to choose the synthetic speech passage closest to the original. Ten listeners and eight sentences with 5 male and 3 female speakers were used.

The system with mixed phase was judged roughly 45% of the time to be closer in quality to the original than its minimum phase counterpart, the minimum phase version 10% of the time was judged closer, and 45% of the time the two were indistinguishable. When preferred, the system with phase was often judged by experienced listeners, to reduce "buzziness" of the minimum phase reconstruction.

With an adaptive cepstral gate of .8P for voiced speech and a fixed 9 msec. cepstral gate for unvoiced speech, no significant differences were noted from use of a fixed 3.2 msec. gate. Consequently, it appears that any phase errors in the impulse response estimate may not be "noticed" because of the overlap-addition property of the Hamming window discussed in section 6.3.3, and since the windowed waveform has a low-time complex cepstrum.

We argued in Section 5.1.3 that the windowing procedure used here may not be suitable for spectral magnitude estimation. From Fig. 5.1 linear interpolation of complex harmonic samples possibly results in a null in the magnitude between harmonics when the phase increment is near π . When the pitch estimate is inaccurate, these nulls may in fact be sampled upon recon-

struction. This unfortunate situation suggests that the magnitude should perhaps be obtained in a different manner from that of the phase. The spectral envelope system to be discussed, in fact, applies this philosophy.

In spite of this possibility, however, we have found that the use of a Hamming window, of duration twice a pitch period, within a minimum phase system derived from the magnitude only, generates higher quality synthetic speech than a system where a longer window of duration 40 msec. is applied [28]. A longer window introduces other forms of degradation which have been investigated by the author in a study conducted in parallel with this thesis [28].

6.4 Spectral Envelope Speech Analysis-Synthesis

In this section we first briefly review the structure of the spectral envelope speech analysis-synthesis system proposed by Paul[26]. This system relies on a magnitude estimate only and is based on the harmonic samples of the Fourier transform of the windowed speech waveform. A phase estimate of the vocal tract system function is then introduced by the linear interpolation scheme of section 5.1.1, also based on harmonic samples. This phase estimation procedure is therefore easily appended to Paul's original system.

6.4.1 Minimum Phase Analysis-Synthesis

Conceptually, the minimum phase spectral envelope analysis-synthesis system is similar to the homomorphic configuration of section 6.3.1. The primary difference lies in the magnitude estimation procedure. An outline of this scheme is given as follows:

- (1) Apply a Hamming window to the speech waveform.
- (11) Find the location and value of the harmonic peaks of the magni-

tude (or logmagnitude) of the resulting Fourier transform by peak picking. A pitch adaptive peak picking algorithm is used.

(iii) Linearly interpolate the harmonic samples of the magnitude (or logmagnitude).

A minimum phase estimate is obtained by applying the causal sequence of the form in (4.13) to the real cepstrum.

6.4.2 Phase Envelope Estimation By Linear Interpolation

Step (ii) above gives the locations of harmonic peaks. We can use these locations to find the values of $H(\omega)$ at harmonics. Given these values we then directly apply the linear interpolation scheme of section 5.1.1 to obtain a phase estimate which preserves the desired phase envelope.

We must of course align the waveform so that condition (5.16) is satisfied. In particular, the heuristic pre-alignment process of section 6.3.3 is utilized. The precise duration of the window, however, is not crucial since we do not rely on the window itself to perform interpolation in the frequency domain. In our analysis, we choose a window length of twice the pitch period which approximately preserves the harmonic values and is short enough so that stationarity of the speech waveform approximately holds.

6.4.3 Short-Time Reconstruction

With our analysis scheme, the synthesizer of section 6.3.1 will in theory recover a periodic waveform exactly, and thus acts as an identity system for such waveforms. This is because the analysis preserves the harmonic values which completely characterizes a periodic waveform. Note

that when pre-alignment is such that condition (5.16) is not satisfied, but the alignment is consistent from frame to frame, the identity, nevertheless, holds.

6.4.4 Informal Listening Tests

When compared with its minimum phase counterpart, the system with mixed phase provided small, but audible, improvement in quality. The two systems were identical in both the analysis and synthesis stages except for the phase estimation procedure in the analysis and the elimination of post-alignment in the minimum phase version. In both schemes the magnitude estimate is derived by linear interpolation of samples of the magnitude of $H(\omega)$ at harmonics.

The informal A/B listening test of section 6.3.4 was performed. The system with mixed phase was judged roughly 53% of the time to be closer in quality to the original than its minimum phase counterpart, the minimum phase version 10% of the time was judged closer, and 37% of the time the two were indistinguishable.

CHAPTER 7

SUMMARY AND FUTURE RESEARCH

In this chapter we summarize the major results of the thesis and discuss a number of directions for future research.

7.1 Summary

In this dissertation we have considered techniques of phase estimation for the purpose of speech analysis and synthesis. Our methods fall roughly within the categories of direct and indirect approaches.

A number of indirect techniques generate the phase of a Fourier Transform from its magnitude and a priori knowledge of the desired sequence or phase. Both closed form and iterative solutions were developed for phase retrieval from magnitude under various constraints. An alternate indirect means of estimating phase from a magnitude relies on transformation of the speech waveform to create a minimum phase impulse response whose magnitude is estimated. An inverse operation provides an estimate of the original phase function.

Direct approaches do not require an estimate of a magnitude function, but require samples of the desired system function or partial knowledge of its phase derived from the speech waveform. In particular, phase estimates based on harmonic samples of the desired system function were incorporated within two speech analysis-synthesis systems with the result of high-quality synthetic speech.

We have also developed an iterative technique for magnitude retrieval from phase. This algorithm provides an alternative to the Hilbert transform for obtaining the magnitude from the phase of a minimum phase sequence. The technique does not require an unwrapped phase, but simply its principal value. The iteration also provides a means for magnitude recovery when

only the phase of a finite length mixed phase sequence is given. Furthermore, the iterative minimum phase signal reconstruction serves as the major component within a new phase unwrapping algorithm which does not require modulo 2π considerations.

This study represents only an initial step in developing techniques of phase estimation in the realm of speech analysis and perhaps, equally important, in other areas such as image restoration where an accurate phase estimate may be particularly significant.

Below we outline a number of possible directions for future research.

7.2 Suggestions For Future Research

Sensitivity Analysis

The methods of phase and magnitude retrieval in sections 3.2 and 4.2 rely on exact knowledge of a magnitude or phase function, or values of a sequence. Sensitivity of either the closed form or iterative solutions to noise or other degradation is not completely understood. A number of observations, however, can be made. When causality and phase continuity constraints are imposed, clearly the uniqueness argument of section 4.2 still holds. That is, given a degraded phase there exists one degraded magnitude which corresponds to a minimum phase sequence. With a similar argument, we conclude that under a finite length constraint there exists one magnitude function for a given distorted phase.

In fact, with phase distortion, our iterative algorithm to recover magnitude from phase converges to the unique predicted magnitude. The magnitude function, however, may differ significantly from the original distortionless magnitude. For example, in minimum phase reconstruction by iteration, a maximum phase zero lying near the unit circle may become

a minimum phase zero in the presence of noise rather than a minimum phase pole as determined by (2.28). This phenomenon, however, is not a function of the iterative procedure, but of the constraints imposed.

The sensitivity of the dual problem of recovery of phase from magnitude is more difficult to characterize. An ambiguity arises since addition of noise may cause the breakdown of our assumed rational models and conditions of section 3.2.

Another sensitivity issue involves the effect of noise or other degradation on the technique of section 4.1.4 where a phase estimate is obtained by transformation of a quasi-periodic waveform. We illustrated one example of this approach to improve the phase estimate over a direct procedure in the presence of noise. However, the mapping of magnitude degradation through an inverse transformation to phase degradation is not understood.

Convergence Issues

The iterative algorithm for phase and magnitude retrieval were found in practice to generate converging solutions sometimes slowly (e.g., after several hundred iterations) and sometimes quickly (e.g., after a few iterations). Consequently, determining rates of convergence in terms of spectral structure, and methods for quickening convergence are useful areas of research.

Another question involves the existence of rigorous proofs demonstrating unique convergence of $\theta_k(\omega)$ and $M_k(\omega)$ derived from the iterative algorithms.

Alternative Constraints for Phase Retrieval

Given a magnitude function, we might consider constraints for unambiguous phase retrieval where specific nonzero values of $h(n)$ are not required.

Consider, for example, the minimum and maximum values of $h(n)$ or $\theta_h(\omega)$. A question which naturally follows is whether such constraints can be imposed within an iterative scheme.

Partial Magnitude - Partial Phase

In sections 3.1 and 4.2.1 samples of the magnitude or phase along with certain nonzero values of $h(n)$ were sufficient for phase or magnitude retrieval. One obvious extension of this result concerns signal reconstruction from samples of both magnitude and phase. We might consider possible ways to distribute such samples to unambiguously recover the entire magnitude and phase functions. In the context of coding, certain sampling distributions may cause little degradation of the reconstructed signal.

"Optimal" Transformations

In section 4.3 we have only touched upon the use of transformations in phase estimation. What is needed is a method for designing "optimal" transformations which create minimum phase sequences with specific desirable properties. For example, in the context of homomorphic deconvolution the real cepstrum of the modified sequence should be as low time, as possible, but not susceptible to quantization noise as occurs with exponential weighting. These properties are also useful in coding samples of the log-magnitude function. In the context of signal enhancement, the transformation should be such that the resulting magnitude function is suitable to estimation by a specific noise reduction scheme such as spectral subtraction.

Phase Modeling

Suppose an accurate estimate of the unwrapped phase is available. If this phase as well as magnitude function is to be efficiently transmitted

in, for example, a speech communications system, a small parameter set must be derived which represents the entire phase function. When the unwrapped phase is bandlimited, samples of the phase are adequate. Alternatively, we may consider ways to directly model the phase, or phase derivative, by a rational function.

A Phase-Only Vocoder

We saw in section 6.3.3 that the homomorphic analysis-synthesis scheme is an identity system with respect to a periodic waveform when an adaptive Hamming window is applied. The quality of the overall system does not significantly differ when a low-time or unity cepstral gate is applied since the complex logarithm is nearly bandlimited. Aligning the Hamming window and adapting its duration to the pitch period yields logmagnitude and unwrapped phase functions whose slowly varying components approximately represent the windowed waveform.

Since the windowed waveform is of finite extent, from section 4.2 the phase is sufficient for complete waveform characterization. Therefore, coding samples or a small set of model parameters of the bandlimited phase function may be adequate for signal reconstruction.

REFERENCES

1. B. S. Atal and N. David, "On Synthesizing Natural-Sounding Speech by Linear Prediction", IEEE Conf. Acoust., Speech, Signal Processing, Washington, D.C., pp. 44-47, April 2-4, 1979.
2. B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", J. Acoust. Soc. Amer., Vol. 50, No. 2, pp. 637-655, 1971.
3. M. J. Cheng, "A Comparative Performance Study of Several Pitch Detection Algorithms", S.M. thesis, Dept. Elec. Eng. and Computer Science, Massachusetts Institute of Technology, Cambridge, Mass., June 1975.
4. J. J. D'Azzo and C. H. Houpis, Feedback Control System Analysis and Synthesis, New York, N.Y.: McGraw-Hill, 1966.
5. J.L. Flanagan, Speech Analysis Synthesis and Perception, 2nd Edition, New York: Springer-Verlag, 1972.
6. R. W. Gerchberg, "Super-Resolution Through Error Energy Reduction", Optica Acta, Vol. 21, No. 9, pp. 709-720, 1974.
7. R.W. Gerchberg and W. O. Saxton, "A Practical Algorithm for The Determination of Phase From Image and Diffraction Plane Pictures", Optik, Vol. 35, No. 2, pp. 237-246, November 1972.
8. B. Gold, "Channel Vocoder Quality Improvements", J. Acoust. Soc. Amer., Vol. 65, Sup. No. 1, pp. 5100, Spring 1979.
9. E. A. Guillemin, The Mathematics of Circuit Analysis, M.I.T. Press, Massachusetts Institute of Technology, Cambridge, Mass. 1969.
10. M. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal Reconstruction from Phase and Magnitude", submitted to IEEE Trans. Acoust., Speech, and Signal Processing.
11. A. S. House, K. N. Stevens, "Estimation of Formant Band Widths From Measurements of Transient Response of the Vocal Tract", Journal of Speech and Hearing Research, Vol. 1, No. 4, pp. 309-315, Dec. 1958.
12. G. E. Kopec, A. V. Oppenheim, J. M. Tribolet, "Speech Analysis by Homomorphic Prediction", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-25, No. 1, pp. 40-49, February 1977.
13. J. S. Lim, "Enhancement and Bandwidth Compression of Noisy Speech by Estimation of Speech and its Model Parameters", Sc.D. Thesis, Dept. Elec. Eng. Comput. Sci., Massachusetts Inst. Technology, Cambridge, MA, 1978.

14. J. S. Lim, Private Communication, Lincoln Laboratory, Massachusetts Inst. Technology, 1979, Cambridge, MA.
15. J. Makoul, "Linear Prediction: A Tutorial Review", Proceedings of the IEEE, Vol. 63, No. 4, pp. 561-580, April 1975.
16. J. Makoul, R. Viswanathan, R. Schwartz, and A.W.F. Huggins, "A Mixed-Source Model for Speech Compression and Synthesis", IEEE Conf. Acoust., Speech, Signal Processing, Tulsa Oklahoma, April 10-12, 1978, pp. 163-167.
17. R. C. Mathes and R. L. Miller, "Phase Effects in Monaural Perception", J. Acoust. Soc. Amer., Vol. 19, No. 5, pp. 780-797, Sept. 1947.
18. J. H. McClellan, "Notes For Advanced Topics in Digital Signal Processing", Summer Session, Massachusetts Inst. Technology, 1977.
19. J. H. McClellan and T. W. Parks, "Eigenvalue and Eigenvector Decomposition of The Discrete Fourier Transform", IEEE Trans. Audio. and Electroacoust., Vol. AU-20, No. 1, pp. 66-74, March 1972.
20. R. M. Mersereau and R. W. Schafer, "Comparative Study of Iterative Deconvolution Algorithms", IEEE Conf. Acoust., Speech, Signal Processing, Tulsa, OK, April 10-12, 1978.
21. H. Nawab, T. F. Quatieri, Private Communication, Massachusetts Inst. Technology, Cambridge, MA, 1979.
22. A. V. Oppenheim and R. W. Schafer, Digital Signal Processing, Englewood Cliffs, N.J.: Prentice-Hall, 1975.
23. A. V. Oppenheim, "Speech Analysis-Synthesis System Based on Homomorphic Filtering", J. Acoust. Soc. Amer., Vol. 45, pp. 458-465, Feb. 1969.
24. A. V. Oppenheim and R. W. Schafer, "Homomorphic Analysis of Speech", IEEE Trans. Audio Electroacoust., Vol. AU-16, pp. 221-228, June 1968.
25. A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation", IEEE Trans. on Circuits and Systems, Vol. CAS-22, No. 9, pp. 735-742, Sept. 1975.
26. D. Paul, "A Robust Vocoder with Pitch-Adaptive Spectral Envelope Estimation and an Integrated Maximum-Likelihood Pitch Estimator", IEEE Conf. Acoust. Speech, Signal Processing, Washington, D.C., pp. 64-68, April 2-4, 1979.
27. T. F. Quatieri, "A Mixed Phase Homomorphic Vocoder", IEEE Conf. Acoust., Speech, Signal Processing, Washington, D.C., pp. 56-59, April 2-4, 1979.

28. T. F. Quatieri, "Minimum and Mixed Phase Speech Analysis-Synthesis by Adaptive Homomorphic Deconvolution", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, No. 4, pp. 328-335, Aug. 1979.
29. J. R. Ragazzini and G. F. Franklin, Sampled-Data Control Systems, New York, McGraw-Hill, 1958.
30. M. Rosenlicht, Introduction to Analysis, Glenview, Ill.: Scott, Foresman, 1970.
31. A. Ruiz, "Windowing Effects in Short Time Homomorphic Deconvolution", S.M. thesis, Dept. Elec. Eng. Comput. Sci., Massachusetts Inst. Technology, Cambridge, MA, 1978, unpublished.
32. R. W. Schafer, "Echo Removal by Discrete Generalized Linear Filtering", Ph.D. thesis, Dept. Elec. Eng., Massachusetts Inst. Technology, Cambridge, MA, 1969.
33. J. M. Tribolet, "Seismic Applications of Homomorphic Signal Processing", Sc.D. thesis, Dept. Elec. Eng. Comput. Sci., Massachusetts Inst. Technology, Cambridge, MA, 1977.
34. J. M. Tribolet, "A New Phase Unwrapping Algorithm", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-25, pp. 170-177, April 1977.
35. J. M. Tribolet, T. F. Quatieri, and A. V. Oppenheim, "Short-Time Homomorphic Analysis", IEEE Conf. Acoust., Speech, Signal Processing, Hartford, CT, May 9-11, 1977.

BIOGRAPHICAL NOTE

Thomas F. Quatieri, Jr., was born in Somerville, Massachusetts, on January 31, 1952. He received the B.S. degree summa cum laude from Tufts University, Medford, Massachusetts, in 1973, the S.M. degree in 1975, and the E.E. degree in 1977, from the Massachusetts Institute of Technology (M.I.T.), Cambridge, Massachusetts, all in electrical engineering.

From 1973 to 1975, he was a Teaching Assistant in the Department of Electrical Engineering and Computer Science at M.I.T. His research for the Master's degree involved the design of two-dimensional digital filters. He was a Research Assistant from 1975 to 1979, also at M.I.T., in the Research Laboratory of Electronics. During this time his studies and publications involved digital signal processing and its application to speech. His current interests encompass image, speech, seismic, and biomedical signal processing.

Dr. Quatieri is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.

Distribution List

Director Advanced Research Project Agency 1400 Wilson Boulevard Arlington, Virginia 22209 Attn: Program Management	DODAAD Code HX 1241	(2)
Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217 Attn: Code 437	DODAAD Code N00014	(3)
ONR Branch Office Bldg. 114, Sec. D 666 Summer Street Boston, Massachusetts 02210	DODAAD Code N 62879	(1)
Mr. John Mahoney Administrative Contracting Officer E19-628 Massachusetts Institute of Technology Cambridge, MA 02139	DODAAD Code N 66017	(1)
Director Naval Research Laboratory Washington, D.C. 20375 Attn: Code 2627	<i>3 Brit 2 Canada 1 DTIC</i> DODAAD Code N 00173	(6)
Dr. Robert Kahn Program Manager Information Processing Techniques 1400 Wilson Boulevard Arlington, Virginia 22209		(1)
Capt. Richard L. Martin, USN Commanding Officer USS Francis Marion (LPA-249) FPO New York 09501		(1)
Director, National Security Agency ATTN: R54, Mr. Page Fort G.G. Meade, MD 20755		(1)

DATE
FILMED
9-8