# "Signal Estimation from Short-time Spectral Magnitude"

Syed Hamid Nawab

TECHNICAL REPORT 494

May 1982

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

This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grant ECS80-07102

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 |
| 4. TITLE *(and Subtitle)*<br>SIGNAL ESTIMATION FROM SHORT-TIME SPECTRAL MAGNITUDE | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Report<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Syed Hamid Nawab | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-81-K-0742 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Research Laboratory of Electronics<br>Massachusetts Institute of Technology<br>Cambridge, MA 02139 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>NR-049-506 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Advanced Research Projects Agency<br>1400 Wilson Boulevard<br>Arlington, Virginia 22217 | | 12. REPORT DATE<br>May 1982<br>13. NUMBER OF PAGES<br>91 |
| 14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*<br>Office of Naval Research<br>Mathematical and Information Sciences Division, 800 North Quincy Street<br>Arlington, Virginia 22217 | | 15. SECURITY CLASS. *(of this report)*<br>Unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

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)*

spectrum
magnitude
time-scale modification
noise reduction
multidimensional

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

see other side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. ABSTRACT

This thesis shows that in many practical situations the processing of a discrete-time signal can be accomplished using only the magnitude of its short-time spectrum. Mild restrictions on the signal and on the analysis window of the short-time spectrum are shown to be sufficient for unique signal representation with the short-time spectral magnitude. Furthermore, various algorithms are developed which reconstruct the signal from appropriate samples of the short-time spectral magnitude. Some of these algorithms are designed to obtain signal estimates from the processed short-time spectral magnitude, which generally does not have a valid short-time structure. These algorithms are successfully applied to the time-scale modification and noise reduction problems in speech processing. However, the results presented here have similar potential for other application areas, including those with multidimensional signals.

# Signal Estimation From Short-Time Spectral Magnitude

*by*

*Syed Hamid Nawab*

*Submitted to the Department of Electrical Engineering and Computer Science, on May 1982, in partial fulfillment of the requirements for the degree of Doctor of Philosophy*

## *ABSTRACT*

This thesis shows that in many practical situations the processing of a discrete-time signal can be accomplished using only the magnitude of its short-time spectrum. Mild restrictions on the signal and on the analysis window of the short-time spectrum are shown to be sufficient for unique signal representation with the short-time spectral magnitude. Furthermore, various algorithms are developed which reconstruct the signal from appropriate samples of the short-time spectral magnitude. Some of these algorithms are designed to obtain signal estimates from the processed short-time spectral magnitude, which generally does not have a valid short-time structure. These algorithms are successfully applied to the time-scale modification and noise reduction problems in speech processing. However, the results presented here have similar potential for other application areas, including those with multidimensional signals.

Thesis Supervisor: Professor Alan V. Oppenheim

Title: Professor of Electrical Engineering

## ACKNOWLEDGEMENTS

My association with Professor Alan V. Oppenheim, starting with an undergraduate project some eight years ago, has been a very valuable educational experience. Besides the intellectual benefits of numerous discussions with him that shaped the direction of this research, I found his enthusiastic attitude and candid advice on various matters to be extremely helpful.

Prof. Jae S. Lim and Dr. Thomas F. Quatieri collaborated at various points in this research. Their invaluable contributions are deeply appreciated.

Group 53 at Lincoln Laboratory, MIT, sponsored the Research Assitantship I recieved during my doctoral program. I am very greatful for their generous support.

I would also like to acknowledge the support of my parents throughout my education. Their pride in almost whatever I do has always been a source of pleasure for me.

Finally, I would like to thank my wife, Martha. She is the one who had to suffer the torturous role of a doctoral student's spouse, which she did splendidly. This not only required her patience and understanding, but also her aid in drafting many of the figures in this thesis.

TABLE OF CONTENTS

# CHAPTER ONE: SIGNAL PROCESSING USING ONLY SHORT-TIME SPECTRAL MAGNITUDE

The time-invariance of spectral processing [1] is a disadvantage in several applications, particularly those involving speech and images. For example, a speech waveform consists of voiced and unvoiced sections [2]. The voiced sections have a periodic structure, whereas the unvoiced sections consist mainly of wideband random noise. The processing requirements of these two types of sections are often quite different. In voiced sections, for example, it is important to preserve the periodicity but no such restriction applies to unvoiced sections. On the other hand, in unvoiced sections it is often essential to preserve the wideband random noise characteristic. Even within the various voiced or unvoiced sections, the signal properties tend to change. For example, within voiced sections, the length as well as the shape of each period is generally changing as a function of time. In fact, speech characteristics such as periodicity are generally assumed to be constant over only short durations on the order of 20 milliseconds [2]. In many cases, therefore, it is inadvisable to apply time-invariant processing to speech over intervals much greater than 20 milliseconds.

To achieve a degree of time dependence in the processing of signals such as speech, spectral processing is often applied independently to various short-time sections of a signal. This type of processing is usually based on the *short−time spectrum* [3]. In section 1.1 of this chapter, we present the definition of the short-time spectrum for discrete-time signals. The magnitude and phase of the short-time spectrum of a signal are usually both required in various signal processing applications. However, as we shall see in section 1.2, there are some applications where it is desirable to accomplish the processing with only the magnitude of the short-time spectrum. This has previously not been possible because of the lack of any practically useful results on the relationship between the short-time spectral magnitude and the corresponding signal. In particular, it is important to develop results on signal reconstruction from the magnitude of the short-time spectrum. Furthermore, since a processed short-time spectral magnitude may not necessarily

correspond to any signal, we would like to be able to obtain reasonable signal estimates in such cases. This thesis presents a number of important results on these problems that make possible the practical implementation of signal processing using only the magnitude of the short-time spectrum. Some previous investigations on this subject are described in section 1.3. Finally, in section 1.4, we outline the major results of this thesis.

### 1.1 Short-Time Spectrum

The short-time spectrum has been developed for continuous as well as discrete-time signals. Excellent references on the subject include the work of C. Weinstein [3], J. Allen [4], and M. Portnoff [5]. In this thesis, we are interested in discrete-time signal processing with the short-time spectrum. For a discrete-time signal $x(n)$, the short-time spectrum is a function of time as well as frequency and it is mathematically expressed as

$$X_w(nL,\omega) = \sum_{m=-\infty}^{\infty} x(m)w(nL-m)e^{-j\omega m} \tag{1.1}$$

where the subscript $w$ in $X_w(nL,\omega)$ denotes the analysis window, $w(n)$. The parameter $L$ is an integer which denotes the separation in time between adjacent short-time sections. This parameter is independent of time and is selected so as to ensure a degree of time overlap between adjacent short-time sections. For a fixed value of $n$, the short-time spectrum $X_w(nL,\omega)$ defined in (1.1) represents the Fourier transform with respect to $m$ of the short-time section $f_n(m)=x(m)w(nL-m)$. The *sliding window* interpretation [5] views $X_w(nL,\omega)$ as being generated by shifting the time-reversed analysis window across the signal. After each shift of $L$ samples, the window is multiplied with the signal and the Fourier transform is applied to the product. There are other interpretations of the short-time spectrum, including a well known filter bank interpretation [2]. However, for the purposes of this thesis, we find the sliding window interpretation to be the most appropriate.

For most signals and analysis windows, the short-time sections $f_n(m)$ generally do not have any symmetry with respect to the origin. Consequently, the short-time spectrum is generally a

complex function. In many short-time spectral processing applications both the magnitude and the phase of the short-time spectrum are used. As illustrated in the next section, however, it is important to determine if short-time spectral processing can be accomplished using only the magnitude of the short-time spectrum.

### 1.2 Applications For Magnitude Only Processing

In this section, we consider two important applications that illustrate the importance of developing practical signal processing techniques which use only the magnitude of the short-time spectrum. Specifically, we consider the problems of noise reduction and time-scale modification. These problems are stated mostly in the context of speech processing. However, it will be clear from the discussion that the same concepts also apply to other applications.

We first consider the problem of noise reduction. Suppose that a discrete-time signal $x(n)$ is the sum of a desired signal $s(n)$ and a noise signal $e(n)$. The signal $x(n)$ may, for example, represent samples of a noisy speech recording. If $e(n)$ originates from a random process that can be appropriately modelled as being stationary, there are some classical spectral processing methods for noise reduction. These include Wiener filtering, power spectrum filtering, and spectral subtraction. A comprehensive survey of such processing is contained in a paper by Lim and Oppenheim on noise reduction for speech signals [6]. These noise reduction procedures have the property that they process only the magnitude of the signal spectrum; the spectral phase of the noisy signal $x(n)$ is retained in the processed signal. For applications such as speech processing, these techniques perform relatively better when applied to the short-time spectrum rather than the (long-time) spectrum of $x(n)$. This way, each short-time section can be filtered according to its own spectral characteristics. Of course, only the spectral magnitudes of the short-time sections are affected by the processing. A problem of interest is whether an estimate of the short-time spectral phase can be obtained from the processed magnitude of the short-time spectrum. This is equivalent to estimating the processed signal from the short-time spectral magnitude alone.

Thus, in addition to obtaining a processed short-time spectral phase estimate, such a technique would have the property of not requiring any spectral phase information on the noisy signal.

Time-scale modification of signals is another area where short-time spectral processing plays an important role. The basic problem is to compress or expand the signal without changing its short-time spectral characteristics. In other words, the rate of change of the spectral characteristics is to be modified without significantly changing the principal frequency locations of spectral energy within the various short-time sections. In speech, such processing corresponds to a change in the apparent rate of articulation without any appreciable degradation of perceptual quality. M. Portnoff [7] has developed a time-scale modification technique based on the short-time spectrum. The technique applies a linear time-scaling to the short-time spectrum and then divides an estimate of the unwrapped phase [1] of the short-time spectrum by a factor proportional to the desired rate of time compression or expansion. Finally, the processed short-time spectrum is used to synthesize an estimate of the time-scale modified signal. Throughout this technique, both the magnitude and phase of the short-time spectrum are used. However, if signal estimation could be done directly from the processed short-time spectral magnitude, the phase processing would be avoided. Thus, in this case, a major incentive for developing techniques for signal estimation from short-time spectral magnitude is to avoid the computational expense associated with phase processing.

### 1.3 Previous Investigations

The magnitude of the short-time spectrum was the subject of investigations even before the short-time spectrum itself. In particular, researchers were motivated to study the short-time spectral magnitude because it was physically easier to estimate for signals such as speech. The first formal definition of the short-time spectral magnitude was introduced by R. Fano [8] in his studies on speech analysis. Investigations by R. Fano, M. Schroeder and B. Atal [9], and A. Kharkevich [10] were responsible for developing many aspects of short-time spectral analysis.

C. Weinstein [3] formally showed that continuous as well as discrete-time signals can be uniquely determined from the short-time spectrum to within a scale factor. Other investigators such as J. Allen [4] and M. Portnoff [5] have further refined the results obtained by C. Weinstein. For example, several different procedures have been established for signal reconstruction from the short-time spectrum. Portnoff has recently introduced an elegant approach to signal reconstruction from short-time spectrum. This approach includes previously established reconstruction procedures as special cases. As a consequence of all these studies, the short-time spectrum has become a very useful signal representation for various signal processing purposes.

The question of unique signal representation with the magnitude of the short-time spectrum has remained mostly unresolved. A study by Weinstein [3] showed the uniqueness of the short-time spectral magnitude only for a very restricted class of signals and analysis windows. In particular, his approach was based on the property that minimum phase signals are uniquely specified by their spectral magnitude. He observed that if each short-time section were minimum phase, we could uniquely reconstruct the short-time sections from their spectral magnitudes. If there is sufficient overlap between short-time sections, all the samples of the original signal may then be obtained by dividing out the analysis window from the various short-time sections.

More recently, an alternative approach to unique signal representation with short-time spectral magnitude was used by R. Altes [11]. This approach places no restriction on the signal to be represented. However, the analysis window is required to satisfy a condition which in practice means that the analysis window has to be longer than the signal. These results were obtained using a relationship between the short-time spectral magnitude and an ambiguity function, which for a discrete-time signal $x(n)$ is defined as

$$A_x(n,\omega) = \sum_{m=-\infty}^{\infty} x(m)x(n-m)e^{j\omega m} \tag{1.2}$$

The results derived by R. Altes show that if the analysis window $w(n)$ is such that $A_w(n,\omega)\neq 0$ for any pair $n,\omega$, then the signal $x(n)$ can be uniquely determined up to a sign factor from the magnitude of $X_w(n,\omega)$. Furthermore, if $x(n)$ is restricted to be a finite-length signal, then the

requirement on $w(n)$ is that $A_w(n,\omega)$ must not be zero for values of $n$ for which $A_x(n,\omega)$ is nonzero. From (1.2), it can be easily observed that the time duration of the ambiguity function is proportional to the time duration of the signal it represents. Hence, this approach gives conditions that are sufficient for unique signal representation with short-time spectral magnitude only for cases where the analysis window is longer than the signal being represented.

In short-time spectral processing we are generally interested in analysis windows whose lengths are much shorter than the signal to be processed. In this thesis, we present results which show that the uniqueness of the short-time spectral magnitude for signal representation can also be extended to such cases.

### 1.4 Outline of Thesis

In this thesis, we show that in many practical situations a discrete-time signal is uniquely represented by its short-time spectral magnitude. The key assumption in these results is that the analysis window is a known finite-length sequence. In such cases, it is seen that if there is sufficient overlap between short-time sections, the problem of determining a signal from its short-time spectral magnitude requires certain results on the extrapolation of finite-length signals from (long-time) spectral magnitude. Such results are derived in chapter 2 of this thesis and then used in chapter 3 for developing conditions under which the short-time spectral magnitude is a unique signal representation.

For practical applications, it is necessary to obtain algorithms for signal reconstruction from samples of the short-time spectral magnitude. In chapter 4, we present a number of such algorithms with various implementation properties. These algorithms have been successfully implemented for the reconstruction of speech signals. They also offer similar potential for other application areas, including those with multidimensional signals.

In general, processing the short-time spectral magnitude results in a function which does not correspond to the short-time spectral magnitude of any signal. An important contribution of this

thesis is the development of signal reconstruction algorithms that yield reasonable signal estimates from the processed short-time spectral magnitude. Some general issues involved in applying the signal reconstruction algorithms to the processed short-time spectral magnitude are discussed in chapter 5.

The final chapters of this thesis consider the application of the various ideas in chapters 2 to 5 to the problems of time-scale modification and noise reduction, particularly in the context of speech processing. For time-scale modification, we have implemented a procedure whose performance is comparable to previous systems based on both the magnitude and the phase of the short-time spectrum. In contrast, however, our technique has significantly less computational complexity.Furthermore, we have also implemented a short-time spectral processing technique for noise reduction that estimates the processed short-time spectral phase from the processed short-time spectral magnitude. The performance thus obtained appears comparable to that obtained with techniques that require both the magnitude and phase of the short-time spectrum of the noisy signal.

# CHAPTER TWO: SIGNAL EXTRAPOLATION FROM SPECTRAL MAGNITUDE

In this chapter, we derive theorems on the extrapolation of discrete-time signals from their (long-time) spectral magnitude. Besides being important theoretical results in their own right, these theorems play a central role in deriving conditions under which the short-time spectral magnitude is a unique signal representation.

In discrete-time signal extrapolation, a signal $x(n)$ known up to $n=n'$ is extended for $n>n'$, maintaining consistency with all *a−priori* knowledge on $x(n)$. The signals considered in this chapter are known to be zero outside an interval $0 \leq n \leq N$ for some positive integer $N$. The particular location of this interval on the n-axis is for notational convenience only ; none of the results derived in this chapter are affected by any shift in this location. Given $x(n)$ for $0 \leq n \leq M$ where $M < N$, we wish to extrapolate $x(n)$ up to $n=N$, using the spectral magnitude, $|X(\omega)|$, where

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)\, e^{-j\omega n} \tag{2.1}$$

Furthermore, we are interested in determining conditions under which the extrapolation is unique. Section 2.2 derives two theorems on such extrapolation for the case in which only the sample $x(N)$ is unknown. This is referred to as single sample extrapolation. Section 2.3 presents a theorem for the more general case, where several samples of $x(n)$ are extrapolated. These theorems are used extensively in Chapter 3 for deriving conditions under which the short-time spectral magnitude is a unique signal representation. The relationship between the theorems in this chapter and the uniqueness of the short-time spectral magnitude is discussed in the following section.

**2.1 Relation to Short-Time Spectral Magnitude**

For a signal $x(n)$ and a positive integer $L$, the short-time spectral magnitude is given by

$$S_w(nL,\omega)=\left|\sum_{m=-\infty}^{\infty} x(m)w(nL-m)e^{-j\omega m}\right|^2 \tag{2.2}$$

where the subscript $w$ in $S_w(nL,\omega)$ refers to the signal $w(n)$, known as the analysis window. In the *sliding−window* interpretation [5] of (2.2), the time-reversed analysis window $w(-n)$ shifts along the n-axis. After each shift of $L$ samples, $w(-n)$ is multiplied with $x(n)$; each product is called a short-time section of $x(n)$. The spectral magnitude of the short-time section for a particular window shift of $n_0L$ gives the frequency variation of $S_w(nL,\omega)$ for $n=n_0$. The *extent* of any particular analysis window position is defined as the region outside which the samples of the window are all zero. Then the *overlap* of two analysis windows is defined as the intersection of their extents. Note that when $L$ has minimum value 1, adjacent analysis window positions have maximum overlap for the allowable positive integer values of $L$. In this case, the short-time spectral magnitude is said to be computed with *maximum analysis window overlap*. Finally, when $L>1$, the short-time spectral magnitude is said to be computed with *partial analysis window overlap*.

If there were a unique correspondence between signals and their spectral magnitudes, the various short-time sections of $x(n)$ could be uniquely determined from their spectral magnitudes in $S_w(nL,\omega)$. However, the theory of all-pass spectral transformations [1] tells us that a signal is not uniquely specified by its spectral magnitude. For example, $x(n)$ and $x(-n)$ have the same spectral magnitude. More generally, when any poles and zeros of $x(n)$ are replaced by the inverse of their complex conjugates, a signal $y(n)$ is obtained which has the same spectral magnitude as $x(n)$. If any of the replaced poles and zeros is not on the unit circle, $y(n)$ is different from $x(n)$. Fortunately, $S_w(nL,\omega)$ has additional information about the short-time sections besides their spectral magnitudes. This information is contained in the overlap of the analysis window positions. For example, if one of the short-time sections is known, then the signals

corresponding to the spectral magnitude of an adjacent section have to be consistent in the region of overlap with the known short-time section. That is, the two sections should be identical in that region after dividing each of their non-zero samples by the corresponding samples of the analysis window. We will show in this chapter that the samples in the region of overlap can be uniquely extrapolated to obtain the entire unknown section.

Suppose $S_w(nL,\omega)$ is computed under conditions such that knowledge of any short-time section leads to the unique extrapolation of its neighboring short-time sections. Then, knowledge of just one particular short-time section triggers a series of extrapolations, where as a new short-time section is extrapolated, it becomes possible to extrapolate a succeeding short-time section that overlaps the one just extrapolated. Once all the short-time sections have been determined in this way, the final step is to combine these sections for obtaining the entire signal. Chapter 3 uses exactly such an extrapolation approach to determine conditions under which $S_w(nL,\omega)$ is a unique signal representation.

From the above discussion, it follows that the major theoretical problem in establishing unique correspondence between $x(n)$ and $S_w(nL,\omega)$ is one of signal extrapolation. Specifically, we wish to extrapolate a short-time section beyond its known samples, using its spectral magnitude. If the analysis window has finite extent, the resulting problem is equivalent to the extrapolation problem considered in this chapter.

## 2.2 Single-Sample Extrapolation

Consider a discrete-time signal $x(n)$ that is zero outside the interval $0 \le n \le N$. Theorems 2.1 and 2.2 of this section show that the sample $x(N)$ can be uniquely obtained from the spectral magnitude , $|X(\omega)|$, and $x(n)$ for $0 \le n < N$. However, the two theorems differ from each other in the number of samples of $x(n)$ and the number of samples of $|X(\omega)|$ actually used to accomplish the extrapolation. Compared to Theorem 2.2, Theorem 2.1 requires fewer samples of $x(n)$. On the other hand, Theorem 2.2 requires fewer samples of $|X(\omega)|$ than

Theorem 2.1.

**Theorem 2.1**

Let $x(n)$ be a sequence that is zero outside the interval $0 \leq n \leq N$. Suppose $x(0)$ is nonzero. Then, $2N$ or more samples of $|X(\omega)|$ over one period of $2\pi$ and the sample $x(0)$ uniquely specify the sample $x(N)$.

**Proof:**

From $|X(\omega)|^2$ the autocorrelation function $R(n)$ of $x(n)$ is obtained through the inverse Fourier transform.

$$R(n) = \sum_{m=-\infty}^{\infty} x(m)\, x(n+m) \tag{2.3}$$

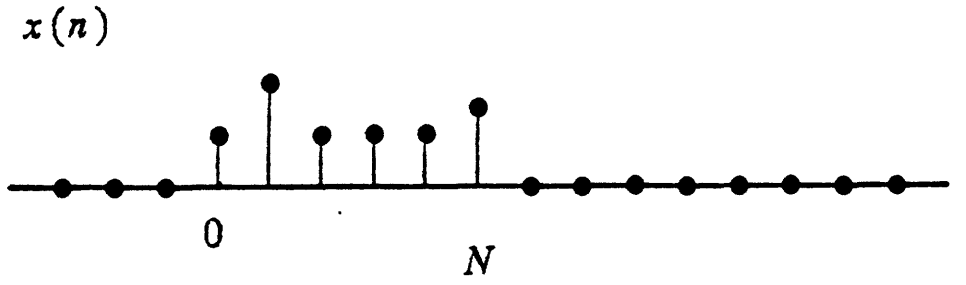
Since $x(0)$ is the first non-zero sample of $x(n)$ and $x(n)=0$ for $n>N$, it follows that ( see Figure 2.1 )

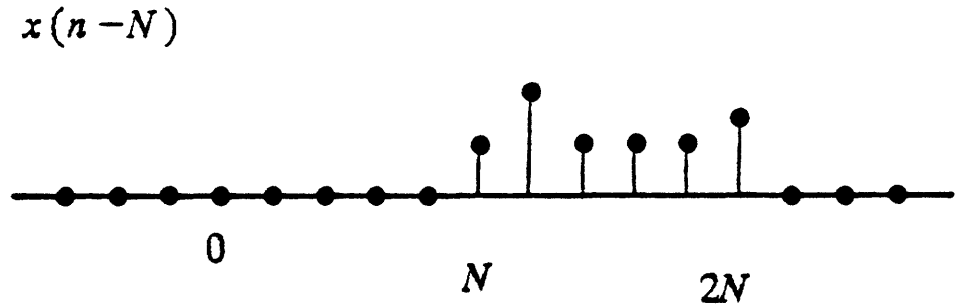
$$R(N) = x(0)\, x(N) \tag{2.4}$$

Therefore, since $x(0)$ is assumed known,

$$x(N) = R(N) / x(0) \tag{2.5}$$

Note that the autocorrelation value $R(N)$ is the only information derived from $|X(\omega)|$. Since, $x(n)$ is $N+1$ points long, $R(n)$ is $2N+1$ points long and an even function of n. Thus, the entire sequence $R(n)$ can be obtained without aliasing with a $2N+1$ point Inverse Discrete Fourier Transform (IDFT) of $|X(\omega)|^2$. However, with a 2N point IDFT, the sample $R(N)$ will be aliased with the sample R(-N-1). Since $R(N)=R(-N)$, it follows that $2R(N)$ can be obtained through a 2N point IDFT, requiring only $2N$ uniformly spaced samples of $|X(\omega)|^2$. This completes the proof of Theorem 2.1.




$$R(N) = \sum_{m=-\infty}^{\infty} x(n)x(n-N) = x(0)x(N)$$

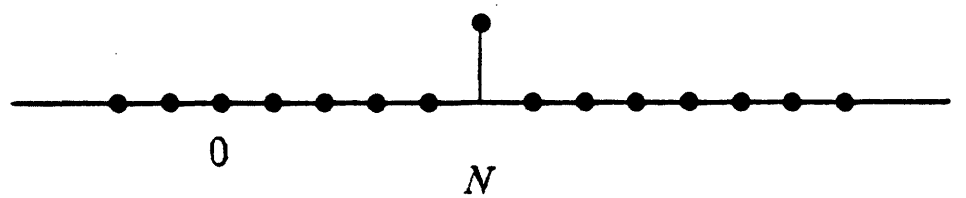



Fig. 2.1 The Computation of $R(N)$

The next theorem also concerns single sample extrapolation of finite-length signals. However, it uses a totally different approach in extrapolating $x(N)$ from the preceding samples of $x(n)$ and $|X(\omega)|$. In contrast to the proof of Theorem 2.1, the values of $|X(\omega)|^2$ are used directly in the extrapolation instead of first obtaining the autocorrelation $R(n)$ of $x(n)$.

**Theorem 2.2**

Let $x(n)$ be a sequence that is zero outside the interval $0 \leq n \leq N$. Assume that there is at least one non-zero sample of $x(n)$ in the interval $0 \leq n < N$. Then, $x(n)$ for $0 \leq n < N$ and two appropriately chosen samples of $|X(\omega)|$, uniquely specify the sample $x(N)$.

**Proof:**

Let $y(n)=x(n)w(n)$ where $w(n)$ is given by

$$w(n) = \begin{cases} 1 & 0 \leq n < N-1 \\ 0 & otherwise \end{cases} \tag{2.6}$$

Let $Y(\omega)$ denote the spectrum of $y(n)$. Then,

$$X(\omega) = Y(\omega) + x(N)\, e^{-j\omega N} \tag{2.7}$$

Taking the magnitude squared of both sides and rearranging the terms,

$$x^2(N) + b(\omega)\, x(N) + c(\omega) = 0 \tag{2.8}$$

where:

$$b(\omega)=2\ \mathrm{Re}\,[Y(\omega)\, e^{j\omega N}] \tag{2.9}$$

$$c(\omega) = |Y(\omega)|^2 - |X(\omega)|^2 \tag{2.10}$$

Note that $b(\omega)$ and $c(\omega)$ can both be determined from $|X(\omega)|$ and the N-1 samples preceding $x(N)$. When (2.8) is solved for $x(N)$, there are two solutions for each value of $\omega$. Consider two distinct values of $\omega$, say $\omega_1$ and $\omega_2$, in the interval $[0, \pi]$. Assume that at least one of $b(\omega)$ and

$c(\omega)$ changes when $\omega$ is changed from $\omega_1$ to $\omega_2$. Then, from the properties of quadratic equations, the two solutions associated with $\omega_1$ cannot be the same as the pair of solutions for $\omega_2$. However, one of the solutions must be identical and that is the true value of $x(N)$. It now remains to show that provided the $N-1$ values preceeding $x(n)$ are not all zero, one can always find $\omega_1$ and $\omega_2$ for which two different quadratic equations are obtained from (2.8).

Two values of $\omega$ giving two distinct equations from (2.8) can be found if $b(\omega)$ is not independent of $\omega$. Our approach here will be to show that $b(\omega)$ is independent of $\omega$ in only one case -- when the $N$ samples preceding $x(N)$ are all zero. The sequence $y(n)$ falls in the region $0 \leq n < N$. Thus, the inverse Fourier transform of $Y(\omega)e^{j\omega N}$, denoted by $\tilde{y}(n)$, falls in the region $-N \leq n < 0$. However, for $b(\omega)$ not to depend on $\omega$, the Fourier transform of $\tilde{y}(n)$ must have a constant real part. That is, y(n) must be of the form $A\,\delta(n) + q(n)$ where A is real, $\delta(n)$ is the unit sample sequence, and q(n) is an odd sequence. Therefore $b(\omega)$ is independent of $\omega$ only when y(n)=0 for all n, i.e., the $N$ values preceeding $x(N)$ are all zero. In this situation, $c(\omega)$ from (2.10) is also independent of $\omega$. Thus, when the $N$ samples of $x(n)$ preceding $x(N)$ are all zero, this is the only situation when (2.8) does not have a unique solution for $x(N)$.

In fact, such values of $\omega$ can be found even when the various frequency functions are sampled at the rate $2\pi/M$ where $M \geq 2N-2$. In such a case, $\tilde{y}(n)$ is replaced by

$$\tilde{\tilde{y}}(n) = \sum_{p=-\infty}^{\infty} \tilde{y}(n+pM) \tag{2.22}$$

The requirement that $b(2\pi r/M)$ be independent of r then becomes the requirement that $\tilde{\tilde{y}}(n)$ be an odd sequence. It can be verified that $\tilde{\tilde{y}}(n)$ is odd if and only if $y(n)=0$ for all n. This completes the derivation.

### 2.3 Multiple Samples Extrapolation

This section presents a theorem on the extrapolation of of a finite-length sequence $x(n)$

with more than one unknown sample, using the spectral magnitude, $|X(\omega)|$. Once again, $x(n)$ is assumed zero outside the interval $0 \leq n \leq N$. As indicated in the beginning of this chapter, the location of this interval on the n-axis may be changed without affecting the results derived here.

It should be noted that theorem 2.3 below uses the autocorrelation function, $R(n)$, of $x(n)$ to determine the unknown samples. This is analogous to the way $x(N)$ was determined from $R(n)$ in the proof of theorem 2.1. In fact, theorem 2.1 can be derived as a corollary of theorem 2.3. However, we chose not to do this in order to emphasize the simplicity of the direct proof of theorem 2.1.

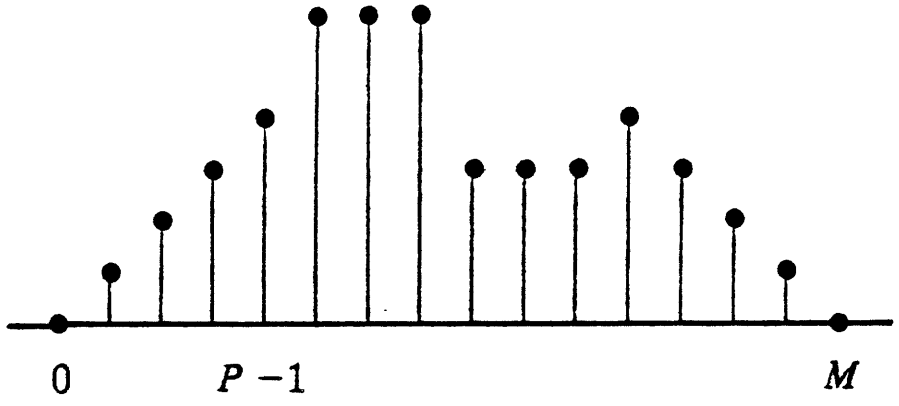
**Theorem 2.3**

Let $x(n)$ be a sequence that is zero outside the interval $0 \leq n \leq N$. Suppose $x(0)$ is non-zero. Then, $2N$ or more samples of $|X(\omega)|$ over one period of $2\pi$ and the $P$ samples of $x(n)$ in the interval $0 \leq n < P$ uniquely specify the entire sequence $x(n)$ if and only if $P \geq \lceil M/2 \rceil$ (where $M = N+1$ and $\lceil \alpha \rceil$ is the smallest integer greater or equal to $\alpha$).

**Proof:**

Throughout this proof, the samples of $x(n)$ for $0 \leq n < P$ will be referred to as the inital $P$ samples of $x(n)$. We first provide a counter-example to show that if $P < \lfloor M/2 \rfloor$, then $x(n)$ cannot in general be uniquely specified by $|X(\omega)|$ and the initial $P$ samples.

With $P < \lfloor M/2 \rfloor$ consider any sequence $x(n)$ such that $x(n) = x(M-1-n)$ for $n = 0,1,\ldots P-1$, and $x(n) \neq x(M-1-n)$ for $n = P, P+1,\ldots, M-P-1$ (See Figure 2.2). Then, the sequences $x(n)$ and $y(n) = x(M-1-n)$ have the same samples for $n = 0,1,\ldots P-1$. Furthermore, since $y(n)$ is a time-reversed version of $x(n)$, the two sequences have the same Fourier transform magnitude [1]. Since $x(n) \neq x(M-1-n)$ for $n = P, P+1,\ldots, M-P-1$, $y(n)$ and



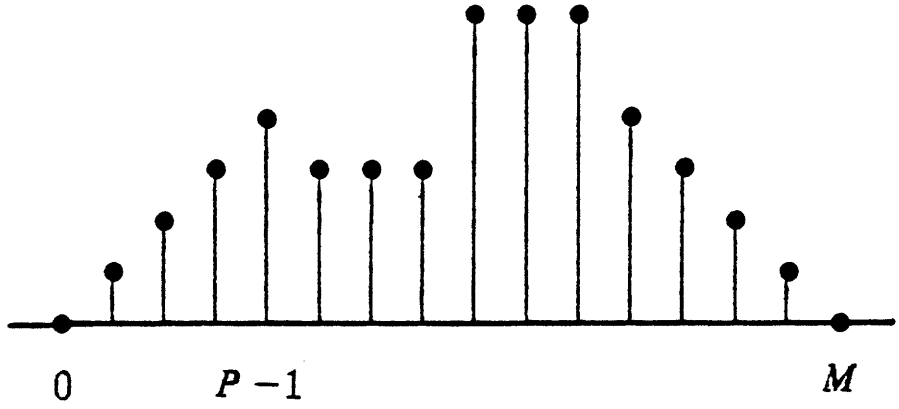




Fig. 2.2 Counter-example for the Proof of Theorem 2.3

$x(n)$ are distinct. Thus, the initial $P$ samples and $|X(\omega)|$ are not sufficient to uniquely represent x(n).

We now develop a procedure for uniquely recovering the unknown samples of $x(n)$ when $P \geq \lceil M/2 \rceil$. From $2N$ uniform samples of $|X(\omega)|$, we saw in the proof of Theorem 2.1 that the autocorrelation $R(n)$ of $x(n)$ can be obtained.

$$R(n) = x(n)*x(-n) = \sum_{m=0}^{M-1-n} x(m)\,x(n+m) \tag{2.11}$$

Consider the case where $M$ is even. From (2.11), $M/2$ linear equations are obtained in $M/2$ unknowns, $x(M/2), x((M/2)+1), \ldots, x(M-1)$. In matrix form these equations are:

$$\begin{bmatrix} x(0) & & & \\ x(1) & x(0) & & \\ x(2) & x(1) & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ x((M/2)-1) & x((M/2)-2) & \cdots & x(0) \end{bmatrix} \begin{bmatrix} x(M-1) \\ x(M-2) \\ \\ \cdot \\ \cdot \\ \cdot \\ x(M/2) \end{bmatrix} = \begin{bmatrix} R(M-1) \\ R(M-2) \\ \\ \cdot \\ \cdot \\ \cdot \\ R(M/2) \end{bmatrix} \tag{2.12}$$

The left matrix is lower triangular with all diagonal elements $x(0)$. Since $x(0) \neq 0$ by assumption, this matrix is invertible. Thus, a unique solution exists for $x(n)$, $n = M/2, (M/2)+1, \ldots M-1$. For $M$ odd, the $\lfloor M/2 \rfloor$ unknowns, $x((M+1)/2), x(((M+1)/2)+1), \ldots, x(M-1)$, are solved for through a set of equations similar to (2.12). Thus, for $P \geq \lceil M/2 \rceil$, the uniqueness of $x(n)$ follows regardless of whether M is even or odd.

The one remaining case is when M is odd and $P = \lfloor M/2 \rfloor$. In this case, our theorem asserts that a unique solution for $x(n)$ does not exist. To show this, consider the sequence $x(n)$ to be such that $x(n) = -x(M-1-n)$ for $n = 0,1,\ldots P-1$, and $x(P) \neq 0$. Then, the sequences $x(n)$ and $y(n) = -x(M-1-n)$ have the same samples for $n = 0,1,\ldots P-1$ and $y(P) = -x(P)$. On the other hand, it is easily seen that $|Y(\omega)| = |X(\omega)|$. This completes the proof of Theorem 2.3.

# CHAPTER THREE: SIGNAL REPRESENTATION WITH SHORT-TIME SPECTRAL MAGNITUDE

In this chapter, we address the problem of uniquely representing a signal by its short-time spectral magnitude. We assume that the analysis window of the short time spectrum is a *known* finite-length sequence. This permits us to use the extrapolation theorems of chapter 2 for developing conditions which ensure unique correspondence between a signal and its short-time spectral magnitude. These conditions place restrictions on the finite-length analysis window as well as the signal being represented. The need for such conditions is discussed in section 3.1. In section 3.2 we present various conditions for unique signal representation with the short-time spectral magnitude. Most of these conditions concern the representation of *one-sided* signals. That is, signals which are always zero either before (right-sided) or after (left-sided) some point on the time axis. These conditions do not represent all the possible situations in which a signal is uniquely specified by its short-time spectral magnitude. However, the conditions we develop are broad enough to be of significant practical interest, as illustrated in later chapters. This chapter closes with section 3.3 which shows how the uniqueness conditions can be easily extended to the short-time spectral magnitude of multidimensional signals.

## 3.1 Uniqueness Problems

For a signal $x(n)$ and a positive integer $L$, the short-time spectral magnitude is given by

$$S_w(nL,\omega)=\left|\sum_{m=-\infty}^{\infty} x(m)w(nL-m)e^{-j\omega m}\right|^2 \tag{3.1}$$

where the subscript $w$ in $S_w(nL,\omega)$ refers to the analysis window, $w(n)$. In the *sliding window* interpretation [5], $S_w(nL,\omega)$ for each $n$ is viewed as representing the spectral magnitude of the short-time section $f_n(m)=x(m)w(nL-m)$. When $L=1$, the short-time spectral magnitude is said to have maximum analysis window overlap. On the other hand, if $L>1$, the short-time spec-

tral magnitude has partial analysis window overlap. In this section, we discuss some situations where $x(n)$ is not uniquely represented by $S_w(nL,\omega)$. This helps us select the conditions developed in the next two sections for ensuring unique specification of $x(n)$ with $S_w(nL,\omega)$.

At least one condition is easily shown to be necessary on $x(n)$ for unique correspondence with the short-time spectral magnitude, $S_w(nL,\omega)$. In expression (3.1) for $S_w(nL,\omega)$, when $x(n)$ is replaced by $-x(n)$, the minus sign is absorbed by the absolute value operation. Thus, $x(n)$ and $-x(n)$ have the same short-time spectral magnitude. This ambiguity may be resolved, for example, by knowing the sign of some non-zero sample of $x(n)$.

In the case of a finite-length analysis window, a gap of zero samples between two non-zero portions of $x(n)$ can also lead to ambiguity in signal representation with $S_w(nL,\omega)$. Suppose $x(n)$ is the sum of two signals, $x_1(n)$ and $x_2(n)$, occupying different regions of the n-axis ( See Figure 3.1 ). Suppose that the gap of zeros between $x_1(n)$ and $x_2(n)$ is large enough so that there is no analysis window position for which the corresponding short-time section includes non-zero contribution from $x_1(n)$ as well as $x_2(n)$. Clearly, in such a situation, the short-time spectral magnitude of x(n) is the sum of the short-time spectral magnitudes of $x_1(n)$ and $x_2(n)$. However, we previously saw that a signal and its negative have the same short-time spectral magnitude. It follows that $x(n)$ has the same short-time spectral magnitude as the signals obtained from the differences $x_1(n) - x_2(n)$ and $x_2(n) - x_1(n)$ ( See Figure 3.1 ). We conclude that if there is a large enough gap of zero samples, there will be sign ambiguities on either side of the gap. Consequently, all the uniqueness conditions developed in this chapter include a restriction on the length of zero gaps between non-zero portions of the signal.

In section 3.2 we will see that $S_w(nL,\omega)$ with $L=1$ uniquely specifies a one-sided signal $x(n)$ under conditions whose only restriction on $x(n)$ is a limit on the size of any zero gaps. The known analysis window is restricted to have no zero samples within its finite length. This condition is satisfied by commonly used rectangular, triangular, Hamming and Hanning windows.

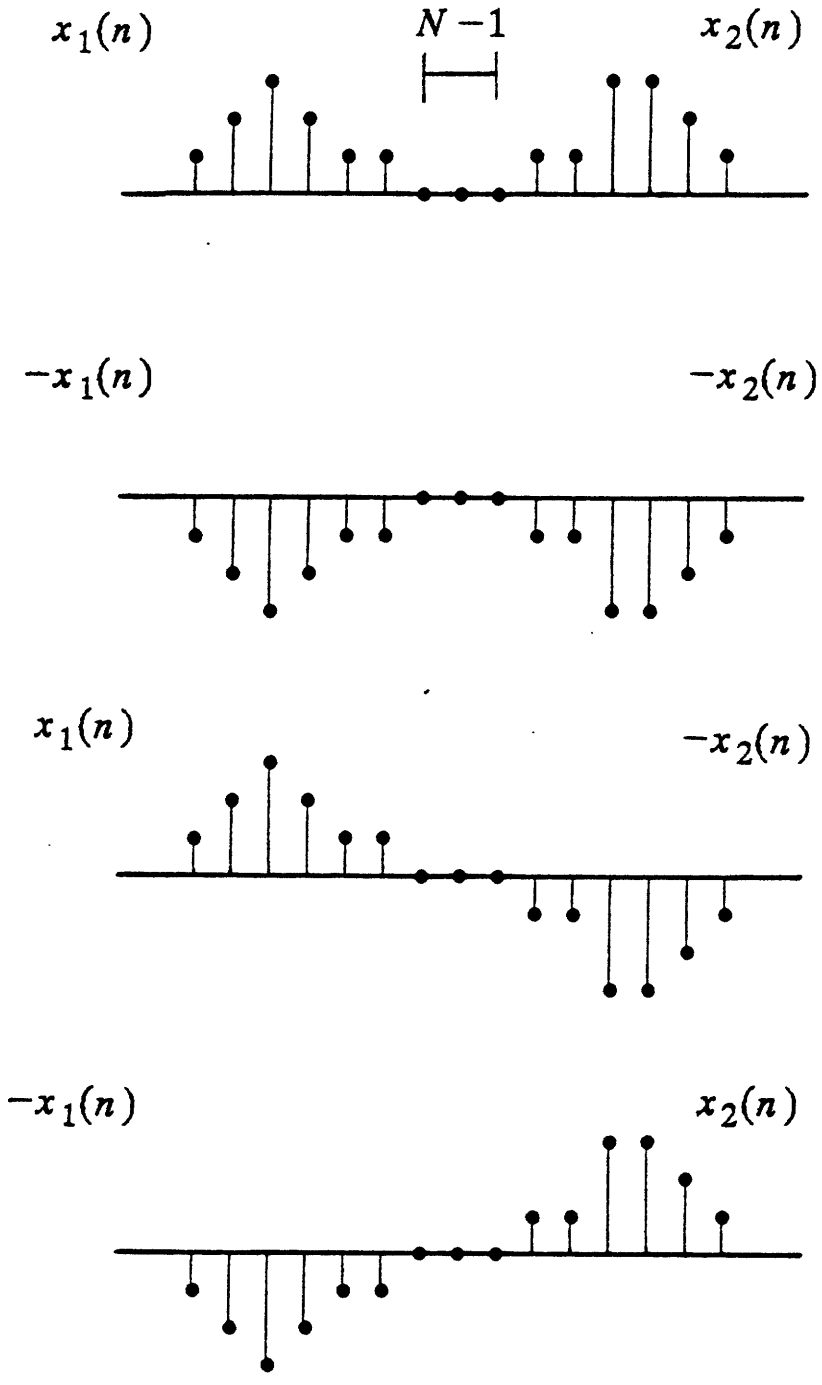



Fig. 3.1 Four Sequences with same $S_w(n,\omega)$

With such analysis windows, we will now see that for $L > 1$ in $S_w(nL,\omega)$, the zero gap restriction on $x(n)$ is not sufficient to guarantee signal specification even up to a sign ambiguity. To show this, we construct a class of sequences that have no zero samples between any two nonzero samples; these sequences have the property that they are not specified even up to a sign factor by $S_w(nL,\omega)$ with $L>1$ and $w(n)$ a rectangular window whose length is a multiple of $L$.

For $M \geq 1$ construct M sequences $x_i(n)$, $i=1,2,\ldots,M$ where each $x_i(n)$ has finite length $L$ and falls in the region $1 \leq n \leq L$. Furthermore, constrain the z-transform $X_i(z)$ of each $x_i(n)$ to have $Q$ of its zeros from an arbitrarily specified set $a_1,a_2,\ldots,a_Q$, none of which lie on the unit circle and $Q<L$. Thus, for each i, $X_i(z)$ can be factored as

$$X_i(z) = [\prod_{j=1}^{Q}(1-a_j z^{-1})]\hat{X}_i(z) \tag{3.2}$$

Now, let

$$x(n)=x_1(n)+x_2(n-L)+\cdots+x_M(n-(M-1)L) \tag{3.3}$$

Then, the z-transform of $x(n)$ is

$$\begin{aligned} X(z) &= X_1(z)+z^{-1}X_2(z)+\cdots+z^{-(M-1)L}X_M(z) \\ &= [\prod_{j=1}^{Q}(1-a_j z^{-1})][\sum_{j=0}^{M-1} z^{-jL}\hat{X}_{j+1}(z)] \end{aligned} \tag{3.4}$$

which also contains the zeros $a_j$ for $j=1,2,\ldots,Q$.

Now consider an analysis window $w_1(n)$ defined for some integer $r>1$ by

$$w_1(n) = \begin{cases} 1 & 0 \leq n < rL \\ 0 & otherwise \end{cases}$$

Observe that $y_k(n)=x(n)w_1(kL-n)$ for any fixed k is given by

$$y_k(n)=x_j(n-(j-1)L)+x_{j+1}(n-jL)+\ldots+x_p(n-(p-1)L) \tag{3.5}$$

for some consecutive integers $j, j+1, \cdots p$ determined from the set {1,2,...,M}. Clearly, the z-transform $Y_k(z)$ of each $y_k(n)$ has $a_1,a_2,\ldots,a_Q$ among its zeros. Thus, if one or more of the $a_i$'s is reflected about the unit circle to $1/a_i^*$ then $|X(\omega)|$ as well as $|Y_k(\omega)|$ for each k remains the same [1]. Thus, there are $2^Q$ distinct sequences with the same $|Y_n(\omega)|=S_w(nL,\omega)$. Of course,

from those $2^Q$ sequences another $2^Q$ sequences with the same $S_w(nL,\omega)$ are obtained by forming the negative of the first $2^Q$ sequences. Thus, there are $2^{Q+1}$ distinct sequences with the same $S_w(nL,\omega)$. Recall the maximum attainable $Q$ is $L-1$. Also note that each $x_i(n)$ can be chosen to guarantee $x(n)$ is non-zero over its finite length. Thus, with $L>1$, a class of sequences with no zeros between nonzero samples have more than just a sign ambiguity in their representation with $S_w(nL,\omega)$. For example, even with $L=2$ their exist finite-length sequences with no zero samples over their duration such that there is an ambiguity of $2^2 = 4$ in the representation with $S_w(nL,\omega)$.

We have established that for unique specification of $x(n)$ by $S_w(nL,\omega)$ with L > 1, we require additional information on $x(n)$ besides the one-sided and zero gap restrictions. In section 3.2.2, knowledge of the $L$ initial samples of $x(n)$ is found to be sufficient for this purpose. This condition arises naturally from the extrapolation approach used in deriving the various results in the remainder of this chapter.

### 3.2 Uniqueness Conditions

In this section, we present various conditions and their derivations for uniquely representing a signal with its short-time spectral magnitude. The analysis window of the short-time spectrum is assumed to be a known finite-length sequence. The uniqueness conditions presented are *sufficient* but not necessary to guarantee unique correspondence between a signal and its short-time spectral magnitude These conditions are divided in this section into two main categories, according to whether or not maximum analysis window overlap is used in the computation of the short-time spectrum.

### 3.2.1 Maximum Analysis Window Overlap

The short-time spectral magnitude $S_w(nL,\omega)$, defined in (3.1), may be viewed for each $n$ as the spectral magnitude of the short-time section $f_n(m) = x(m)w(nL-m)$. When $n$ is incremented by one, the time-reversed analysis window $w(nL-m)$ shifts $L$ sample positions. Since (3.1) is defined for positive integer values of $L$, it is clear that with $L=1$, adjacent analysis window positions have maximum overlap. In this case, we denote the short-time spectral magnitude by $S_w(n,\omega)$.

We are interested in developing conditions that guarantee unique signal representation with $S_w(n,\omega)$ when the analysis window is a known finite-length sequence. For this purpose, Theorem 2.1 on single sample extrapolation of finite-length sequnces is extremely useful. For easy reference, we restate this theorem from chapter 2.

**Theorem 2.1**

Let $x(n)$ be a sequence that is zero outside the interval $0 \le n \le N$. Suppose $x(0)$ is nonzero. Then, 2N or more samples of $|X(\omega)|$ over one period of $2\pi$ and the sample $x(0)$ uniquely specify the sample $x(N)$.

Although the theorem is stated for $x(n)$ in the interval $0 \le n \le N$, it also holds for $x(n)$ in any other interval on the n-axis. This is accomplished by a change of reference on the n-axis such that the first non-zero sample of $x(n)$ falls at the origin of the new coordinate system.

We now state our first set of conditions for uniquely specifying a signal $x(n)$ with $S_w(n,\omega)$. In this case we restrict the signal $x(n)$ to be one-sided. That is, $x(n)=0$ for $n<n'$ or $n>n'$ for some integer $n'$. Of course, the analysis window must have at least one non-zero sample so that $S_w(n,\omega)$ is not zero for all signals. Furthermore, we restrict $w(n)$ to be non-zero over its finite

length, $N_w$. This simplifies the type of restriction imposed on $x(n)$ for avoiding the zero gap ambiguities discussed in section 3.1.

**Conditions 3.1 : For Representing $x(n)$ Uniquely With $S_w(n,\omega)$**

| | |
|---|---|
| w(n): | a) Known sequence of finite length $N_w$ |
| | b) No zeros within length $N_w$ |
| x(n): | a) One-sided |
| | b) At most $N_w - 2$ consecutive zero samples between any two non-zero samples |
| | c) Sign of first non-zero sample known |

To show that $S_w(n,\omega)$ uniquely specifies the signal $x(n)$ under Conditions 3.1, let us consider the case when the analysis window $w(n)$ is restricted to the interval $0 \leq n < N_w$. We do not lose any generality with this assumption because it can be easily accounted for by a change of reference on the n-axis. Under Conditions 3.1, we now show a procedure for recovering $x(n)$ from $S_w(n,\omega)$. The derivation is completed by showing that $x(n)$ is the only sequence that could have been obtained from $S_w(n,\omega)$ under Conditions 3.1.

We will consider only the case with $x(n)$ right-sided. The case with $x(n)$ left-sided can be proved analogously. Let $n'$ be the smallest value of $n$ such that $x(n')$ is non-zero. Then, with $L=1$ in (3.1) and $w(n)$ as assumed above, it follows that $S_w(n,\omega)$ is zero for all $n < n'$. Furthermore,

$$S_w(n',\omega) = w^2(0)\, x^2(n') \quad \text{for } \textit{all } \omega \tag{3.6}$$

We then have

$$x(n') = \pm \frac{\sqrt{S_w(n',0)}}{w(0)} \tag{3.7}$$

The sign ambiguity in this equation can be resolved since Conditions 3.1 specify the sign of the first non-zero sample, $x(n')$. Having determined $x(n')$, the next step is to use Theorem 2.1 for obtaining $x(n'+1)$. The short-time section $f_{n'+1}(m) = x(m)w(n'+1-m)$ has zero samples outside the interval $n' \leq m \leq n'+1$ and all its samples are known except at $m=n'+1$ where it equals $x(n'+1)\, w(0)$. The spectral magnitude $S_w(n'+1,\omega)$ of this section is known. Thus, applying Theorem 2.1 with N=2, $x(n'+1)w(0)$ can be extrapolated. Since, $w(n)$ was assumed known and non-zero over $0 \leq n \leq N_w$, we divide $x(n'+1)w(0)$ by $w(0)$ to obtain $x(n'+1)$. We now continue such a procedure to determine each unknown sample of $x(n)$ after the samples preceding it have been determined. However, Theorem 3.1 requires that at least one of the $N_w-1$ preceding samples be non-zero. This recursive procedure for determining $x(n)$ for $n>n'$ can be easily expressed in closed form. For each n, let $r_n(m)$ denote the autocorrelation function corresponding to $S_w(n,\omega)$. The autocorrelation function is given by

$$r_n(m) = \sum_{k=n-(N_w-1)}^{n} x(k)w(n-k)\, x(k-m)w(n-(k-m)) \tag{3.8}$$

Solving this equation for $x(n)$, we obtain

$$x(n) = \frac{r_n(m) - \sum_{k=n-(N-1)}^{n-1} w(n-k)w(n-(k-m))x(k)\, x(k-m)}{w(0)w(m)x(n-m)} \tag{3.9}$$

This is a valid equation only for values of $m$ for which $w(m)x(n-m)$ is non-zero. Since $w(m)$ is non-zero only for $0 \leq m < N_w$, we require that $x(n-m)$ be non-zero for some $m$ in $0 < m < N_w$. This leads to the requirement that $x(n)$ have no more than $N_w-2$ zero samples between any two non-zero samples. This is consistent with our observation in section 3.1 that there should not be a zero gap separating two non-zero portions of the signal such that no analysis window position has contributions from both the non-zero portions. Since Conditions 3.1 include this requirement, it follows that the signal $x(n)$ can be obtained from $S_w(n,\omega)$ using the procedure we have just outlined.

Suppose there is another signal $x'(n)$ satisfying Conditions 3.1 and for which the sign of

the first non-zero sample is the same as the sign of the first non-zero sample of $x(n)$. Since $S_w(n,\omega)$ has its first non-zero value at $n=n'$, it follows that $x'(n')$ must be the first non-zero value of $x'(n)$. We can then use the same reconstruction procedure for obtaining $x'(n)$ from $S_w(n,\omega)$ as we used for obtaining $x(n)$ from $S_w(n,\omega)$. However, that procedure only yielded one answer. It follows that $x(n)=x'(n)$. We conclude that $x(n)$ is uniquely represented by $S_w(n,\omega)$ under Conditions 3.1.

From section 3.1, we know that $-x(n)$ has the same $S_w(nL,\omega)$ as $x(n)$. It follows that under Conditions 3.1, $-x(n)$ can be uniquely obtained from $S_w(n,\omega)$. However, the only difference in obtaining $x(n)$ and $-x(n)$ using the procedure outlined above is that different signs are selected for $x(n')$ in (3.7). It follows that without the a-priori sign knowledge in Conditions 3.1, $x(n)$ could have been obtained up to a sign ambiguity from $S_w(n,\omega)$.

The following set of conditions deal with the sign ambiguity in the representation with $S_w(n,\omega)$ by restricting the class of signals under consideration to be non-negative. In this case, the sign ambiguities due to any zero gaps also disappear.

**Conditions 3.2 : For Representing $x(n)$ Uniquely With $S_w(n,\omega)$**

| | |
|---|---|
| w(n): | a) Known sequence of finite length and at least one nonzero sample |
| x(n): | a) One sided |
| | b) Non-negative |

This set of conditions as well as Conditions 3.1 restrict $x(n)$ to be one sided. Lets consider extending the class of signals we can uniquely specify with the short-time spectral magnitude.

To start the recursion of (3.9), knowledge of $N_w - 1$ consecutive samples of $x(n)$ is sufficient, provided one of those known samples is non-zero. Therefore, the requirement that $x(n)$ be one sided is not necessary. Furthermore, although the recursion was derived for increasing n, a similar procedure can be derived for decreasing n. Using these observations, the following conditions for unique signal representation with $S_w(n,\omega)$ can be derived that apply to a wider class of signals than Conditions 3.1.

**Conditions 3.3 : For Representing $x(n)$ Uniquely With $S_w(n,\omega)$**

| | |
|---|---|
| w(n): | a) Known sequence of finite length $N_w$ |
| | b) No zeros within length $N_w$ |
| x(n): | a) $N_w - 1$ consecutive samples known, at least one of which is nonzero |
| | b) At most $N_w - 2$ consecutive zero samples between any two nonzero samples |

### 3.2.2 Partial Analysis Window Overlap

We will now develop a set of conditions that are sufficient for uniquely specifying a signal with its short-time spectral magnitude which is computed with partial analysis window overlap ( i.e. $L > 1$ in $S_w(nL,\omega)$ ). The signal $x(n)$ is restricted to be one-sided. Furthermore, the analysis window $w(n)$ is assumed to be a known sequence with no zero samples over its finite length. As shown in section 3.1, even if we do not allow any zero samples within finite-length $x(n)$, there are signals which are not specified even up to a sign ambiguity by $S_w(nL,\omega)$ with $L > 1$. In the conditions below, we counter those ambiguities with knowledge of L consecutive samples of the signal, starting from the first non-zero sample.

**Conditions 3.4 : For Representing $x(n)$ Uniquely With $S_w(nL,\omega)$**

| | |
|---|---|
| L : | a) $1<L\leq\lfloor N_w/2\rfloor$, |
| w(n): | a) Finite length $N_w>2$ |
| | b) No zeros within length $N_w$ |
| x(n): | a) One-Sided |
| | b) At most $N_w-2L$ consecutive zeros between any two nonzero samples |
| | c) L consecutive samples known, starting from the first non-zero sample |

In the above conditions $\lceil x\rceil$ denotes the smallest integer greater or equal to $x$. The derivation of these conditions relies heavily on Theorem 2.3 of chapter 2, restated below for easy reference.

**Theorem 2.3**

Let $x(n)$ be a sequence that is zero outside the interval $0\leq n\leq N$. Suppose $x(0)$ is non-zero. Then, $2N$ or more samples of $|X(\omega)|$ in an interval of $2\pi$ and the $P$ samples of $x(n)$ in the interval $0\leq n<P$ uniquely specify the entire sequence $x(n)$ **if and only if** $P\geq\lceil M/2\rceil$ ( where $M=N+1$ and $\lceil\alpha\rceil$ is the smallest integer greater or equal to $\alpha$).

As indicated in chapter 2, this extrapolation theorem holds regardless of the position of $x(n)$ on the n axis. Let $n'$ be the smallest $n$ for which $x(n)\neq 0$. Without loss of generality, assume $1\leq n'\leq L$ (See Figure 3.2a). Let $x_L(n)$ denote a sequence which equals $x(n)$ for $n'\leq n<n'+L$ and is zero otherwise. Thus, $x_L(n)$ represents the L known initial samples required by Conditions 3.4. Without loss of generality we assume that $w(n)$ occupies the region $0\leq n<N_w$. Since $x_L(n)$ is known, it follows that $x(n)$ under the analysis window $w(L-n)$ is

known. The first objective is to recover any unknown samples of $x(n)$ over the duration of $w(2L-n)$.

In order to recover the unknown samples of $x(n)$ under $w(2L-n)$, consider the sequence $y_2(n)=x(n)w(2L-n)$ illustrated in Figure 3.2b. Since $P \geq L$ and $L \leq \lfloor N_w/2 \rfloor$, knowledge of $x_L(n)$ assures that at least $L$ samples of $y_2(n)$ beginning at $n=n'$ are known. Furthermore, the length of $y_2(n)$ is $2L-n'+1$ and $L \geq \lceil (2L-n'+1)/2 \rceil$. Therefore, applying Theorem 2.3, the unknown samples of $y_2(n)$ are uniquely determined by $S_w(nL,\omega)$ and the initial conditions $x_L(n)$. Since $w(n)$ is nonzero over its duration, the unknown values of $x(n)$ under $w(2L-n)$ are obtained by division.

We have now determined $x(n)$ up to $n=2L$. We will next show that if $x(n)$ is known up to $n=(k'-1)L$, then $x(n)$ is uniquely determined up to $n=k'L$ under Conditions 3.4. By induction, $x(n)$ is then uniquely determined for all $n \geq n'$.

Consider the short-time segment $y_{k'}(n)=x(n)w(k'L-n)$ for a particular $k=k'$. Suppose further that $x(n)$ is known up to the last sample of $w((k'-1)L-n)$, that is, up to $n=(k'-1)L$. Then beginning at $n=k'L-N_w+1$ (See Figure 3.3), $N_w-L$ consecutive samples of $y_{k'}(n)$ are known. The next objective is to recover the last $L$ samples from the first $N_w-L$ samples of $y_{k'}(n)$. Clearly, the ability to do so depends on the value of $L$.

Suppose $L > \lfloor N_w/2 \rfloor$. Then $N-L < \lfloor N_w/2 \rfloor$. Consequently, from Theorem 2.3, the unknown L samples of $y_{k'}(n)$ are not uniquely specified by $S_w(k'L,\omega)$.

Suppose $1 < L \leq \lfloor N_w/2 \rfloor$. Furthermore, suppose that the initial value $y_{k'}(k'L-N_w+1)$ is non-zero. The $N_w-L$ values of $y_{k'}(n)$ starting from $n=k'L-N_w+1$ are known. Since $N-L \geq \lceil /2 \rceil$ and $S_w(k'L,\omega)$ is known, $y_{k'}(n)$ is completely determined by using Theorem 2.3. Now consider the cases when the first non-zero value of $y_{k'}(n)$ occurs beyond $n=k'L-N_w+1$. In particular, suppose that there are at most J consecutive zeros in $y_{k'}(n)$ starting at $n=k'L-N_w+1$

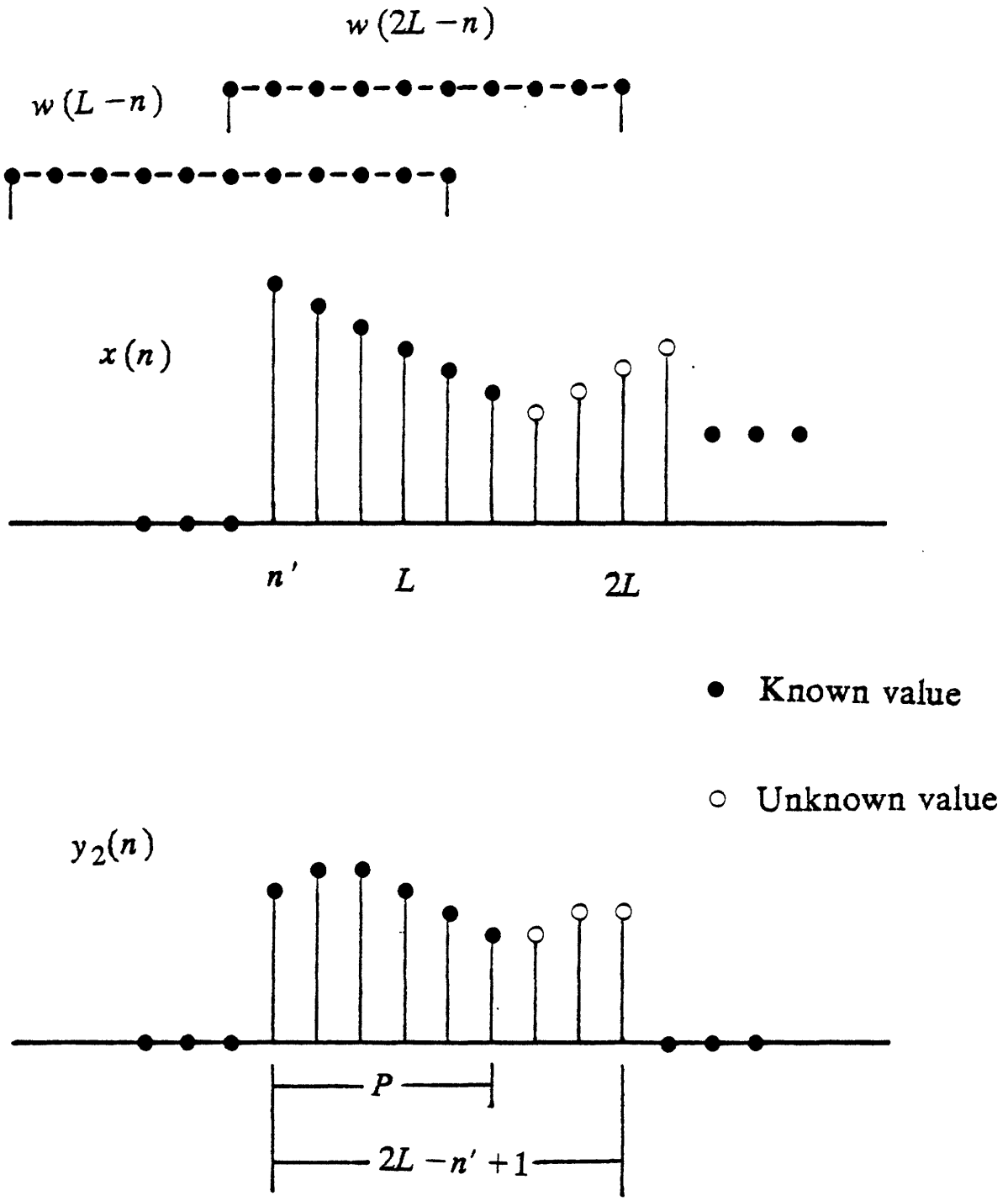



Fig. 3.2 Sequences for Proof of Conditions 3.4

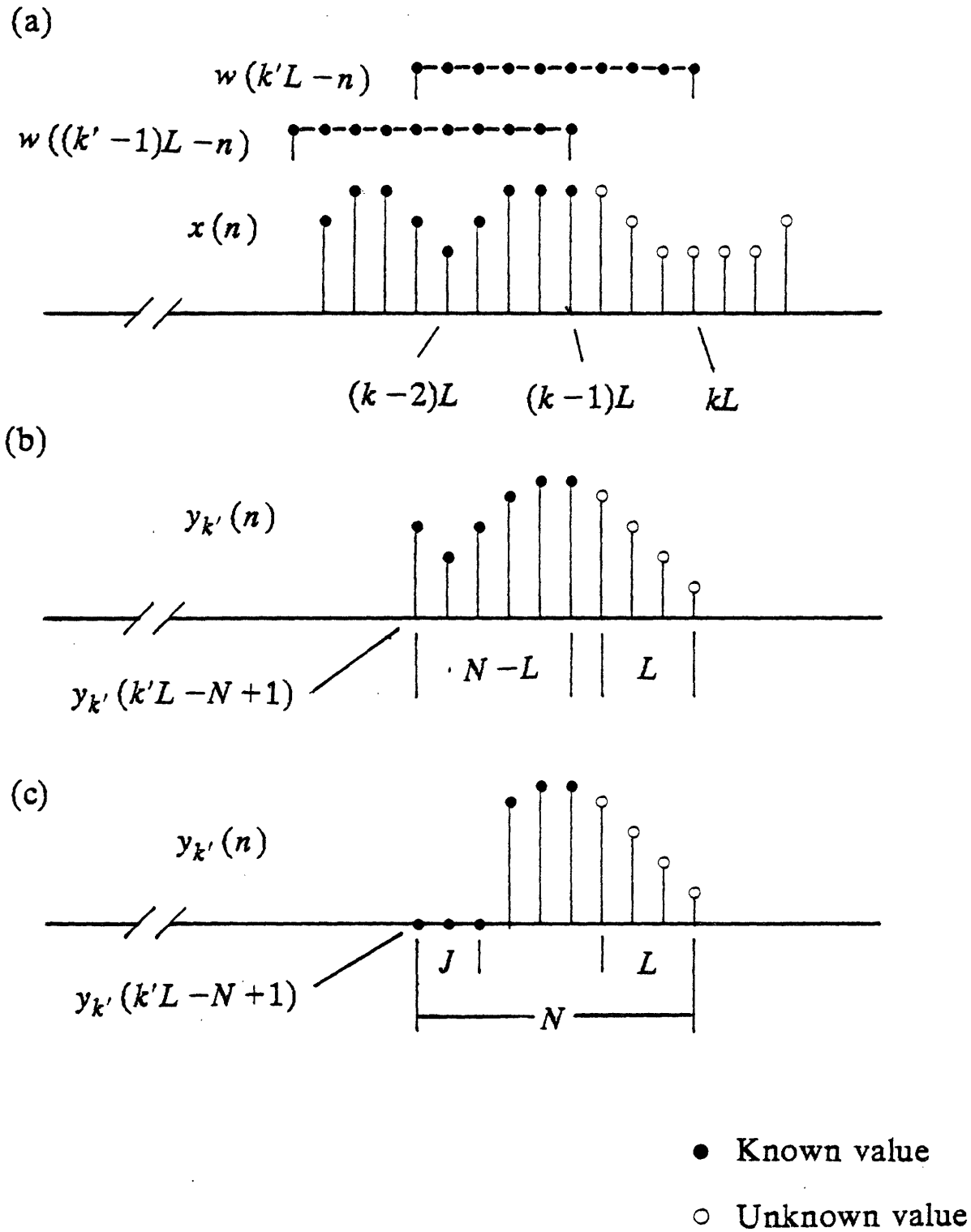



Fig. 3.3 Sequences for Proof of Conditions 3.4

(See Figure 3.3c). Lets find the largest J for which the $L$ unknown samples of $y_{k'}(n)$ can be determined. Theorem 2.3 requires at least $L$ known samples preceeding the $L$ unknown samples. Thus, the maximum allowable value of J is $N_w-[L+L]=N_w-2L$. This is consistent with Conditions 3.4.

We have shown that $x(n)$ can be uniquely determined from $S_w(nL,\omega)$ under Conditions 3.4. Suppose another signal $x'(n)$ also had the short-time spectrum $S_w(nL,\omega)$ and satisfied Conditions 3.4 with the same initial $L$ known samples as $x(n)$. Applying the procedure outlined above, we obtain $x'(n)$. However , the procedure is identical to the one used for obtaining $x(n)$. Since the procedure gives a unique answer, it follows that $x'(n)=x(n)$. Thus, under Conditions 3.4, a signal is uniquely represented by its short-time spectral magnitude.

### 3.3 Multidimensional Extension

This section extends signal representation with short-time spectral magnitude to multidimensional discrete-time signals. Since the extension is conceptually straightforward but notationally cumbersome, it will be presented here only for the short-time spectral magnitude with maximum analysis window overlap and for two-dimensional signals with finite support. For a two dimensional signal $x(n,m)$, the short-time spectral magnitude is given by

$$S_w(n,m;\omega,\nu)=\left|\sum_{m_1=-\infty}^{\infty}\sum_{m_2=-\infty}^{\infty} x(m_1,m_2)w(n-m_1,m-m_2)e^{-j\omega m_1}e^{-j\nu m_2}\right|^2$$

where $w(n,m)$ is the two dimensional analysis window.

Let $[x(n,m)]_N$ represent the class of two-dimensional signals whose finite regions of support contain no blocks of zeros larger than (N-2) x (N-2). This is a generalization of the one-dimensional condition of finite length with no gaps of more than N-2 zeros within the length. Then, the following conditions are sufficient for reconstruction up to a sign ambiguity.

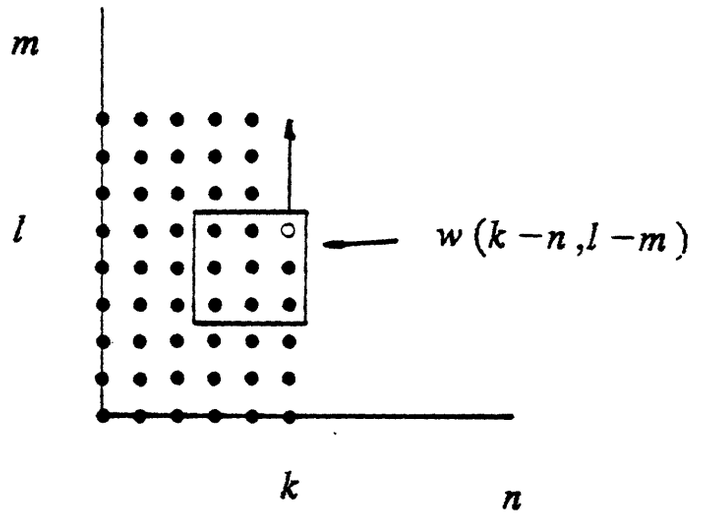
| | |
|---|---|
| w(n,m): | a) Non-zero over its NxN rectangular support |
| x(n,m): | a) Belongs to $[x(n,m)]_N$ |
| | b) Sign of one non-zero sample known |

The derivation of these conditions is analogous to the ones used for one dimensional signals. In particular, a sequential reconstruction procedures can be easily designed in a manner similar to the sequential extrapolation procedures based on the theorems in chapter 2. One such procedure for obtaining $x(n,m)$ proceeds along successive columns (rows). Suppose, in particular, that $x(n,m)$ has been computed up to the (k-1th) column and (r-1)th row (See Figure 3.4a). Then the next value can be determined from the autocorrelation of the region shown in the box along with all the known samples within the box in a manner analogous to that for one-dimensional signals. An alternative method of computation proceeds along successive lines of the form $m=-n+n'$ for some constant $n'$. This approach is illustrated in Figure 3.4b.

(a)

● Computed value

○ Next value to be computed




(b)

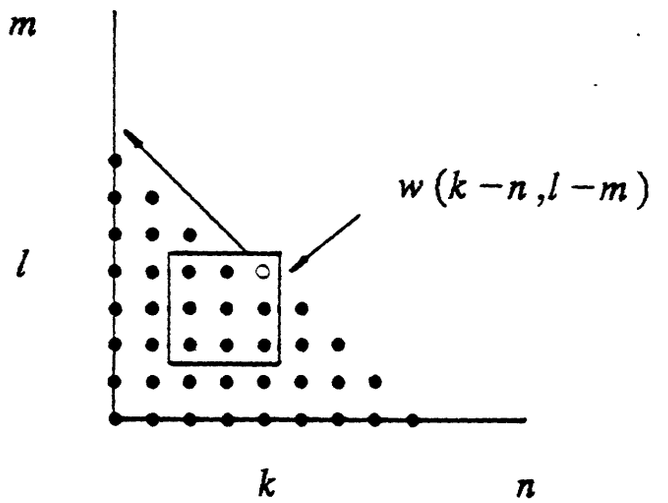



Fig. 3.4 2-D Reconstruction Procedures

# CHAPTER FOUR: SIGNAL RECONSTRUCTION FROM SHORT-TIME SPECTRAL MAGNITUDE

We have established a number of conditions under which a signal is uniquely represented by its short-time spectral magnitude. However, for such a signal representation to be practical, we need techniques that *reconstruct* a signal from its short-time spectral magnitude. In this chapter, we develop such techniques, particularly for reconstructing finite-length signals because of their importance in practical applications. In chapter 3, we introduced one such technique while developing conditions for unique signal correspondence with the short-time spectral magnitude. That technique belongs to a more general class of techniques described in section 4.1 which reconstruct the short-time sections of a signal in an order determined by their positions on the time axis. We call this the sequential extrapolation approach.

The main characteristic of sequential extrapolation techniques is that they extrapolate each short-time section using only its own spectral magnitude. A number of theorems were presented in chapter 2 for such extrapolation and used in the reconstruction procedure of chapter 3. However, in those cases only a portion of the known information was used to perform the extrapolation. In sections 4.2 and 4.3, we consider techniques which use more of the known information. In particular, we develop techniques which require the extrapolated short-time section to match the entire known information using various error criteria. This is particularly useful when the known information is not exact. For example, we will see in section 4.4 that reconstruction techniques of this chapter are less sensitive to round-off errors when compared to the extrapolation procedures of chapter 2. Furthermore, in later chapters, we will see that these techniques give better signal reconstructions when the short-time spectral magnitude is purposely modified for accomplishing signal processing tasks such as noise reduction and time-scale modification of speech.

The final section of this chapter presents an alternative reconstruction approach that is referred to as *simultaneous extrapolation*. Rather than matching the known information for each

short-time section individually, the idea is to choose all the unknown samples in a way that minimizes an error criterion defined over the entire short-time spectral magnitude. However, the resulting algorithms require the simultaneous solution of as many equations as there are samples to be reconstructed. This becomes computationally prohibitive even for average length signals in various applications. For reducing this computational complexity, one may consider extrapolating several short-time sections simultaneously, but extrapolating each such group in sequential order. Such techniques have not been implemented in this thesis. However, they are expected to perform even better than the sequential techniques of this chapter when applied to the time-scale modification and noise reduction applications of chapters 6 and 7.

### 4.1 The Sequential Extrapolation Approach

The short-time spectral magnitude of a signal $x(n)$ for a positive integer L and an analysis window $w(n)$ is given by

$$S_w(nL,\omega) = \left| \sum_{m=-\infty}^{\infty} x(m)w(nL-m)e^{-j\omega m} \right|^2 \tag{4.1}$$

We assume that $w(n)$ is a known sequence with no zero samples over its finite-length, $N_w$. Furthermore, these nonzero samples are in the region $0 \le n < N_w$. The signal $x(n)$ finite-length with no more than $N_w - 2L$ consecutive zeros separating any two nonzero samples for $L>1$. If $L=1$, at most $N_w - 2$ consecutive zeros are allowed between two nonzero samples of $x(n)$. It is also assumed that the first nonzero sample of $x(n)$ falls at $n=0$. Finally,we assume that the $L$ samples of $x(n)$ for $0 \le n < L$ are known. These assumptions are necessary for all the algorithms described in this chapter.

The sequential extrapolation approach to signal reconstruction from short-time spectral magnitude is illustrated in Figure 4.1. The $L$ known samples of $x(n)$ completely determine the short-time section corresponding to $S_w(nL,\omega)$ for $n=1$. The short-time section corresponding to $S_w(nL,\omega)$ for $n=2$ can then be extrapolated from its spectral magnitude and its known samples

$x(n)$ : Finite Length with $x(0)$ the first nonzero sample

$w(n)$ : Non-zero over $0 \le n < N_w$

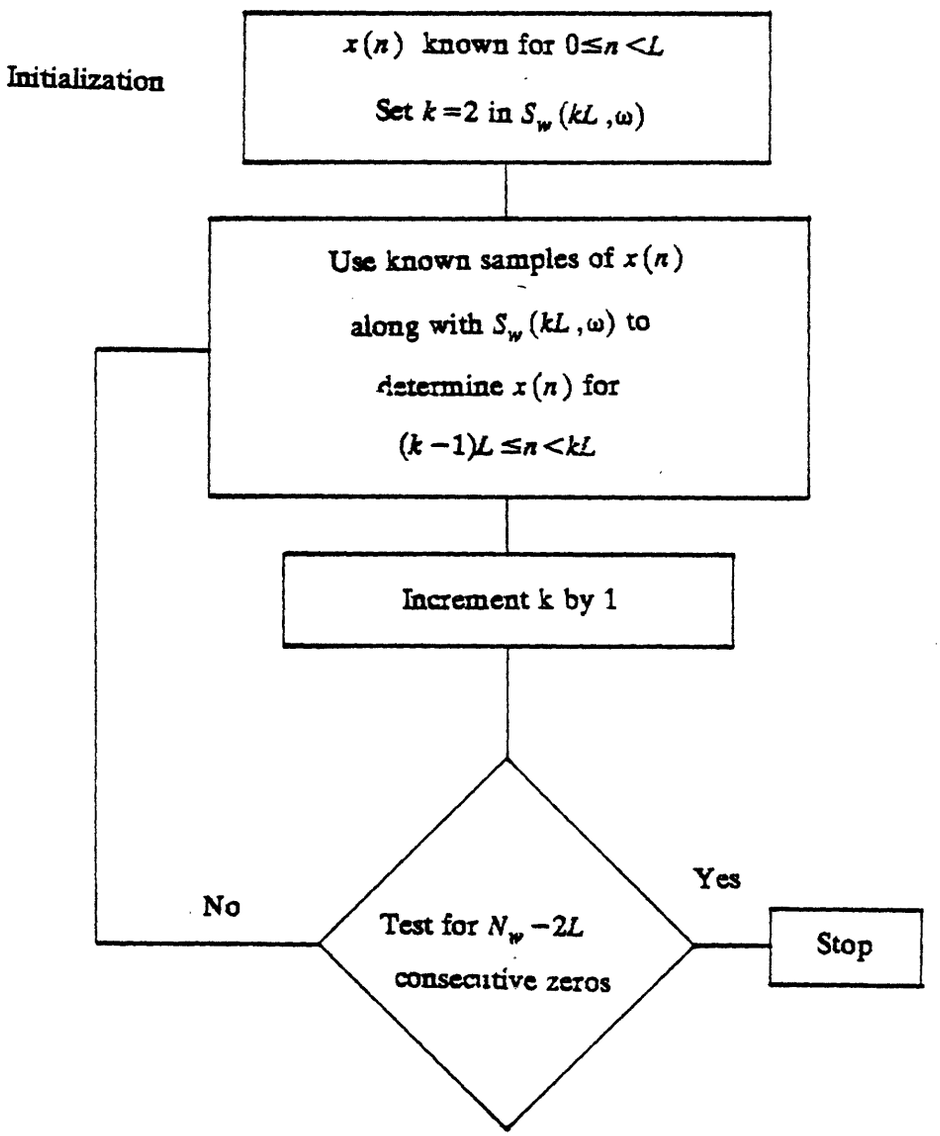



Fig. 2. Sequential Extrapolation Approach

in the region of overlap with the previously determined short-time section. This process continues as the complete extrapolation of each new short-time section makes possible the extrapolation of the next overlapping short-time section. The reconstruction stops when a short-time section is encountered for which the known samples are not sufficient to complete the extrapolation. For the conditions outlined at the beginning of this section we know from chapter 3, that the reconstruction stops only after all the non-zero short-time sections have been extrapolated. Furthermore, since the analysis window is non-zero over the length of each short-time section, dividing the short-time sections by the analysis window yields the required samples of the signal $x(n)$.

The techniques used by the proofs of the various theorems in chapter 2 can be used for accomplishing the extrapolation step of the sequential extrapolation approach. For example, in section 3.1 of chapter 3 we used techniques in the proofs of Theorems 2.1 and 2.3 for the extrapolation. In this section we apply the technique of the proof of Theorem 2.2 to the extrapolation step. In that theorem we saw that the last sample of a finite-length signal ( i.e. a sample after which the signal is always zero) can be extrapolated from the preceding samples and two appropriately chosen samples of the spectral magnitude. In the proof of that theorem, it was also shown that the two appropriate samples of the spectral magnitude can be found even if the spectral magnitude is uniformly sampled in frequency with a rate greater than $2\pi/(2N-3)$.

For each $n$, let the short-time section of $x(n)$ whose last sample falls at $n$ be denoted by $f_n(m)$. If the sample of $f_n(m)$ at $m=n$ is replaced by zero, the resulting sequence is denoted by $g_n(m)$ and its spectrum is denoted by $G_n(\omega)$. Then, Theorem 2.2 solves for the sample $x(n)$ through the quadratic equation:

$$x^2(n) + b(n,\omega)x(n) + c(n,\omega) = 0 \tag{4.2}$$

where

$$b(n,\omega) = 2Re[G_n(\omega)e^{j\omega(n+N-1)}] \tag{4.3}$$

and

$$c(n,\omega) = |G_n(\omega)|^2 - S_w(n,\omega) \tag{4.4}$$

Observe that this technique uses only two frequency samples of $S_w(n,\omega)$ for each value of n. With speech waveforms we have found that any arbitrary selection of the two frequency values generally yields two distinct quadratic equations. In particular, to reduce the computational load, a good choice is $\omega=0$ and $\omega=\pi$. In this case,(assume M even)

$$|G_n(2\pi r/M)|_{r=0} = \sum_m g_n(m) \tag{4.5}$$

$$|G_n(2\pi r/M)|_{r=M/2} = \sum_m (-1)^m g_n(m) \tag{4.6}$$

If the analysis window is rectangular, the computational load can be reduced further because $G_n(\omega)$ can be computed recursively,

$$G_n(\omega)=G_{n-1}(\omega)e^{-j\omega}-g_n(n-N_w+1)e^{-j\omega(n-N_w+1)}+x(n-1)e^{-j\omega(n-1)} \tag{4.7}$$

We still have to address the problem of synthesizing the entire reconstructed signal from its short-time sections. We have assumed that the analysis window is non-zero over its length $N_w$. It follows that we can divide each short-time section by the analysis window to obtain the corresponding samples of the reconstructed signal. Alternatively, we can select the analysis window $w(n)$ such that

$$\sum_{n=-\infty}^{\infty} w(nL-m) = 1.0 \quad \text{for } all\ m \tag{4.8}$$

In such a case, the entire signal can be reconstructed by simply adding all its short-time sections.

### 4.2 Least-Squares Sequential Extrapolation

In this section we develop a least-squares technique for the short-time extrapolation step of the sequential signal reconstruction procedure of the previous section ( See Figure 4.1 ). The major idea here is to use more information from the short-time spectral magnitude than is strictly necessary to reconstruct the signal. This makes the reconstruction algorithm more robust to errors in the short-time spectral magnitude as will be seen in section 4.4 and later chapters. The

analysis window of the short-time spectrum and therefore the corresponding short-time sections of the signal are assumed to be finite-length. From section 4.1, each short-time section is extrapolated from a set of its known samples and the spectral magnitude of the section.

Let $f(n)$ be the short-time section being extrapolated. For simplifying notation, we assume that $f(0)$ is the first non-zero sample of $f(n)$. However, the technique developed here is not affected by the particular location of the first non-zero sample. Assume that the analysis window is $N_w$ points long and thus $f(n)$ is known to be zero for $n \geq N_w$. In the sequential extrapolation approach, as outlined in Figure 4.1, the known samples of $f(n)$ are in the range $0 \leq n < M$ where $M \geq N_w / 2$ for $N_w$ even, and $M \geq (N_w - 1) / 2$ for $N_w$ odd. The problem is to extrapolate the unknown samples of $f(n)$ in the range $M \leq n < N_w$. For this we use a least-squares algorithm that minimizes

$$E = \sum_{m=-\infty}^{\infty} (r(m) - s(m))^2 \tag{4.9}$$

where $r(m)$ is the autocorrelation function obtained by taking the inverse Fourier transform of the squared spectral magnitude of $f(n)$. The function $s(m)$ represents the inverse Fourier transform of the squared spectral magnitude of the reconstructed $f(n)$. By Parseval's Theorem [1], minimizing the above expression is equivalent to minimizing the integral over the squared difference between the squared spectral magnitude of $f(n)$ and the squared spectral magnitude of the reconstructed version of $f(n)$. Both $r(m)$ and $s(m)$ are autocorrelation functions of real sequences that are at most $N_w$ samples long. It follows that $r(m)$ and $s(m)$ are even sequences of maximum duration $2N_w - 1$. Under such conditions, the minimization of (2.12) is equivalent to minimizing

$$E = \sum_{m=0}^{N_w - 1} (r(m) - s(m))^2 \tag{4.10}$$

To minimize E, we set its derivative with respect to the unknown samples of $f(n)$ to zero. If there are L unknown samples, this procedure yields a system of L simultaneous cubic equations in the $L$ unknowns. For example, if $f(N_w - 1)$ is the only unknown, we get the following cubic

equation:

$$2f^3(N_w-1)-(2r(0)-3t(0))f(N_w-1)-\sum_{m=1}^{N_w-1}(r(m)-t(m))f(N_w-1-m)=0 \qquad (4.11)$$

where $t(m)$ is the autocorrelation of the sequence obtained from $f(n)$ by setting $f(N_w-1)$ equal to zero. Generally, this equation will have two complex conjugate roots and one real root. If the signal being reconstructed is known to be real, we clearly select the real root.

For situations with more than one unknown sample, the system of simultaneous cubic equations is difficult to solve. One possible approach to simplify the equations is to neglect some of the terms in (4.10). If there are L unknowns and we neglect the terms for $0 \leq m < L$, we obtain a set of L simultaneous *linear* equations in the L unknowns. For example, if $L=1$ we obtain the following linear equation for $f(N_w-1)$.

$$f(N_w-1)=\frac{\sum_{m=1}^{N_w-1}(r(m)-t(m))f(N_w-1-m)}{\sum_{m=1}^{N_w-1}f^2(N_w-1-m)} \qquad (4.12)$$

where $t(m)$ is the autocorrelation of the sequence obtained from $f(n)$ by setting $f(N_w-1)$ equal to zero.

### 4.3 Iterative Sequential Extrapolation

In this section we develop an iterative technique for extrapolating a finite-length sequence from certain of its known samples and the spectral magnitude of the sequence. This procedure can be used for the extrapolation of each short-time section in the sequential extrapolation technique ( See Figure 4.1 ) for signal reconstruction from short-time spectral magnitude. As in the least-squares technique, the main idea here is to develop a reconstruction algorithm that uses more information than is strictly necessary to reconstruct the signal. In the following section as well as in later chapters, this algorithm proves to be very robust to errors in the short-time spectral magnitude information.

Following the notation of section 4.2, let $f(n)$ be the short-time section being extrapolated. For simplifying notation, we assume that $f(0)$ is the first non-zero sample of $f(n)$. Assuming the analysis window is $N_w$ samples long, $f(n)$ is known to be zero for $n \geq N_w$. In the sequential extrapolation approach, as outlined in Figure 4.1, the known samples of $f(n)$ are in the range $0 \leq n < M$ where $M \geq N_w / 2$ for $N_w$ even and $M \geq (N_w - 1) / 2$ for $N_w$ odd. In this section, we present an iterative technique that goes back and forth between the time and frequency domains, imposing the known constraints in each domain ( See Figure 4.2 ). The constraints imposed in the time domain are all the known samples of $f(n)$ outside the region $M \leq n < N_w$. On the other hand, in the frequency domain we impose the known spectral magnitude of $f(n)$. The goal is to have the technique converge to the correct answer for the unknown samples of $f(n)$ in the region $M \leq n < N_w$.

The problem of mathematically showing whether or not the iterative procedure outlined in Figure 4.2 converges to any kind of answer has not been addressed in this thesis. However, we have empirically observed that the procedure appears to converge to the correct answer in many cases. In other instances, however, the procedure does appear to converge but not to the samples we seek. In section 4.4 we will see that for signal reconstruction from short-time spectral magnitude, the failiure to converge to the right answer in some of the short-time sections leads to a reconstructed signal quite different from the original signal. On the other hand, for speech signals, the reconstruction is quite successful in retaining most of the perceptual quality of the original signal.

### 4.4 Reconstruction Examples

This section presents results of experiments conducted to test the reconstruction algorithms of this chapter on speech. In particular, we have tested the algorithms on the short-time spectral magnitude of the speech waveform in Figure 4.3. This waveform corresponds to the sentence "The bowl dropped from his hand", spoken by a female speaker. The processing was carried out

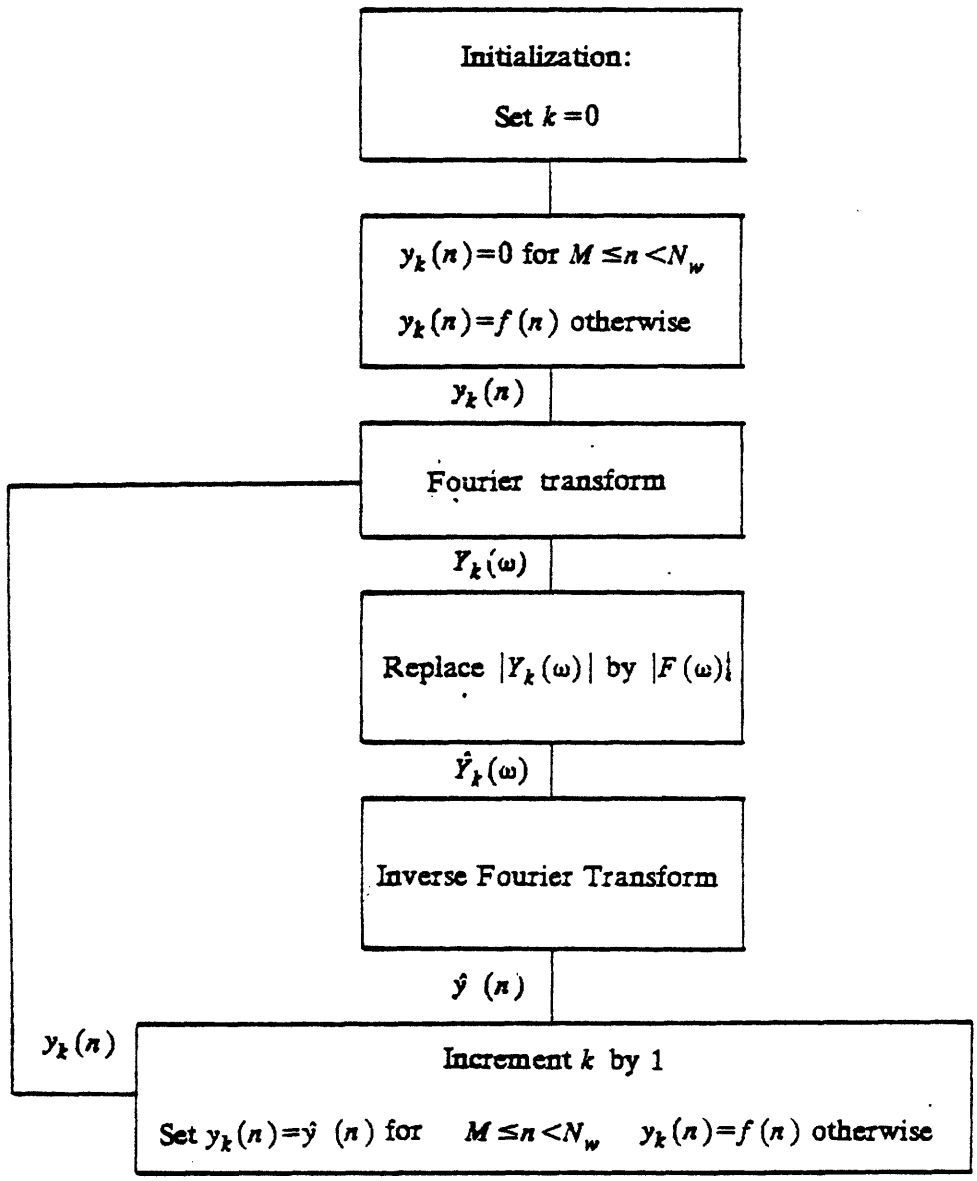



Fig. 4.2 Iterative Extrapolation

on a PDP 11/50 with floating point arithmetic. For this processing, the waveform is sampled at 10kHz and the sampling quantization rate is 12 bits.

In the first experiment, the goal is to reconstruct the signal from $S_w(n,\omega)$ using the sequential extrapolation approach based on the proof of Theorem 2.1. Specifically, the one unknown sample in each short-time section is solved for by using just one sample from the autocorrelation of the same short-time section. This approach was applied with rectangular as well as Hamming analysis windows of various lengths. Using double precision (64 bits) floating point computation, the reconstruction was successful to within the 12 bit precision of the original speech signal of Figure 4.3. For the case of a rectangular window of 32 points, the reconstruction from $S_w(n,\omega)$ is shown in Figure 4.4. Signal reconstruction was also successfully accomplished for the cases when the analysis window spacing $L$ was slightly larger than unity. In these cases, we applied the sequential extrapolation procedure based on the proof of Theorem 2.3 of chapter 2. However, when the analysis window overlap was greater than 4, this reconstruction algorithm failed very early in the signal. The failiure appears to occur due to computational errors that arise because of successive divisions by very small signal values within a short-time section.

For larger analysis window spacing, we next tried signal reconstruction using the linear version of the sequential least-squares technique of section 4.2. The analysis window is a 128-point rectangular window. Using double precision, the computation was quite successful for window spacings up to $L=30$. For example, the reconstruction for $L=20$ is shown in Figure 4.5. However, as L approaches N/2=64, there are not too many extra autocorrelation coefficients to make the computation robust. The algorithm therefore fails for values of L much higher than 40.

Finally, we applied the sequential iterative algorithm to reconstruct signals with large analysis window spacing. As indicated in section 4.3, this algorithm does not reconstruct the signal exactly. However, for speech applications the signal obtained from the iterative algorithm is perceptually close to the original signal. For example, Figure 4.6 shows the reconstruction of the speech in Figure 4.3 using the iterative reconstruction algorithm.

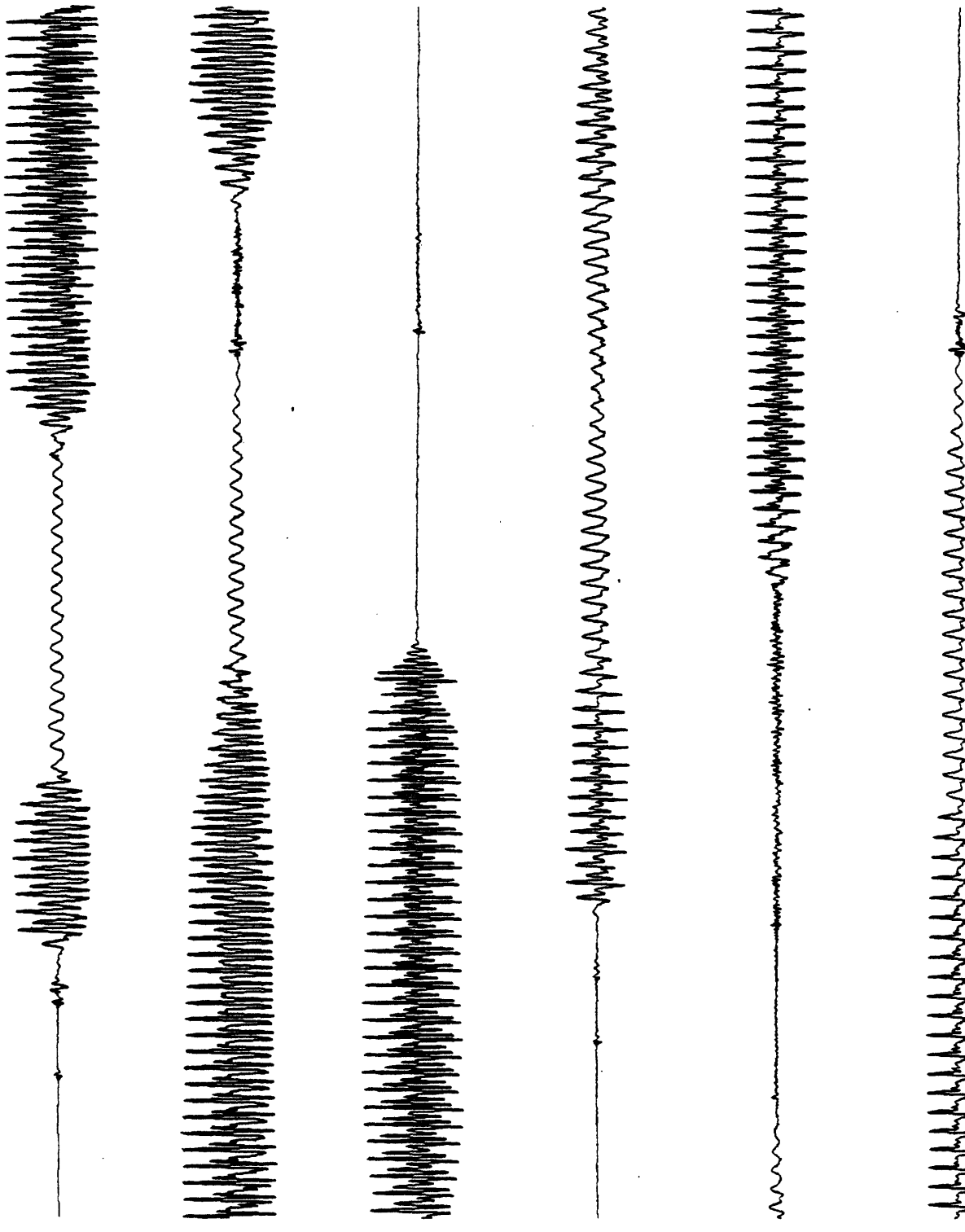


Fig. 4.3 Test Speech Waveform

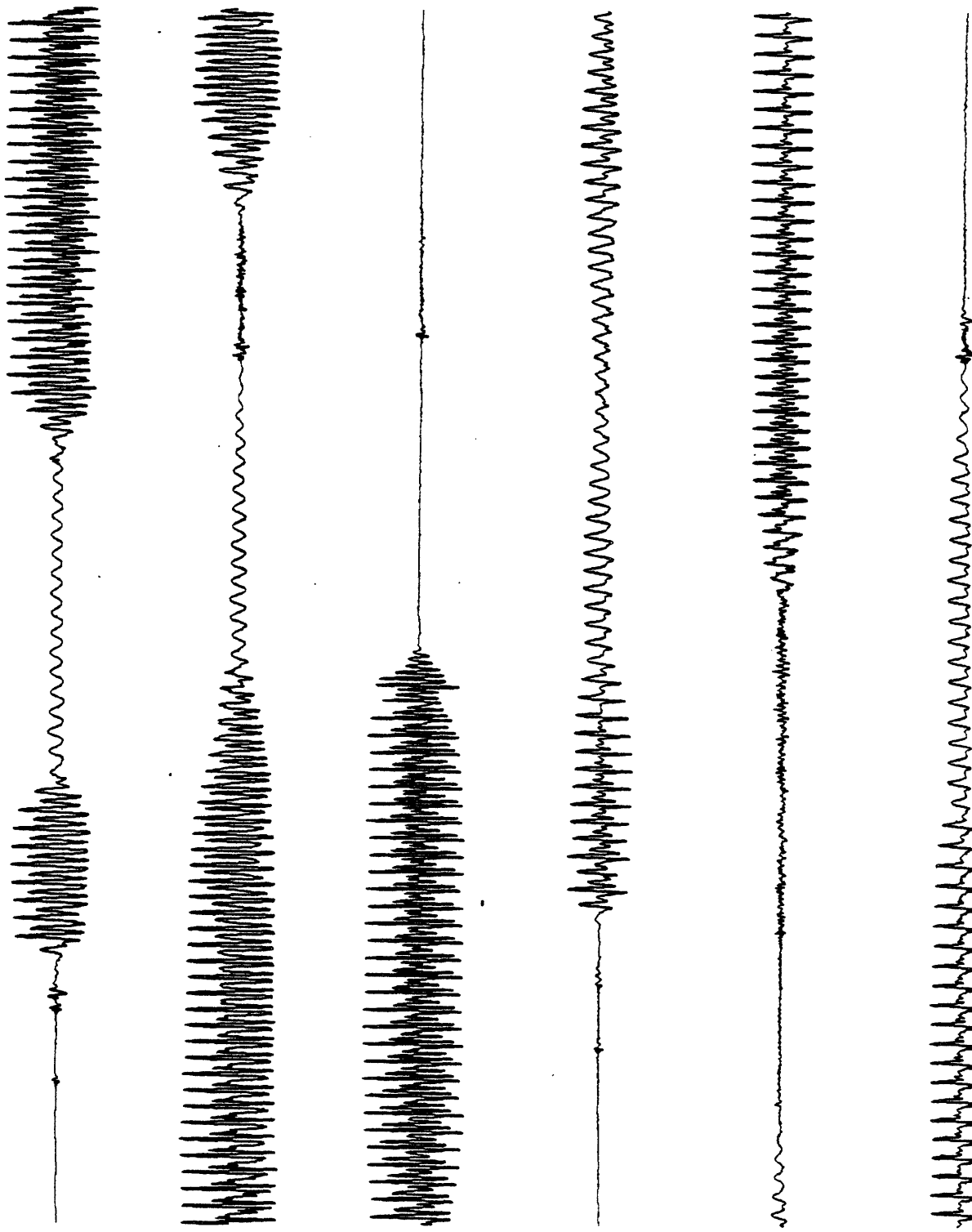


Fig. 4.4 Sequential Reconstruction Based on Theorem 2.1

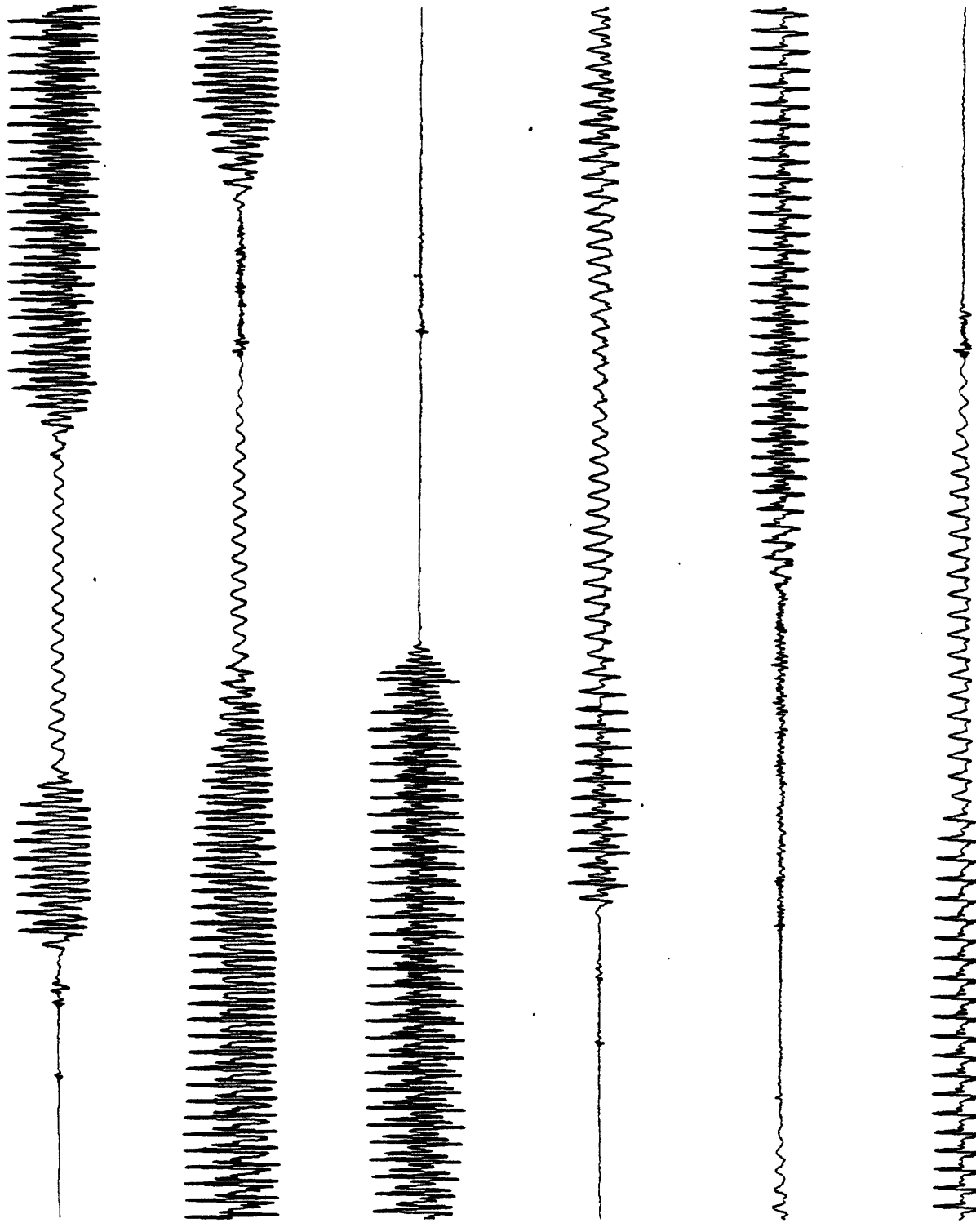


Fig. 4.5 Sequential Least-Squares Reconstruction

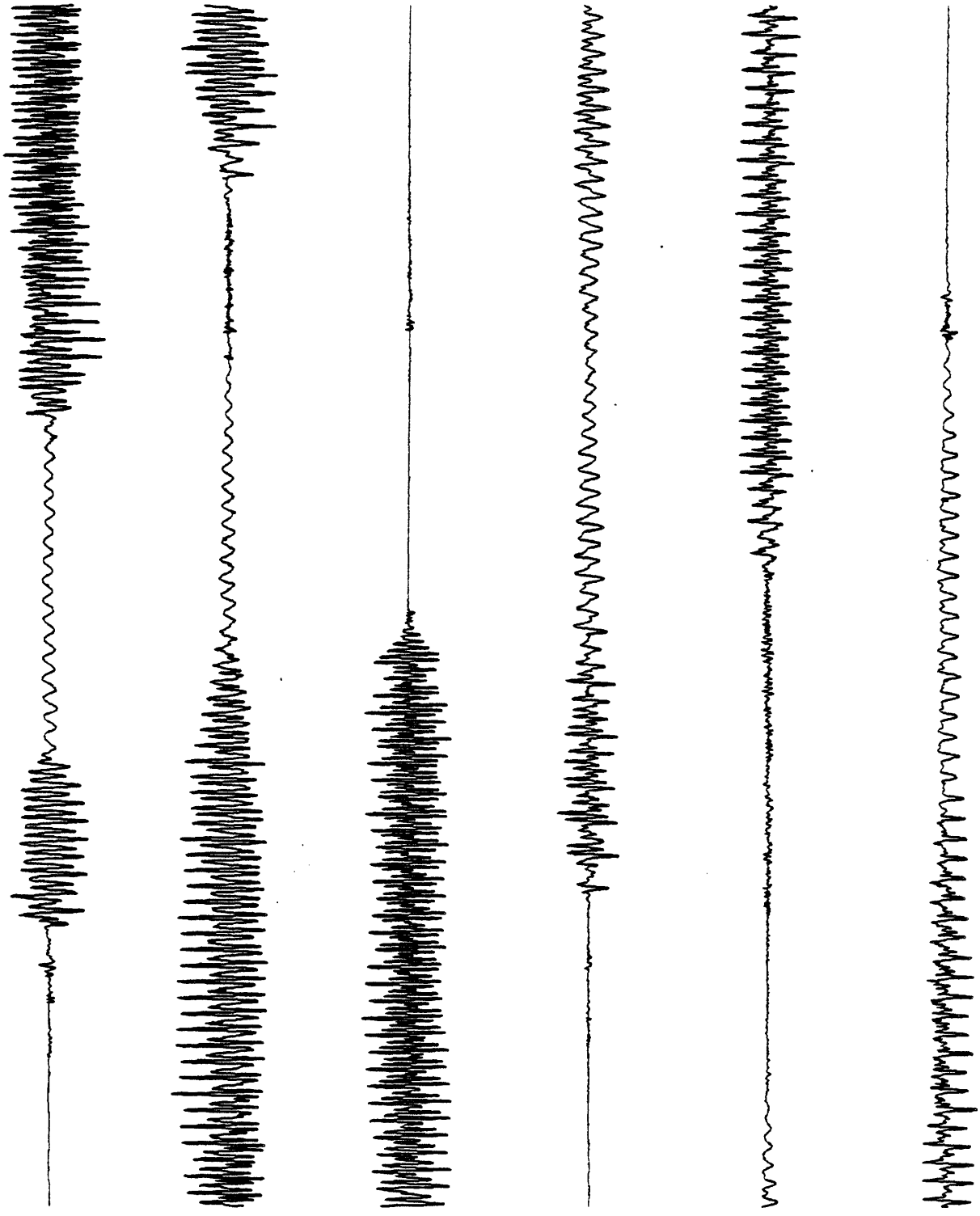


Fig. 4.6 Sequential Iterative Reconstruction

The analysis window is a rectangular window of 128 points and the window spacing $L$ is 64.

### 4.5 Simultaneous Extrapolation Approach

The emphasis in this thesis is on the sequential extrapolation algorithms of the previous sections for signal reconstruction from short-time spectral magnitude. However, other approaches can be designed for reconstructing a signal from its short-time spectral magnitude. In this section, we outline an approach which we refer to as *simultaneous extrapolation*. The main idea in this approach is to use the spectral magnitudes of several ( possibly all ) short-time sections for determining their unknown samples simultaneously. This is in contrast to the sequential extrapolation approach where each short-time section is extrapolated only on the basis of its own spectral magnitude. Of course, we have seen that the spectral magnitude of just the one section is sufficient to uniquely extrapolate the section under conditions we have been assuming in this chapter. However, in case of errors or purposeful modifications in the short-time spectral magnitude, we have seen previously that it is useful to incorporate extra information in the reconstruction procedures. For example, the least-squares and iterative techniques of the previous section used much more of the spectral magnitude of each short-time section than the techniques based on the proofs of the theorems in chapter 2. In the simultaneous extrapolation approach, we also wish to incorporate the spectral magnitude information on other short-time sections in the extrapolation of any particular short-time section.

We will illustrate the simultaneous extrapolation approach by developing it as an extension to the least-squares technique of section 4.2. The problem is to reconstruct a finite-length signal $x(n)$ from its short-time spectral magnitude, $S_w(nL,\omega)$, under the conditions developed in chapter three. In section 4.2, we showed that a set of L equations can be developed for L unknowns in each short-time-section using least-squares error criteria. These equations were either cubic or linear according to the particular error criterion used. Since the short-time sections overlap with each other, solving for those L samples in each short-time section was shown to be

sufficient to reconstruct $x(n)$. However, in section 4.2 we solved the equations separately for each short-time section. In solving those equations, we used the already determined samples of the short-time section immediately preceeding in time. Clearly, such a solution neglects the structure of the short-time spectrum that is contained in the overlap of any paticular short-time section with the short-time section that follows it in time. This structure is important to exploit when there are errors in the short-time spectral magnitude. In fact, the structure of the short-time spectrum extends over the entire time duration of the signal because of the overlap between all the short-time sections. Therefore, in the simultaneous extrapolation approach, we simultaneously solve several sets of L equations corresponding to a set of overlapping short-time sections. In the extreme, one may solve for all the sets of equations for the entire signal simultaneously. However, this would generally be computationally prohibitive.

The simultaneous extrapolation techniques have not been implemented for this thesis. However, since they exploit more of the structure of the short-time spectrum, they are expected to perform better than the sequential extrapolation techniques. On the other hand, we will see in the following chapters that the sequential techniques perform quite reasonably in various speech processing applications. The sequential techniques generally have the advantage of a simpler computational structure.

# CHAPTER FIVE: SIGNAL ESTIMATION FROM MODIFIED SHORT-TIME SPECTRAL MAGNITUDE

In many applications it is desirable to modify the short-time spectral magnitude of a signal. For example, to smooth a noisy signal, the spectral magnitudes of the short-time sections may be filtered independently according to their frequency characteristics. As discussed in section 5.1, the structure of the short-time spectrum is very sensitive to such modifications [3,5]; the modified function is generally not a valid short-time spectrum. It is of interest to estimate a signal in some reasonable way from the modified short-time spectral magnitude. For example, we would like to obtain a smoothed signal estimate from the filtered short-time spectral magnitude of a noisy signal.

In section 5.2, we consider the issues involved in applying the signal reconstruction algorithms of the previous chapter for signal estimation from modified short-time spectral magnitude. In section 5.3, we discuss certain artifacts associated with signal estimates from modified short-time spectral magnitude. Specifically, such estimates may contain abrupt changes at certain locations corresponding to the boundaries of short-time sections. The discussion includes possible ways of suppressing these artifacts. In fact, for speech processing it is found that the sequential iterative algorithm of the previous chapter is quite successful in suppressing the artifacts. Furthermore, even better performance is to be expected from simultaneous extrapolation algorithms.

## 5.1 Short-time Spectral Structure

An arbitrary function of time and frequency does not necessarily represent the short-time spectral magnitude of a signal [3,5]. This is because the definition of the short-time spectrum imposes a structure on its time and frequency variations. To see this structure, let us examine the definition of the short-time spectrum

$$X_w(nL,\omega) = \sum_m x(m)w(nL-m)\, e^{-j\omega m} \qquad (5.1)$$

This expression for $X_w(nL,\omega)$ can be viewed for a fixed $\omega$ as the convolution in $n$ of $x(n)e^{-j\omega n}$ with $w(n)$. On the other hand, for a fixed $n$, we can view $X_w(nL,\omega)$ as a convolution in frequency through the following equivalent definition [5]

$$X_w(nL,\omega) = \int_{-\pi}^{\pi} X(\psi)W(\psi-\omega)e^{j(\psi-\omega)nL}d\psi$$

where $X(\omega)$ and $W(\omega)$ are the Fourier transforms of $x(n)$ and $w(n)$ respectively.

Another illustration of the structure in the short-time spectrum is obtained from the interpretation of $X_w(nL,\omega)$ as a collection of Fourier transforms obtained as window $w(-n)$ slides across $x(n)$. In particular, consider the case when the analysis window $w(n)$ is unity over $0 \le n < N$ and zero otherwise. Then $X_w(n'L,\omega)$ for a particular $n = n'$ is the Fourier transform of the portion of $x(n)$ over $n'L - N < n \le n'L$. Similarly, $X_w((n'+1)L,\omega)$ is the Fourier transform of the portion of $x(n)$ over $(n'+1)L - N < n \le (n'+1)L$. Then, the inverse Fourier transforms of $X_w(n'L,\omega)$ and $X((n'+1)L,\omega)$ are the same over $(n'+1)L - N < n \le n'L$. Clearly, any two arbitrary Fourier transforms are unlikely to have such a property. Similarly, with two arbitrary Fourier transform magnitudes, it is unlikely that any of the various sequences corresponding to one Fourier transform magnitude overlaps in the desired way with any of the sequences corresponding to the other Fourier transform magnitude.

### 5.2 Signal Estimation Algorithms

As in previous chapters, we define the short-time spectral magnitude of a sequence $x(n)$ by

$$S_w(nL,\omega) = \left| \sum_{m=-\infty}^{\infty} x(m)w(nL-m) \right|^2$$

In chapters 3 and 4, we found sufficient conditions and corresponding algorithms for reconstructing the signal $x(n)$ from $S_w(nL,\omega)$. In this section, we consider using those reconstruction algorithms for obtaining signal estimates from modified versions of $S_w(nL,\omega)$. We will denote any modified versions of $S_w(nL,\omega)$ by $M_w(nL,\omega)$.

From section 5.1, we know that $M_w(nL,\omega)$ is generally not a valid short-time spectral magnitude. Consequently, any algorithm that relies critically on the validity of the short-time spectral magnitude performs poorly. This is the case, for example, in the sequential extrapolation algorithms that use the extrapolation techniques in the proofs of theorems 2.1 and 2.3 of chapter two. In those algorithms, only a part of the autocorrelation of each short-time section is used for extrapolation of the unknown samples. This ensures that the extrapolated samples of each short-time section are consistent with just a portion of that section's autocorrelation. When the spectral magnitude is unmodified, the remaining portion of the section's autocorrelation is also consistent with the extrapolation. However, if the spectral magnitude of the short-time section is modified, there is no guarantee that the extrapolated samples will be consistent with the unused portion of the autocorrelation. It is therefore desirable in such cases to use algorithms that extrapolate each short-time section in a way that ensures as much consistency as possible with the given autocorrelation. The least-squares and iterative extrapolation algorithms of the previous chapter were designed for this purpose. Therefore, we will use the same techniques for signal estimation from modified short-time spectral magnitude.

Another difficulty encountered in applying the techniques of chapter 4 for signal estimation from $M_w(nL,\omega)$ is that those techniques require a-priori knowledge of some initial samples of the signal estimate. Our approach is to use a reasonable guess for those initial samples. For example, the initial samples of the unprocessed signal, if known, may be used. The techniques of chapter 4 were designed in order to be not too sensitive to modifications in the known information such as the initial samples. We find this to be the case in speech applications such as those discussed in chapters 6 and 7. The effect of errors in the initial signal samples as well as errors in the short-time structure of $M_w(nL,\omega)$ is discussed in the next section.

### 5.3 Short-Time Boundary Artifacts

As discussed in section 5.1, processing the short-time spectral magnitude results in a function that in general does not correspond to the short-time spectral magnitude of any signal. Furthermore, the algorithms of chapter 4 for signal reconstruction from the magnitude of the short-time spectrum require a-priori knowledge of a certain number of initial signal samples. However, in most applications, such information is impossible to obtain accurately. As a result of such inaccuracies, certain artifacts arise at the boundaries of short-time sections in signal estimates from the modified short-time spectral magnitude. In particular, we find abrupt changes in signal value at certain locations corresponding to the boundaries of short-time sections. In this section, we will study the origin of these artifacts and discuss ways of avoiding them. In fact, in chapters 6 and 7 we will see that the sequential iterative algorithm of chapter 4 performs quite well in this regard for speech processing applications.

To investigate the cause for such artifacts, let us consider a discrete-time signal $x(n)$ with short-time spectrum $X_w(nL,\omega)$. We now replace the short-time spectral phase of $X_w(nL,\omega)$ by some other arbitrarily selected phase function. In particular, consider any two overlapping short-time sections of $x(n)$. When the spectral phases of these sections are replaced by some other phase functions, the time distribution of the two short-time sections changes. This distribution may range anywhere between minimum phase energy ( concentrated near smaller values of $n$ in the short-time section ) to maximum phase energy ( concentrated near larger values of $n$ in the short-time section ). Consequently, if the new phase functions were selected arbitrarily, there is no guarantee that at the boundaries of the short-time sections their time distributions will match. For example, in Figure 5.1 we show the test waveform of Figure 4.3 with its short-time spectral phase replaced by zero. The analysis window is a 128-point rectangular window. Clearly, there are very abrupt transitions within this signal that were not present in the original signal of Figure 4.3. Furthermore, the abrupt changes occur periodically and they actually correspond to boundaries of short-time sections.

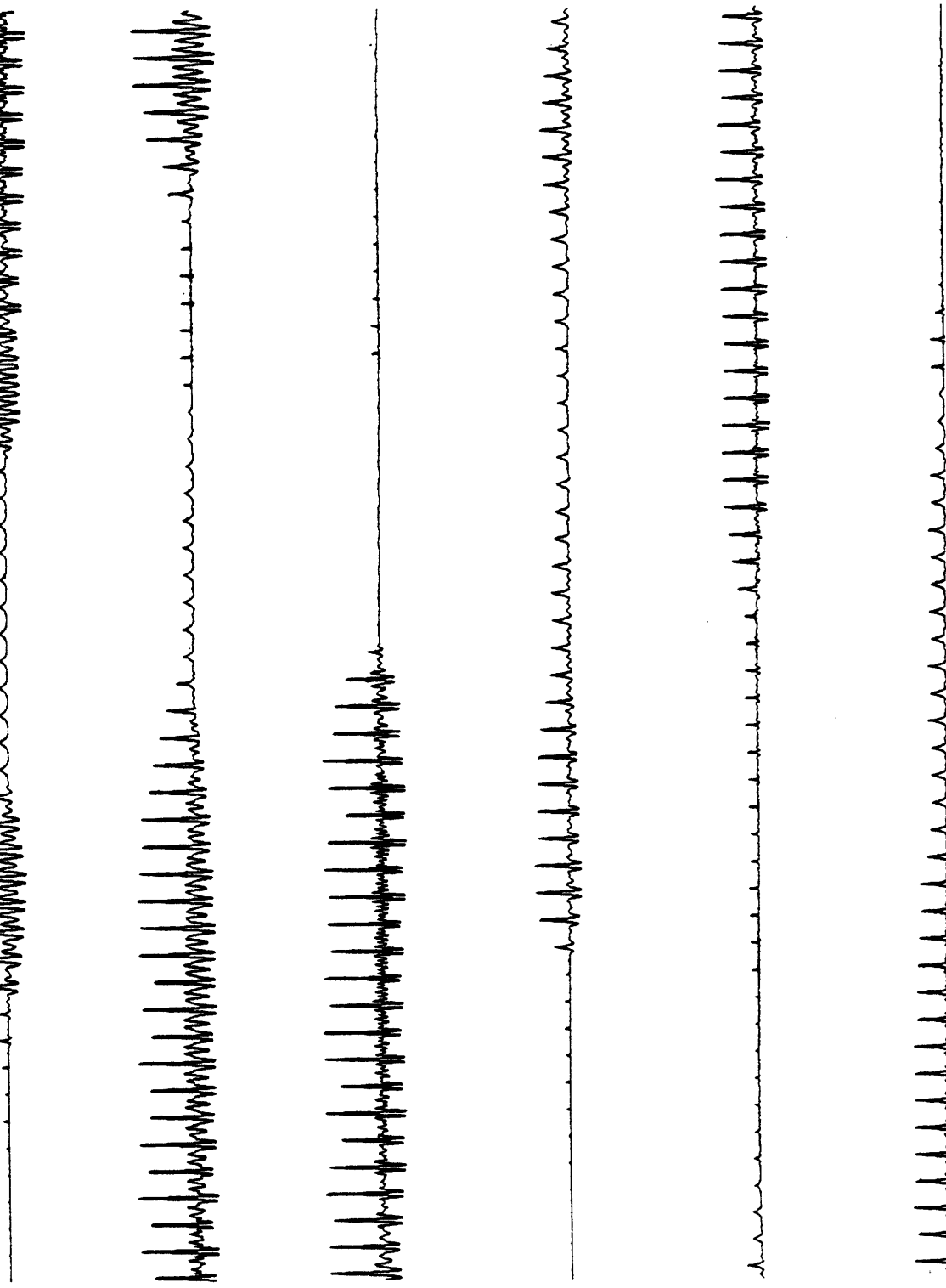


Fig. 5.1 Effect of Zero Short-Time Spectral Phase

We have thus established that given any short-time spectral magnitude, if a phase function is selected for it arbitrarily, this will give rise to short-time boundary artifacts. It is therefore important that any algorithm for signal estimation from short-time spectral magnitude should attempt to select the phase function in a way that minimizes the short-time boundary artifacts. Clearly, if there is no error in the input to a reconstruction algorithm and if there is no computational error, the algorithm will select the unique phase function corresponding to the short-time spectral magnitude. There will therefore be no short-time boundary artifacts.

We have seen that in various short-time spectral processing applications, the processed short-time spectral magnitude does not correspond to the short-time spectral magnitude of any signal. Furthermore, the initial samples of the processed signal are usually impossible to determine exactly. Consequently, the time distribution of the short-time section is generally incorrect. It is important that any algorithm for signal estimation should choose the remaining short-time sections in a way that minimizes the short-time boundary artifacts.

In this thesis we have found that the sequential iterative algorithm significantly suppresses the short-time boundary artifacts for speech applications. However, the algorithm is limited by its sequential nature. Specifically,the alignment of short-time sections is accomplished by considering pairs of short-time sections independently and in an order determined by their location on the time axis. Thus, given the distribution of the one short-time section, the distribution of the short-time section immediately following it is determined. In aligning the two sections, no information on the other short-time sections is incorporated. Thus, the minimization of short-time boundary artifacts is accomplished only over localized regions of the short-time spectral magnitude. It is expected that the performance of sequential algorithms can be improved upon by using simultaneous extrapolation algorithms.

Although sequential extrapolation algorithms can be designed to significantly suppress the short-time boundary artifacts, they often do not yield the type of time distribution in the short-time sections that is consistent with that of the unprocessed signal. For example, when noise

reduction processing is applied to speech in chapter 7, we find that the detailed shapes within short-time sections of the processed signal are significantly different from those of the original undegraded signal. However, the sections do preserve such important attributes as the periodicity of the voiced sections. In fact, perceptually we find that the processed speech is almost identical to the original undegraded speech. It is concluded that although the actual short-time spectral phase is not important for speech perception, it is essential that the phase be chosen so as to avoid short-time boundary artifacts. Finally, it must be observed that the significant change in detailed short-time signal shapes seems largely to be a consequence of the sequential character of the algorithms we have implemented for this thesis. For applications where such change in detailed shape is not acceptable, it is suggested that simultaneous extrapolation algorithms be used.

# CHAPTER SIX: TIME-SCALE MODIFICATION

## 6.1 Introduction

Signal estimation from short-time spectral magnitude is applied in this chapter to the problem of time-scale modification of speech. Time-scale modification procedures aim at maintaining the perceptual quality of the original speech while changing the apparent rate of articulation. This is essentially equivalent to preserving the instantaneous frequency locations while changing their rate of change in time. Efficient procedures for such processing have a number of important applications. Controls for time-scale modification on a tape recorder, for example, would allow users to pace the playback according to their own convenience. Thus, sections of the recording can be scanned over rapidly or played slowly depending on the listeners needs. This gives the recorded medium additional flexibility that previously only printed text could provide. For the blind, this is a particularly encouraging prospect, since even normal recorded speech offers a "reading rate" that is typically 2 to 3 times that for Braille [12].

Efficient time-scale modification of speech is also applicable in the areas of signal coding/decoding and speech recognition systems. In the former case, speech may be time-compressed at the coding stage to reduce the data rate and then appropriately time-expanded at the decoding stage. In speech recognition systems, time-scale modification could be used to normalize the duration of utterances before applying recognition algorithms.

A simple time-scaling that replaces $x(n)$ by $x(an)$ introduces significant degradation for the above applications. For example, such a scaling is obtained when a recording is played back faster than the original recording rate. The resulting " Mickey Mouse " effect is amusing but distorts most of the original perceptual characteristics of the speech. This degradation is caused by changes in the pitch of voiced sections and by the shifting of vocal tract resonances (formants). Thus, more sophisticated time-scale modification techniques are required to keep the pitch and formant locations as invariant as possible.

Many of the techniques devised in the past for time-scale modification of speech are based on an approach first used in a technique known as Fairbank's method [13]. This approach and the various techniques based on it are described in section 6.2. An alternative approach, based on the Phase Vocoder [14], was used by Portnoff [7] to develop a very successful time-scale modification technique. This approach, outlined in section 6.3, processes both the magnitude and the phase of the short-time spectrum. The resulting time-scale modification is generally considered to be of acceptable quality [7,15] for many applications. However, from a practical point of view this technique has the major disadvantage of a complicated computational structure.

The time-scale modification procedure developed in section 6.4 combines the techniques for signal estimation from short-time spectral magnitude with the basic idea behind Fairbank's method. The resulting time-scale modifications are found to be comparable to those achieved with the Phase Vocoder technique developed by Portnoff. However, the technique proposed in section 6.4 has a much simpler computational structure and can be used to design practical time-scale modification systems.

### 6.2 Fairbanks Approach

Fairbanks' approach [13] to time-scale modification mainly consists of discarding or replicating short-time sections of the speech depending upon whether time compression or time expansion is desired. Provided the short-time sections are short enough, portions of all the phonemes [2] are preserved but their durations are changed. Furthermore, the pitch and formants in the voiced sections are retained. However, a major difficulty is that the transitions between the short-time sections is not smooth. These sharp transitions introduce a periodic degradation at the frame rate, perceptually perceived as a "burbling" distortion [12].

Various strategies based on pitch detection have been designed to overcome the smooth transition problem in Fairbanks approach. These include the pitch-synchronous technique

[16,18] and the pseudo-pitch-synchronous technique [17]. The pitch-synchronous technique chooses the short-time sections so that they correspond to multiples of the pitch period in voiced sections. This ensures a smoother transition between adjacent short-time sections. In order to select such short-time sections it is necessary to first apply pitch marking algorithms. Any errors in the pitch marking introduces objectionable artifacts in the speech [18]. This is particularly a problem with noisy speech, since in that case pitch marking algorithms have very poor performance. The pseudo-pitch-synchronous technique attempts to avoid this problem by requiring only a rough estimate of the pitch periods. The algorithm repeats or discards sections of the speech equal in length to the average pitch period, then smooths together the edges of the remaining sections. This algorithm has better performance than the pitch-synchronous method, particularly in the presence of noise.

The desire to obtain a time-scale modification technique that is not dependent on pitch extraction and voiced/unvoiced decisions prompted the work on Phase Vocoder based techniques [7,12]. This led to the development of a very successful technique described in the next section.

### 6.3 Phase Vocoder Approach

An alternative to the Fairbanks time-scale modification approach is to use classical vocoder techniques. The speech is coded in the vocoder analysis stage with time-dependent parameters. The idea is to appropriately time-scale those parameters before the resynthesis of the speech signal. However, most of the classical vocoder techniques require voiced/unvoced decisions and pitch extraction. Thus, any time-scale modification technique based on such vocoders would suffer the same kind of pitch detection artifacts as those found in pitch-synchronous refinements of the Fairbanks approach. One exception to this is the Phase Vocoder [14,2,19,20]. This vocoder uses the short-time spectrum (both magnitude and phase) for representing the speech signal. Furthermore, it does not require voiced/unvoiced decisions or any pitch extraction procedures.

Portnoff [7,12] has developed a very successful time-scale modification technique based on the Phase Vocoder. Toward this, he first developed a mathematical representation for the sampled speech signal based on the usual model for speech production. This representation is used as the basis for a definition of rate-changed speech. Finally, Portnoff showed how the short-time spectral representation used in the Phase Vocoder provides a mechanism for modifying the speech time-scale. In the remainder of this section we briefly outline Portnoff's procedure for obtaining time-scale modified speech from the short-time spectrum.

Let $x(n)$ be the discrete time signal which is to be time-scale modified by a factor of $\beta$. We restrict $\beta$ to be a rational number. This is not a practical restriction since any real number can be approximated by a rational number with arbitrary precision. In the phase vocoder approach $x(n)$ is first transformed into its short-time spectrum for M frequency locations chosen appropriately to avoid aliasing [7,12]. In the expression below for the short-time spectrum we assume that $\omega$ is evaluated at just those M frequency values.

$$X_w(n\beta,\omega) = \sum_{m=-\infty}^{\infty} x(m)\, w(n\beta-m)\, e^{-j\omega m}$$

If $\beta$ is not an integer, this computation is accomplished through an interpolating procedure [2]. The next step is to estimate the unwrapped phase [1] of $X_w(n\beta,\omega)$. A good description of the phase estimation process is given in [15]. For time-scale modification, we want the pitch frequency locations to remain the same but their time variation to change by the factor $\beta$. Portnoff showed that this can be accomplished by dividing the unwrapped phase of $X_w(n\beta,\omega)$ by $\beta$. After this division, the phase vocoder approach synthesizes the time-scale modified speech from the processed short-time spectrum.

A major problem with the phase vocoder approach is its computational complexity. It generally requires sophisticated indexing and rather large memory space for its implementation. For example, Portnoff had to introduce significant memory management to implement the technique on a PDP 11/50. On the other hand, Holtzman [15] has developed an alternative implementation that significantly reduces the memory requirements but at the expense of greater programming

complexity. The technique presented in the next section is based on iterative signal estimation from the short-time spectral magnitude. Compared to the phase vocoder approach, that technique has considerable computational advantages. Furthermore, it appears to have comparable performance in terms of the quality of the time-scale modified speech.

### 6.4 Short-Time Spectral Magnitude Approach

This section describes a technique for time-scale modification of speech using signal estimation from modified short-time spectral magnitude. The performance of this technique appears to be comparable to the quality achieved by Portnoff's technique. On the other hand, as noted before, this scheme is computationally much simpler and requires very little memory.

The basic idea for the technique in this section is similar to the Fairbanks approach where various short-time sections are discarded or replicated according to whether compression or expansion is desired. However, the difference is that in this case the spectral magnitudes of various short-time sections are discarded or repeated in the short-time spectral magnitude of the speech. In the Fairbank approach, the remaining short-time sections are merely concatenated with each other, possibly taking into account any pitch information that may be available. In contrast, the strategy here is to consider the set of spectral magnitudes of the remaining short-time sections as representing a modified short-time spectral magnitude. The techniques of signal estimation from modified short-time spectral magnitudes are then used to obtain the time-scale modified signal. As discussed in chapter 5, such signal estimation techniques can be designed such that they significantly suppress the short-time boundary artifacts in the signal estimate. This results in the kind of alignment between short-time sections that is attempted by the pitch-synchronization implementations of the Fairbank's approach. However, in contrast to those techniques, this approach does not depend on pitch detection or pitch marking algorithms. It thus tends to be much more robust to noise in the speech.

For the signal estimation component of our approach to time-scale modification, we have found the sequential iterative technique of chapter 4 to be particularly attractive for speech. It has the advantage of a simple computational structure along with a high quality performance in the tests we have conducted. As an example, consider time compression of the test sentence in this thesis: "The bowl dropped from his hand". The entire waveform of the sentence is shown in Figure 4.3. The short-time spectral magnitude of the waveform is computed with a 128-point Hamming window and a window spacing $L$ of 32. Every other spectral magnitude is discarded in order to obtain a 2:1 time compression. The iterative algorithm of chapter 4 then yields the waveform shown in Figure 6.1. Clearly, the duration of the sentence has been cut by half. Furthermore, the pitch of the various segments is the same as in Figure 4.3 and there are very few short-time boundary artifacts.

As another example consider time expansion of the test sentence. In this case the short-time spectral magnitude is computed with a 128 point rectangular window and window spacing $L$ of 32. To obtain the time-scale modified signal estimate, this short-time spectral magnitude is considered to correspond to a window spacing $L$ of 64. Clearly, this results in a signal that is twice as long as the original signal. The result obtained using the iterative technique is shown in Figure 6.2. Once again, the pitch of the various segments is preserved and there are very few short-time boundary artifacts. The quality of the resulting speech is comparable to that obtained with Portnoff's technique.

Finally, note that different rates of expansion and compression can be obtained than those used in the examples above. For example, discarding two out of every three short-time segments results in time compression by a factor of 3. On the other hand, if one out of every three segments is discarded, the processed signal is two thirds as long as the original speech. Similarly, a time expansion by a factor of three can be obtained by computing the short-time spectral magnitude at a time-sampling rate three times higher than the maximum rate -- half the analysis window length. The result is then processed as if it were sampled at the maximum rate. Clearly,

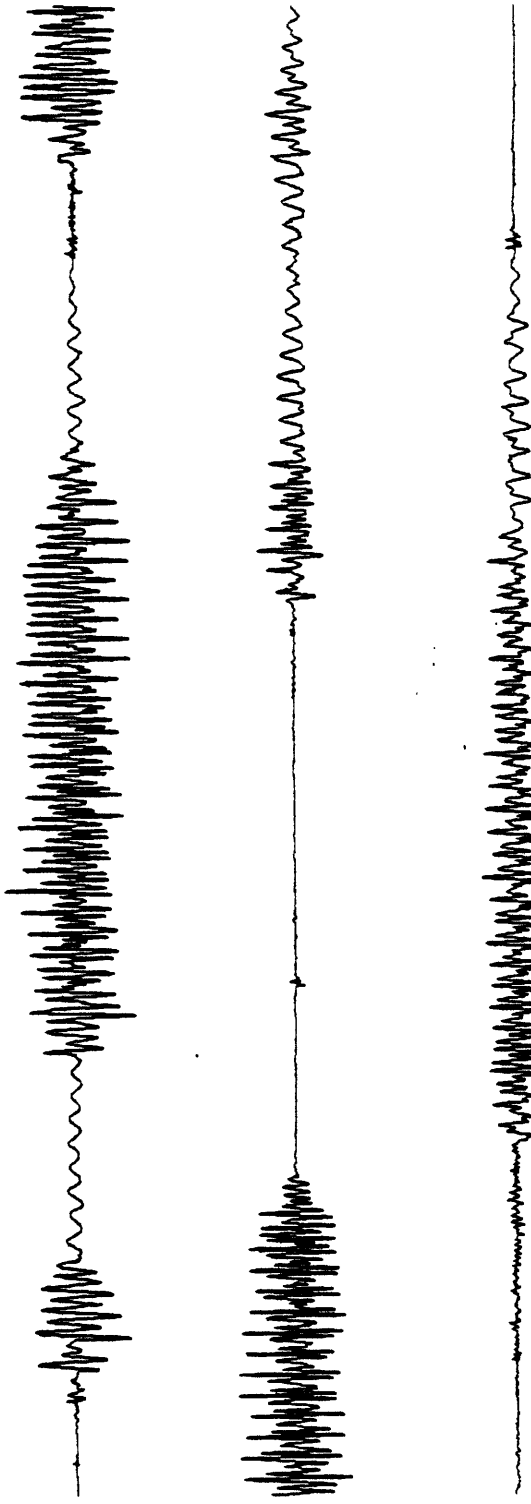


Fig. 6.1 Time-Scale Compression of 2:1

Fig. 6.2 Time-Scale Expansion of 1:2

many such strategies can be devised for discarding or adding new short-time segments for various time-scale modification rates.

# CHAPTER SEVEN: NOISE REDUCTION

The problem of noise reduction arises in numerous signal processing contexts. The reduction of noise in a signal may either be the final goal or an intermediate step. For example, speech communication between a pilot and an air traffic tower is typically degraded by background noise. In such a case, the reduction of noise is the final signal processing step for ensuring clear communication. On the other hand, a radar image may be processed for noise reduction as only a preliminary step for target detection.In all such cases, the noise reduction is an essential element of the entire system.

In this chapter, we consider the processing of signal-independent additive noise. Many problems in speech and image processing fall into this category. Furthermore, problems involving multiplicative or convolutional noise can be converted into an additive noise problem by a homomorphic transformation [1,21]. Sometimes, even signal dependent noise may be converted to signal independent additive noise. For example, pseudo-noise techniques [25] have been used for such a transformation in the quantization noise associated with PCM signal coding.

In the problem of noise filtering of speech or image data, it is often preferable to use short-time spectral processing [6,22]. This is primarily because long-time filtering tends to smooth out local variations that are often important attributes of the signal. In contrast, short-time spectral processing attempts to preserve such attributes and is therefore generally considered a better alternative. Section 7.1 discusses in greater detail the advantages of short-time spectral processing for noise reduction. A number of short-time spectral processing techniques exist for noise reduction in speech and images. The spectral subtraction technique has been shown [6] to have good performance and relatively simple implementation. We describe the standard version of this technique in section 7.2. One characteristic of the standard spectral subtraction technique and other short-time spectral processing techniques for noise reduction is that the short-time spectral phase of the noisy signal is left unprocessed.

In section 7.3, we present a modification for the short-time spectral subtraction technique, in which signal estimation from the processed short-time spectral magnitude is used to obtain a processed version of the short-time spectral phase. This modification is also applicable to the other short-time spectral processing techniques mentioned in section 7.1. We find that the performance of the modified short-time spectral subtraction technique is comparable to that of standard short-time spectral subtraction. However, unlike the standard technique, the modified technique does not require the short-time spectral phase of the noisy signal.

Both the standard spectral subtraction technique and its modified version introduced in this thesis produce certain undesirable artifacts in the processed signal. In section 7.4, possible causes for those artifacts are discussed and techniques are developed for suppresseing them. These artifact suppression techniques can be applied to both the standard and modified versions of short-time spectral subtraction.

### 7.1 Short-Time Spectral Processing Techniques

To establish a mathematical framework for our discussion, let $s(n)$ denote the discrete-time signal we want to estimate from another signal $x(n)$ which is the sum of $s(n)$ and a noise signal $e(n)$. It is assumed that $e(n)$ is a sample sequence of a stationary stochastic process with known spectrum $P_e(\omega)$. Although for convenience we use the notation of one dimensional signals, the entire discussion in this section is also applicable to multidimensional signals [24].

A number of classical techniques exist for filtering stationary stochastic processes [23]. In particular, a problem that has been widely considered is that of filtering additive stationary noise from stationary stochastic processes. This has resulted in the well-known non-causal Wiener filter and numerous other related techniques. It is therefore not surprising that such techniques have been considered for the noise reduction problem in numerous application areas. The most successful applications are those where the desired signal $s(n)$ can be adequately modelled as a sample sequence of a stationary stochastic process with a known spectrum. However, in areas such as

speech and image processing such a model is generally inadequate. For example, speech is commonly modelled as the output of a time-varying linear system driven by either white noise or quasi-periodic pulses [2]. Therefore, it is inappropriate to consider the output of such a system as a stationary process.

The linear system mentioned above for the modelling of speech signals is *slowly* time-varying. This has lead investigators to consider short-time sections of speech signals to have stationary spectral characteristics. This approximation has been used successfully in a variety of engineering contexts, including noise reduction, speech synthesis, and bandwidth compression. Unfortunately, there is no comparable model for images. However, the inspection of any typical image shows many rapidly space-varying characteristics. In fact, much of the information in images lies in sharp changes such as those at the boundaries of objects. These types of characteristics generally render useless any attempt at modelling images as outputs of stationary stochastic systems. However, except near sharp changes such as those at object boundaries, short-space ( 2-D equivalent of short-time ) modelling of images with stationary processes has been relatively successful [22,24]. Furthermore, signal processing based on such short-space modelling often does not appreciably degrade object boundaries and other sharp details.

The short-time spectrum has proved to be particularly convenient for the short-time processing of speech and images. The central idea is to process the spectrum of each short-time section separately. Since the signals are assumed to be stationary at the short-time level, classical noise reduction techniques based on spectral filtering can be used. A common characteristic of all these techniques is that they yield zero-phase filters. Thus, the overall processing affects only the short-time spectral magnitude.

### 7.2 Standard Short-Time Spectral Subtraction

A number of short-time spectral processing techniques have been developed over the years for the reduction of additive noise in speech and image signals [6,24]. The performance of such

techniques is generally of the same order as that of a technique known as short-time ( or short-space for images ) spectral subtraction. However, short-time spectral subtraction offers the advantage of simpler implementation. In this section, we review the short-time spectral subtraction technique as it is generally implemented. As indicated in the previous section, such an implementation of short-time spectral processing retains the short-time spectral phase of the noisy signal. In section 7.3, on the other hand, we will use the theory and techniques of this thesis to develop a different implementation of short-time spectral subtraction. That implementation has the property that it estimates a processed short-time spectral phase from the processed short-time spectral magnitude.

We first review the classical spectral subtraction procedure for processing stationary random signals without utilizing short-time techniques. Let $s(n)$ be the stationary random signal we wish to estimate from another signal $x(n)$ which is the sum of $s(n)$ and uncorrelated noise $e(n)$. Assume that the power spectral density $P_e(\omega)$ of $e(n)$ is known. The power spectral density $P_s(\omega)$ of $s(n)$ is then estimated from the observations of $x(n)$ and the known $P_e(\omega)$. Specifically, since $x(n)$ is the sum of $s(n)$ and the uncorrelated $e(n)$, it follows that

$$P_x(\omega) = P_s(\omega) + P_e(\omega) \tag{7.1}$$

A reasonable estimate for $P_s(\omega)$ is obtained by subtracting the known spectrum $P_e(\omega)$ from an estimate of $P_x(\omega)$. The estimate of $P_x(\omega)$ is usually computed as the magnitude squared of the Fourier transform of the observed $x(n)$. The subtraction process sometimes gives negative values in the estimate of $P_s(\omega)$. The most common approach for such situations is to replace the negative values by zero [6]. Finally, the square root of the estimate of $P_s(\omega)$ is used as the Fourier transform magnitude for the estimate of the signal $s(n)$. This estimate of the Fourier transform magnitude is then combined with the Fourier transform phase of the noisy signal $x(n)$ to yield the standard spectral subtraction estimate of the desired signal $s(n)$. It has been shown [6] that this procedure implicitly performs a type of parametric Wiener filtering.

The implementation of the standard spectral subtraction technique using the short-time

spectrum is relatively straightforward. The basic idea is to consider each short time section as an observation of a stationary stochastic process and apply the spectral subtraction procedure separately to each section. Thus, for example, if $x(n)$ is the signal corrupted by additive noise $e(n)$ and $S_w(nL,\omega)$ is the short-time spectral magnitude of $x(n)$, the spectral subtraction procedure yields the following function:

$$\hat{S}_w(nL,\omega) = \begin{cases} S_w(nL,\omega) - \alpha P_e(\omega) & \text{if } S_w(nL,\omega) > \alpha P_e(\omega) \\ 0 & \textit{otherwise} \end{cases} \tag{7.2}$$

where the parameter $\alpha$ serves as a control for the degree of noise smoothing to be achieved. In practice, it has been found that values of $\alpha$ between 2 and 3 produce acceptable results [6,24]. The analysis window, w(n), and the sampling interval, L, are chosen so that

$$\sum_{k=-\infty}^{\infty} w(kL-n) = 1 \qquad \text{for } \textit{all } n \tag{7.3}$$

This is done to make the mapping to the time domain easier. In the standard technique, $\hat{S}_w(nL,\omega)$ is combined with the short-time spectral phase of $x(n)$, to give a function $D_w(nL,\omega)$. To map back to the time domain, we take the inverse Fourier transform of $D_w(nL,\omega)$ for each $n$. The various time functions thus obtained are simply added to each other in the time domain to give the spectral subtraction estimate of $s(n)$. However, if (7.2) is not satisfied, some additional processing is necessary before the addition of the final short-time sections in order to avoid short-time boundary artifacts in the estimate of s(n).

### 7.3 Magnitude-Only Short-Time Spectral Subtraction

In the previous section, we introduced the standard noise reduction technique for short-time spectral subtraction. In this section we introduce a different short-time implementation of spectral subtraction that uses results on signal estimation from short-time spectral magnitude. The principal difference from the standard implementation is that the short-time spectral phase of the noisy signal is not required by this technique. Instead, a phase function is estimated from the processed short-time spectral magnitude.

As in the previous section, we consider the processing of a discrete time signal $x(n)$ which is the sum of a desired signal $s(n)$ and an uncorrelated stationary noise signal $e(n)$ with known power spectral density $P_e(\omega)$. The initial processing of the short-time spectral magnitude $S_w(nL,\omega)$ of $x(n)$ is identical to that performed in the standard technique of section 7.2. Specifically, we obtain a modified short-time spectral magnitude given by

$$\hat{S}_w(nL,\omega) = \begin{cases} S_w(nL,\omega) - \alpha P_e(\omega) & \text{if } S_w(nL,\omega) > \alpha P_e(\omega) \\ 0 & \textit{otherwise} \end{cases}$$

where the parameter $\alpha$ serves as a control for the degree of noise smoothing to be achieved. The next step in the standard technique is to combine $\hat{S}_w(nL,\omega)$ with the short-time spectral phase of the noisy signal $x(n)$. However, from chapter 5 we know that we can obtain a signal estimate directly from the modified short-time spectral magnitude $\hat{S}_w(nL,\omega)$. For the signal estimation algorithms of the previous chapter we require a-priori knowledge of $L$ consecutive samples of $x(n)$, starting from the first non-zero sample. Our approach, as described in chapter 5, is to use some reasonable estimate for those samples. For example, one approach is to use the corresponding $L$ samples of the noisy signal $x(n)$. This has produced reasonable results in the processing of noisy speech.

In our experiments with magnitude-only short-time spectral subtraction, we have applied the sequential iterative technique of chapter 4 for the signal estimation from processed short-time spectral magnitude. We selected this particular technique because of its simple implementation requirements. Furthermore, as indicated in chapter 4, it performs well compared to the other sequential reconstruction techniques that have been tested in this thesis.

For noise reduction in speech signals, it appears from our experiments that the performance of magnitude-only short-time spectral subtraction is comparable to that of standard short-time spectral subtraction. For relatively high signal to noise ratios (above 10dB), both techniques significantly reduce the noise without any appreciable degradation in speech quality. Figure 7.1 shows the waveform of the sentence " The bowl dropped from his hand" (See Figure 4.3 for original waveform) in additive white noise, giving a signal to noise ratio of 15 dB. Figure 7.2 shows

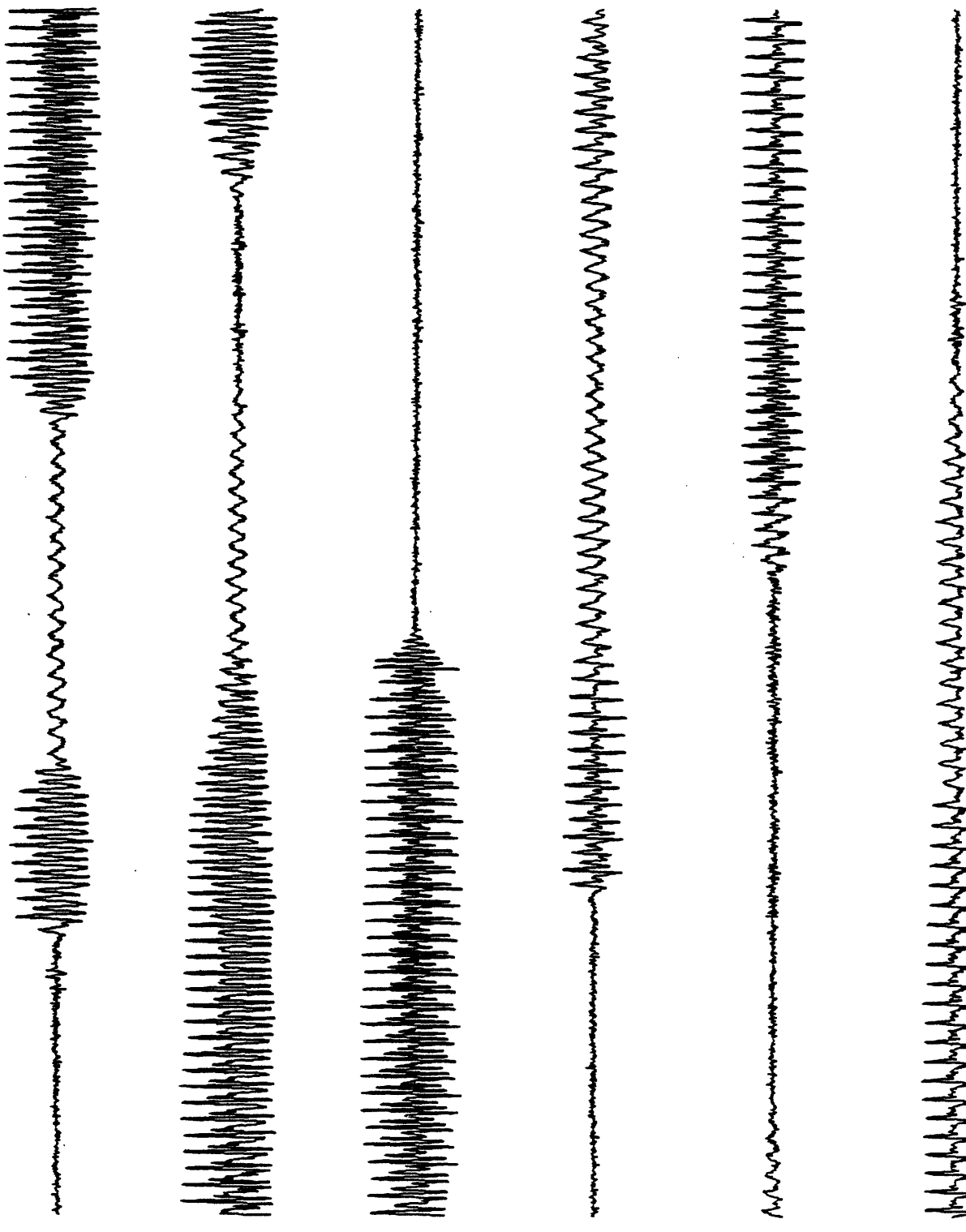


Fig. 7.1 Test Sentence in 15*dB* Additive White Noise

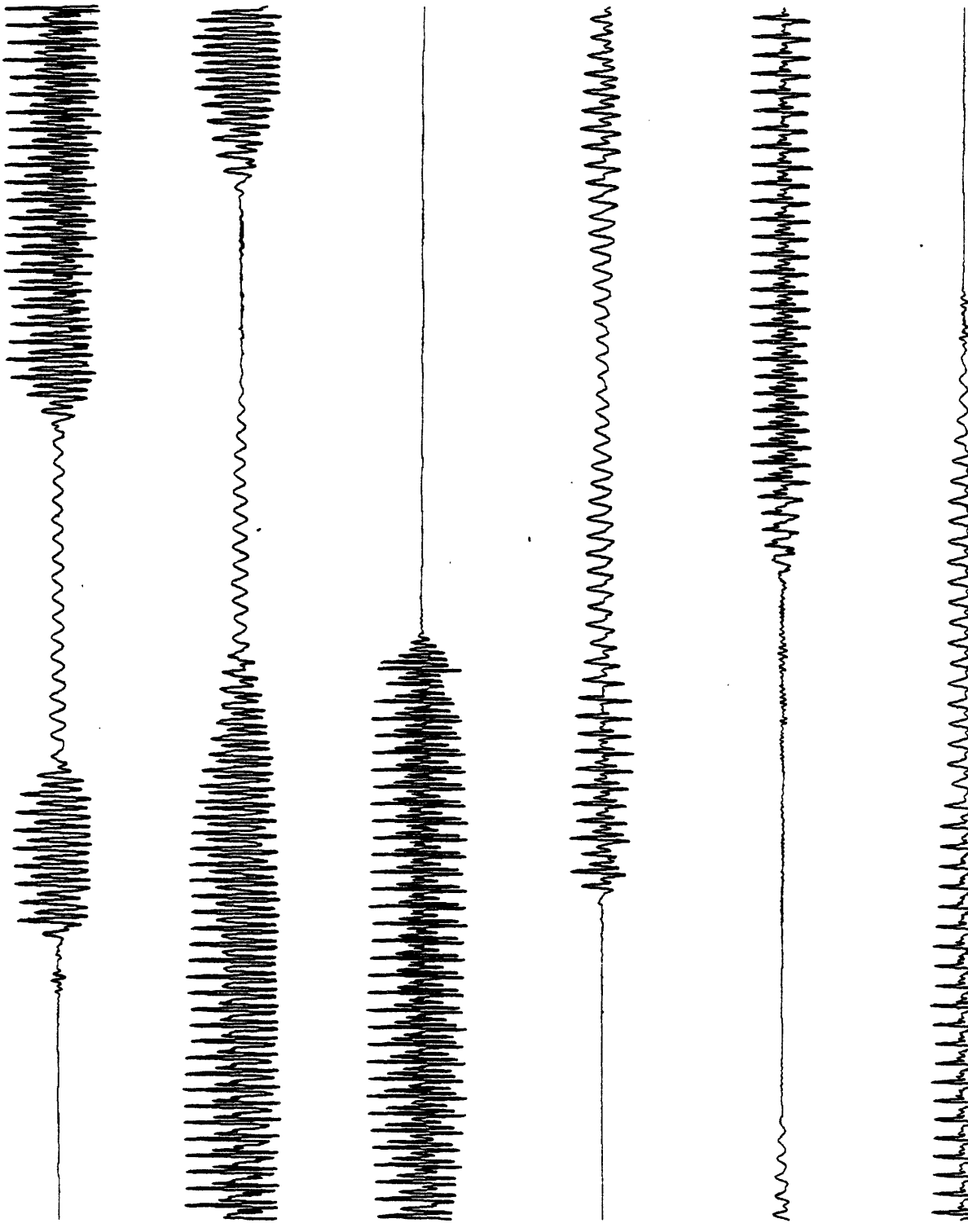


Fig. 7.2a Standard Short-Time Spectral Subtraction

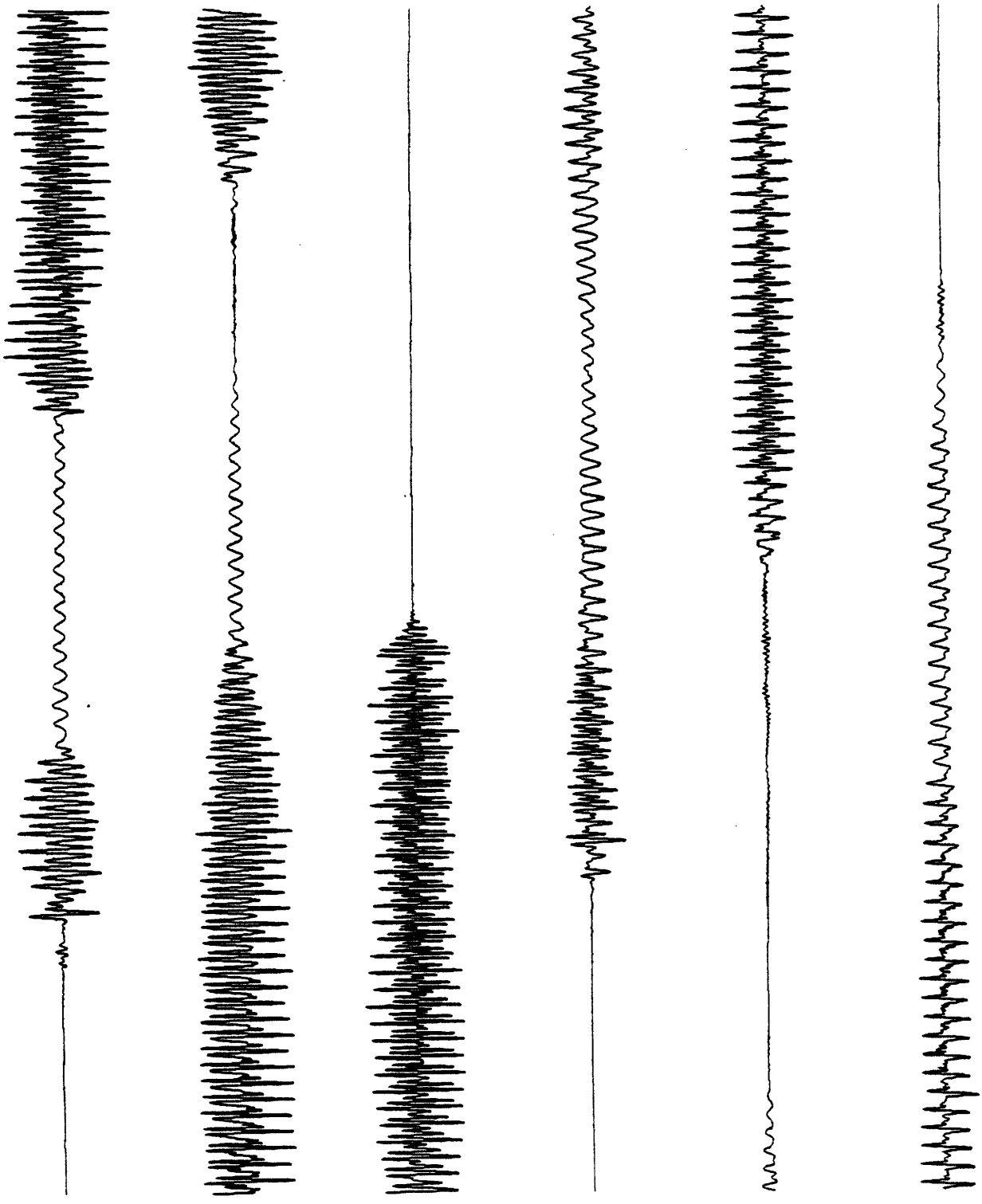


Fig. 7.2b Magnitude-Only Short-Time Spectral Subtraction

the results of processing that waveform with standard as well as magnitude-only spectral subtraction. In both cases, a 128-point triangular analysis window was used. Clearly, both the processed waveforms of Figure 7.2 have significantly reduced noise levels.

For signal to noise ratios below 10 dB, both versions of spectral subtraction introduce significant processing artifacts in the signal. In the next section, we describe these artifacts, discuss their causes, and present some techniques for suppressing them.

### 7.4 Artifacts in Short-Time Spectral Subtraction

When short-time spectral subtraction is applied to signals with low signal to noise ratios such as below 10 dB, certain processing artifacts are generally observed. In Figure 7.4, we illustrate these artifacts in an image that has been processed with standard short-time spectral subtraction. This image was obtained by adding 6 dB of white noise to the image of Figure 7.3 and then processing it with the two dimensional version of standard short-time spectral subtraction. Evident in Figure 7.4 are two types of distortion. One is the presence of an apparently harmonic pattern, particularly in the large high brightness region of the picture. Also noticeable are "ripple" blurring effects near high contrast sharp edges such as between the clock and the background. Although generally not as severe as the distortion represented by the harmonic pattern, this is also a quality limiting artifact. Similar distortions are also apparent in applying standard short-time spectral subtraction to speech. In this case, the processing typically results in the presence of objectionable short tone bursts of varying frequency. In our experiments with magnitude-only spectral subtraction applied to speech, we have also observed the same artifacts. In this section we will discuss the causes of these artifacts. In the next section, we propose various techniques for suppressing the artifacts. In particular, the proposed techniques can be incorporated into both the standard as well as the magnitude-only versions of short-time spectral subtraction.

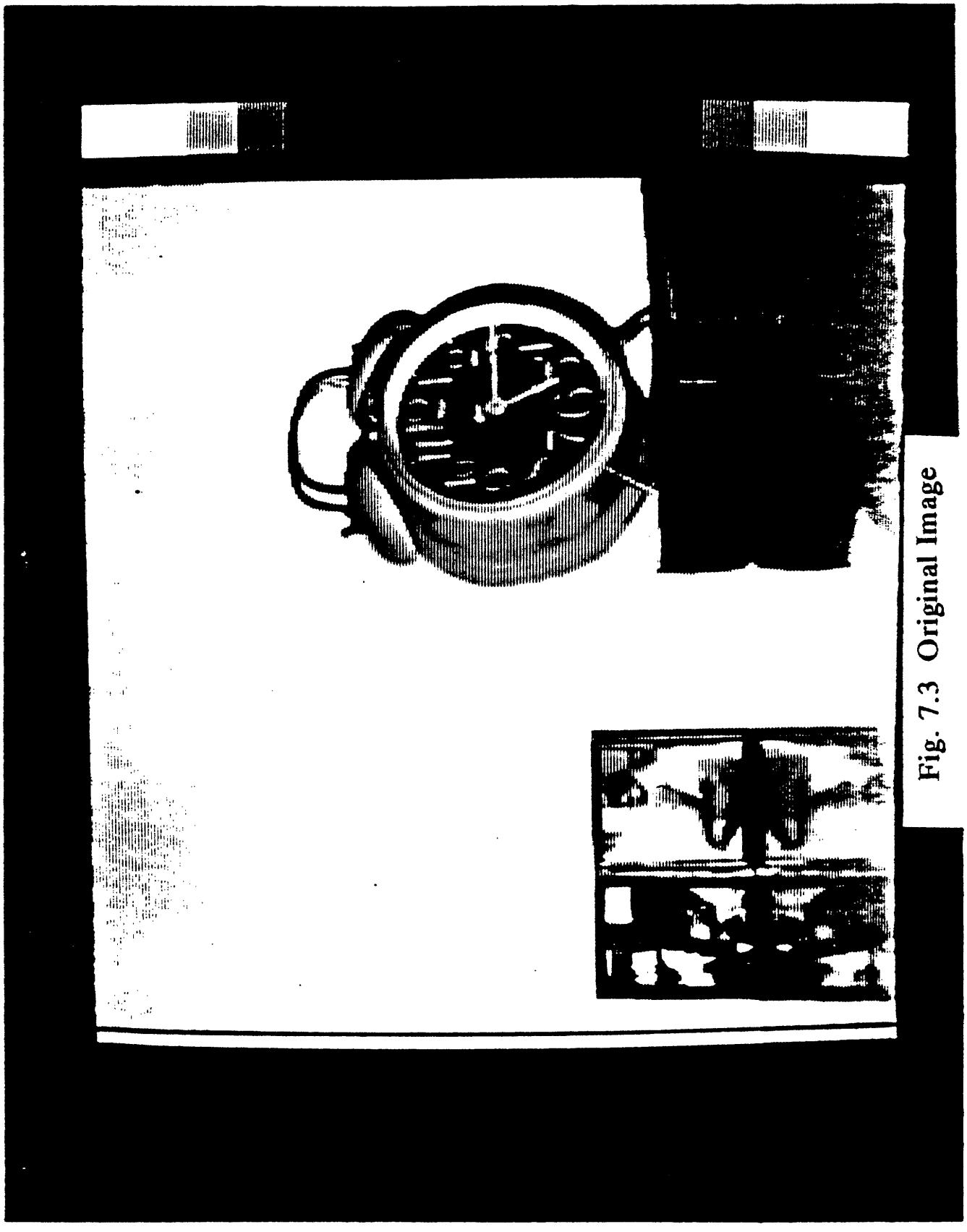


Fig. 7.3 Original Image

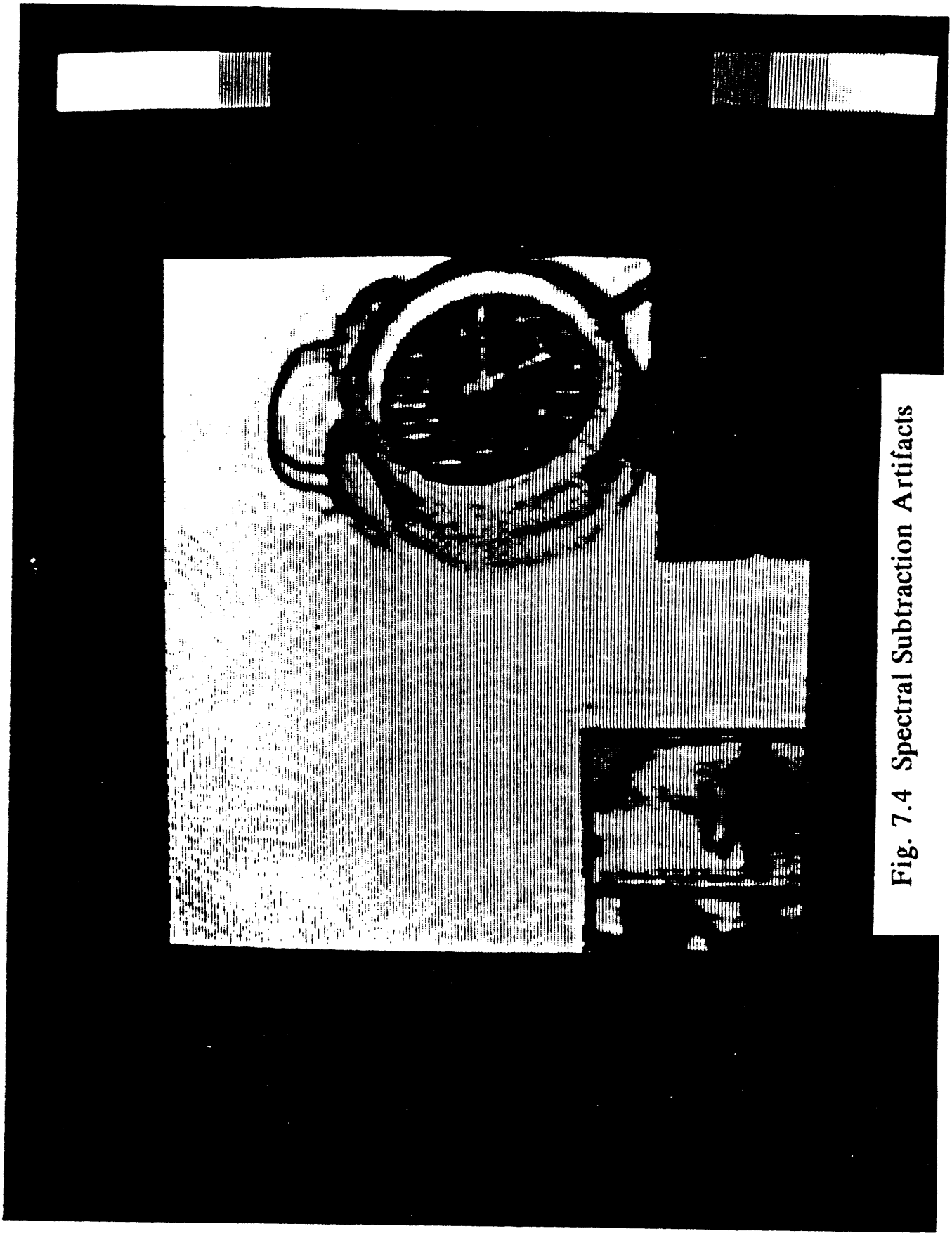


Fig. 7.4 Spectral Subtraction Artifacts

Experiments with short-time spectral subtraction on noisy speech as well as image data indicate that the objectionable harmonic pattern artifacts arise primarily because of a few, large amplitude narrowband peaks of noise energy remaining after spectral subtraction in the spectrum of each short-time section. Specifically, the spectral magnitudes of short-time sections in the noise signal $e(n)$ deviate randomly about the assumed power spectrum, $P_e(\omega)$. Such deviations result in residues of noise energy remaining after spectral subtraction. For wideband random noise with $\alpha$ values in (6.1) in the range generally used [6,24], the residues tend to be dominated by a few narrowband peaks of relatively large amplitude. Of these peaks, the most undesirable are the ones at frequencies where there is little or no signal energy. These give rise to harmonic variations in the short-time sections. Since the noise component of the spectrum has independent deviations from short-time section to short-time section, the dominating frequency of the harmonic patterns also changes randomly from short-time section to short-time section. The resulting artifact is clearly apparent in the image of Figure 7.4. In the next section, we propose specific techniques for suppressing this artifact. One of the techniques, referred to as multi-window spectral smoothing also reduces the rippling effect near large discontinuities. This particular artifact is due to the inherent blurring associated with signal characteristics which change rapidly in relation to the duration of the analysis window.

### 7.5 Artifact Suppression Techniques

In the previous section, we observed two particular artifacts associated with short-time spectral subtraction in both its standard and magnitude-only implementations. The most prominent artifact is the harmonic pattern which is clearly visible in the processed image of Figure 7.4. The other artifact, more important for images rather than speech, is a rippling effect near large discontinuities such as those at object boundaries. In this section, we propose three techniques for the suppression of the harmonic pattern artifact. However, one of the techniques, multi-window spectral subtraction, also reduces the rippling effect at sharp discontinuities. Throughout the remainder of this section the term short-time spectral subtraction without any other qualification

will refer to both the standard as well as the magnitude only implementations.

**Multipass Spectral Subtraction**

The implementation of this technique for suppressing artifacts consists of repeated application of the entire short-time spectral subtraction procedure. Specifically, on each of the total of K passes $\alpha/K$ is used in (6.1) and a new estimate of s(n) is obtained. Each pass uses the estimate of s(n) from the previous pass as its input.

The key to this procedure seems to lie in the post-subtraction mapping to the time domain at the end of each pass. Based on experiments conducted for this thesis, it is conjectured that the mapping to the time domain causes a spectral magnitude smoothing between overlapping short-time sections. Thus, as noise energy is being subtracted, a smoothing process is taking place simultaneously between overlapping sections. Furthermore, as K increases, the spectrum of each short-time section begins to affect the smoothing of distant spectral magnitudes. The idea of smoothing between the spectral magnitudes of different short-time sections is more directly explored in the Neighborhood Smoothing technique described next.

**Neighborhood Smoothing**

This approach is based on the assumption that the spectral magnitudes of neigboring short-time sections in the degraded signal have larger deviations with respect to each other than similar sections in the undegraded signal. Thus, if the spectral magnitudes of neighboring segments of the noisy signal are averaged, then, in principle, the effect of the noise is reduced. This, in effect, corresponds to time smoothing of the short-time spectral magnitude. The neighborhood smoothing can be carried out using either linear smoothing or median smoothing [26] techniques.

**Multi-Window Smoothing**

This technique capitalizes on the flexibility in the choice of the analysis window. In its most general form, the idea is to obtain signal estimates using different analysis windows for the short-time spectrum (but the same amount of spectral subtraction). This is followed by some kind of spectral smoothing between the different estimates. In particular, it is found that using

the same window shape but shifted locations for each estimate is very successful. In the experiments conducted for this thesis, short-time spectral magnitudes of the various estimates were median averaged in the final step, using a rectangular analysis window.

All three techniques listed above significantly reduce the harmonic pattern artifact significantly. Furthermore, the multi-window technique is also successful in reducing the rippling artifacts at sharp discontinuity. The performance of these techniques is illustrated in Figure 7.5. The image on the left side of the figure is the same image as that shown in Figure 7.4. This image was processed with standard short-time spectral subtraction without any modifications for suppression artifact. On the other hand, the other image in Figure 7.5 represents the effect of short-time spectral subtraction implemented with the multi-pass and multi-winow procedures. Specifically, multi-window spectral smoothing is carried out in each pass of the multipass implementation of short-time spectral subtraction. It is apparent from Figure 7.5 that the application of artifact suppression procedures is quite successful in reducing the harmonic pattern as well as the rippling effects near high contrast edges. We have also applied these techniques to noise reduction in speech processing. We find that the objectionable short tone bursts of varying frequency are significantly suppressed.

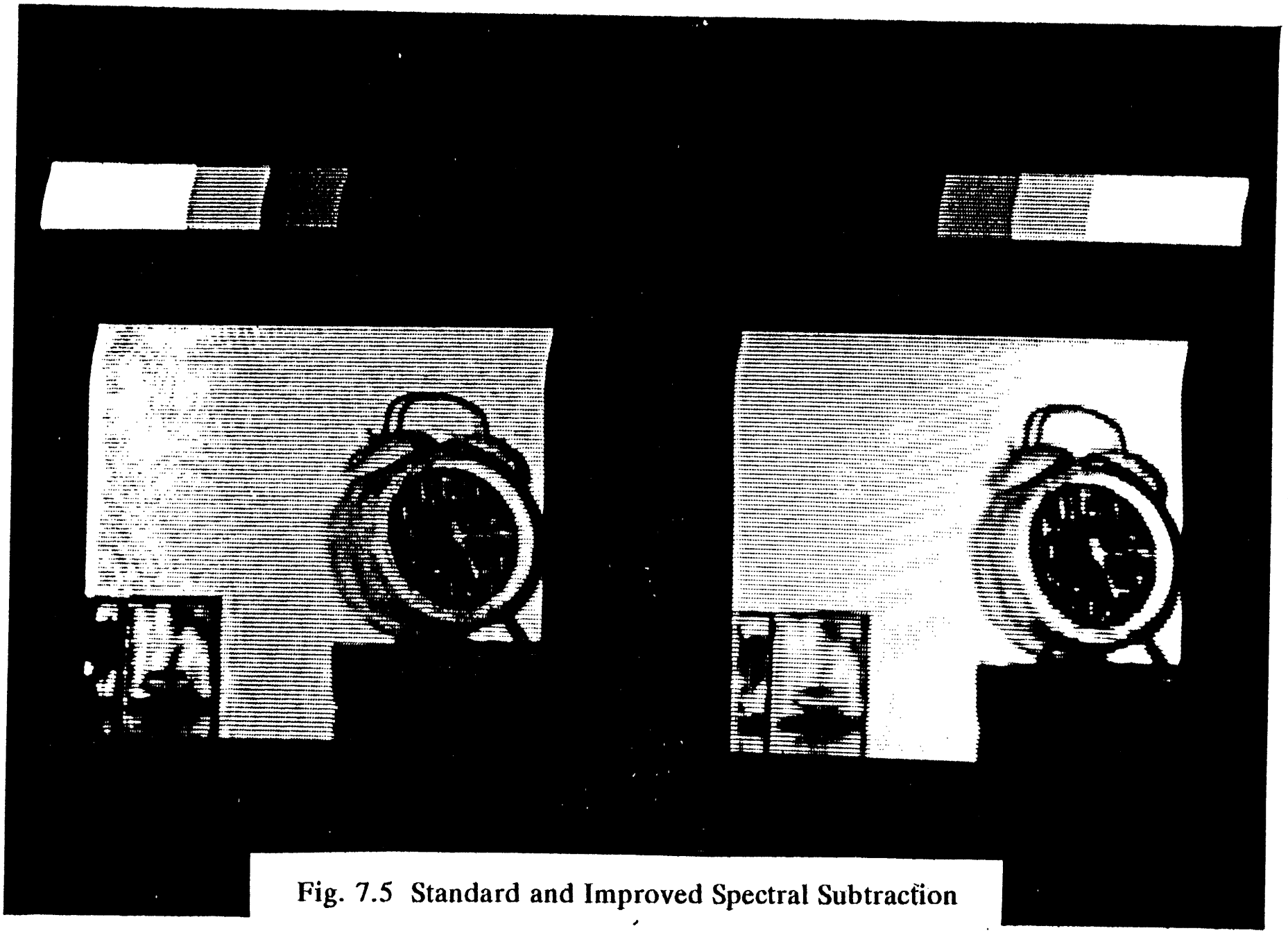


Fig. 7.5 Standard and Improved Spectral Subtraction

# CHAPTER EIGHT: CONCLUSIONS

In this thesis, we have shown that discrete-time signal processing can be accomplished using only the magnitude of the short-time spectrum. In particular, large classes of signals were found to be uniquely representable with the short-time spectral magnitude under conditions that are often satisfied in practical applications. Furthermore, several algorithms were derived for reconstructing a discrete-time signal from samples of its short-time spectral magnitude. These algorithms include some that are designed to yield reasonable signal estimates from a processed short-time spectral magnitude which does not correspond to the short-time spectral magnitude of any signal. This is an important result since almost any kind of processing violates the structure imposed in the definition of the short-time spectral magnitude.

To illustrate the practical usefulness of the results in this thesis, we considered the problems of noise reduction and time-scale modification of speech. The magnitude-only short-time spectral processing technique we have developed for time-scale modification is considerably simpler and computationally more efficient than previous short-time spectral processing techniques. Furthermore, in terms of speech quality, the magnitude-only technique appears comparable to the other techniques. In the case of noise reduction, standard short-time spectral processing techniques generally affect just the magnitude of the short-time spectrum. Thus,the short-time spectral phase of the noisy signal is retained in the processed signal. It is therefore of interest to develop techniques that estimate a processed short-time spectral phase. One approach is to estimate the processed phase directly from the processed short-time spectral magnitude. This is easily accomplished with the techniques developed in this thesis for signal estimation from processed short-time spectral magnitude. Our initial experiments on such noise reduction in speech signals have given results that appear comparable to those obtained with traditional short-time spectral processing techniques. This result is potentially useful in designing systems for combined noise and bandwidth reduction of speech. Such systems perform bandwidth compression on a noisy signal, transmit it over a possibly noisy channel, and finally estimate the original undegraded

signal. The results in this thesis may be used to achieve bandwidth compression of the noisy signal by efficiently coding its short-time spectral magnitude. Time-scale modification may also be used at this stage. Once the transmitted signal has been received at the other end, magnitude-only short-time spectral processing may be applied for noise reduction. This is an important application which deserves more research in the future.

There is a considerable amount of theoretical and applied research that needs to be persued in light of the results presented in this thesis. The most obvious problem is to use these results in application areas other than those considered in this thesis.This includes other applications within speech processing such as vocoder design as well as applications in other areas such as image, acoustical and geophysical signal processing. In the theoretical realm, it is of interest to further extend the conditions under which a signal is uniquely specified by its short-time spectral magnitude. For example, the uniqueness conditions derived in this thesis for signal representation with short-time spectral magnitude generally require the knowledge of a few initial samples of the signal. It is of interest to determine other ways of guaranteeing unique signal specification that require a different type of information about the signal. It should be observed that the need for the initial samples condition was established through a counterexample that was based on a special class of signals and a rectangular analysis window. The question is whether excluding the rectangular analysis window and that special class of signals can in fact be sufficient to remove the requirement of a-priori knowledge on the initial samples.

All the algorithms for signal estimation from short-time spectral magnitude that were implemented in this thesis estimated short-time sections of the signal in a sequential order. However, it was indicated that improved performance may be obtained by the simultaneous extrapolation of several short-time sections at a time. For example, such algorithms may be less sensitive to errors in the knowledge of the initial signal samples. The implementation and study of simultaneous extrapolation algorithms should therefore be an important part of further research.

**References**


1. A.V.Oppenheim and R.W.Schafer, *Digital Signal Processing*, Prentice Hall (1975).
2. L.R.Rabiner and R.W.Schafer, *Digital Processing of Speech Signals*, Prentice Hall (1978).
3. C.J.Weinstein, "Short-Time Fourier Analysis and its Inverse," *S.M. Thesis, Massachusetts Institute of Technology* (1966).
4. J.B.Allen, "Short-Term Spectral Analysis and Synthesis and Modification By Discrete Fourier Transform," *IEEE Trans. ASSP* ASSP-25(3), pp.235-238 (1977).
5. M.R.Portnoff, "Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis," *IEEE Trans. ASSP* ASSP-28, pp.55-69 (Feb. 1980).
6. J.S.Lim and A.V.Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proceedings of the IEEE* 67(12), pp.1586-1604 (Dec. 1979).
7. M.R.Portnoff, "Time-Scale Modification of Speech Based on Short-Time Fourier Analysis," *IEEE Trans. ASSP* ASSP-29(3), pp.374-390 (June 1981).
8. R.M.Fano, "Short-Time Autocorrelation Functions and Power Spectra," *J. Acoustical Society of America* 22(5), pp.546-550 (Sept. 1950).
9. M.R.Schroeder and B.S.Atal, "Generalized Short-Time Power Spectra and Autocorrelation Functions," *J. Acoustical Society of America* 34(11), pp.1679-1683 (Nov. 1962).
10. A.A.Kharkevich, *Spectra and Analysis*, Consultants Bureau Enterprises, NewYork (1960).
11. R.A.Altes, "Detection, Estimation, and Classification with Spectrograms," *J. Acoustical Society of America* 67, pp.1232-1246 (April 1980).
12. M.R.Portnoff, *PhD Thesis, Massachusetts Institute of Technology*, 1978.
13. G.Fairbanks, W.L Everitt, and R.P Jaeger, "Method for Time or Frequency Compression-Expansion of Speech," *IRE Trans.* AU 2(1), pp.7-12 (Jan.-Feb. 1954).
14. J.L. Flanagan and R.M.Golden, "Phase Vocoder," *Bell Syst. Tech. J.* 45, pp.1493-1509 (Nov. 1966).
15. S. Holtzman, "Non-Uniform Time-Scale Modification of Speech," *S.M. Thesis, Massachusetts Institute of Technology* (1980).
16. R.J.Scott and S.E.Gerger, "Pitch Synchronous Time Compression of Speech," *Proc. Conf. Speech Comm. Processing*, pp.63-65 (April 1972).
17. E.P. Neuberg, "Simple Pitch-Dependent Algorithm for High Quality Speech Rate Change, Abstract," *J. Acoustical Society of America* 61 (Spring 1977).
18. H.D.Toong, "A Study of Time-Compressed Speech," *PhD Thesis, Massachusetts Institute of Technology* (1974).
19. M.R.Portnoff, "Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform," *IEEE Trans. on ASSP* 24(3), pp.243-248 (June 1976).
20. J.A.Moorer, "The Use of the Phase Vocoder in Computer Music Applications," *55th Convention of the Audio Engineering Society, Preprint*(1146 (E-1)) (Oct. 1976).
21. A.V.Oppenheim, R.W.Schafer, and J.G.Stockham,Jr., *IEEE Proceedings* 56, p.64 (1968).
22. H.J.Trussel and B.R.Hunt, "Sectioned Methods for Image Restoration," *IEEE Trans. on ASSP* 26, p.157 (1978).
23. H.VanTrees, *Detection, Estimation and Modulation Theory*, John Wiley (1968).
24. J.S.Lim, "Image Restoration by Short-Space Spectral Subtraction," *IEEE Trans. on ASSP* 28, p.191 (1980).

25. L.G.Roberts, "Picture Coding Using Pseudo-Random Noise," *IRE Trans. Inf. Theory* 8(2), , pp.145-154 (Feb. 1962).

26. L.R.Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice Hall (1975).