Signal Reconstruction from Short-Time Fourier Transform Magnitude

S. HAMID NAWAB, THOMAS F. QUATIERI, MEMBER, IEEE, AND JAE S. LIM, MEMBER, IEEE

Abstract—**In this paper, a signal is shown to be uniquely represented by the magnitude of its short-time Fourier transform (STFT) under mild restrictions on the signal and the analysis window of the STFT. Furthermore, various algorithms are developed which reconstruct signal from appropriate samples of the STFT magnitude. Several of the algorithms can also be used to obtain signal estimates from the processed STFT 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. Finally, the results presented here have similar potential for other application areas, including those with multidimensional signals.**

I. Introduction

THE short-time Fourier transform (STFT) is a widely used time-frequency representation for signals such as speech [1] and images [2]. In this paper, we develop certain properties of the magnitude of the STFT and demonstrate the usefulness of these results for various speech processing applications.

For a discrete-time signal $x(n)$, the STFT is defined as

$$X_w(nL, \omega) = \sum_{m=-\infty}^{\infty} x(m)\, w(nL - m)\, e^{-j\omega m} \tag{1}$$

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 $n, X_w(nL, \omega)$ 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 [1] 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 STFT, including a well-known filter bank interpretation [1]. For the purposes of this paper, we find the sliding window interpretation to be the most appropriate.

The STFT, which is generally a complex function, is a complete signal representation in the sense that a signal can be uniquely determined from its STFT [3]-[5]. The problem of representing a signal with its STFT magnitude alone, however, has remained unresolved despite significant efforts [3], [6] and its considerable practical importance. For example, the STFT magnitude of signals such as speech is often considerably easier to determine and is more meaningful in practice than the STFT phase [3]. A central result of this paper is that under certain mild conditions, a discrete-time signal is uniquely specified by its STFT magnitude. In addition, several algorithms are developed that reconstruct the signal from its STFT magnitude.

In many signal processing applications it is desirable to estimate a signal from a modified STFT magnitude. For example, in speed transformation of speech, a technique developed by Portnoff [7] estimates the transformed speech from the modified STFT of the original speech. The modification of the STFT in this case includes a computationally difficult phase unwrapping operation [8]. If speech could be synthesized from the modified STFT magnitude alone, a speed transformation system may be developed which is similar in style but requires significantly less computation than Portnoff's method. A second example of the usefulness of signal reconstruction from STFT magnitude is in the area of speech enhancement. Conventionally, speech enhancement methods such as Wiener filtering and spectral subtraction [9] are used to estimate the STFT magnitude of the original speech from a noisy version, while the STFT phase of the noisy speech is retained. If speech could be synthesized from the modified STFT magnitude alone, this would be an alternative approach to speech enhancement which does not require the STFT phase of the noisy speech. In this paper, we include algorithms for such signal estimation from the modified STFT magnitude.

This paper is organized as follows. In Section II, we develop results on the extrapolation of a finite-length signal from its (long-time) Fourier transform magnitude. These results are used in Section III to develop conditions under which the STFT is a unique signal representation. In Section IV, we present algorithms that reconstruct a signal from its STFT magnitude. These algorithms have different implementation properties and have been tested for the reconstruction of speech signals. In Section V, we discuss the application of the algorithms developed in Section IV to signal estimation from the modified STFT magnitude. In particular, we consider the applica-

Manuscript received October 20, 1981; revised July 1, 1982 and December 5, 1982. This work was supported in part by the Department of the Air Force and in part by the National Science Foundation under Grant ECS80-07102, and in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506.

The authors are with the Lincoln Laboratory and Research Laboratory of Electronics, Massachusetts Institute of Technology, Lexington, MA 02173.

0096-3518/83/0800-0986$01.00 © 1983 IEEE

tion of these algorithms to the problems of time-scale modification and enhancement of speech.

II. Signal Extrapolation from Fourier Transform Magnitude

In this section, we derive theorems on the extrapolation of a discrete-time signal from its (long-time) Fourier transform magnitude. Besides being important theoretical results in their own right, these theorems play a central role in deriving conditions under which the STFT 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 section are known to be zero outside an interval $0 \leqslant n \leqslant N$ 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 section are affected by a shift in this location. Given $x(n)$ for $0 \leqslant n \leqslant M$ where $M < N$, we wish to extrapolate $x(n)$ up to $n = N$, using the spectral magnitude, $|X(\omega)|$, where

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)\, e^{-j\omega n}. \tag{2}$$

Furthermore, we are interested in determining conditions under which the extrapolation is unique. Section II-B derives a theorem on such extrapolation for the case in which only the sample $x(N)$ is unknown. This is referred to as single sample extrapolation. Section II-C presents a theorem for the more general case, where more than one sample of $x(n)$ is extrapolated. These theorems are used extensively in Section III for deriving conditions under which the STFT magnitude is a unique signal representation. The relationship between the theorems in this section and the uniqueness of the STFT magnitude is first discussed in Section II-A. We then present the results of signal extrapolation from Fourier transform magnitude in Sections II-B and II-C.

A. Relation to STFT

For a signal $x(n)$ and a positive integer L, the STFT magnitude is defined as

$$S_w(nL, \omega) = \left| \sum_{m=-\infty}^{\infty} x(m)\, w(nL - m)\, e^{-j\omega m} \right| \tag{3}$$

where the subscript w in $S_w(nL, \omega)$ refers to the analysis window $w(n)$. The Fourier transform magnitude of the short-time section for a particular window shift of $n_0 L$ 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 STFT magnitude is said to be computed with *maximum analysis window overlap*. Finally, when $L > 1$, the STFT 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 [8] 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 nonzero samples by the corresponding samples of the analysis window. We will show in this section that the samples in the region of overlap can, under mild conditions, 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, i.e., as a new short-time section is extrapolated, it becomes possible to extrapolate a succeeding short-time section. 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. Section III 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 section.

B. Single-Sample Extrapolation

Consider a discrete-time signal $x(n)$ that is zero outside the interval $0 \leqslant n \leqslant N$. Theorem 1 of this section shows that the sample $x(N)$ can be uniquely obtained from the Fourier transform magnitude, $|X(\omega)|$, and $x(n)$ for $0 \leqslant n < N$.

Theorem 1: Let $x(n)$ be a sequence that is zero oustide the interval $0 \leqslant n \leqslant N$. Suppose $x(0)$ is nonzero. Then, $|X(\omega)|$ 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, where

$$R(n) = \sum_{m=-\infty}^{\infty} x(m)\, x(n+m). \tag{4}$$

Since $x(0)$ is the first nonzero sample of $x(n)$ and $x(n) = 0$ for $n > N$, it follows that

$$R(N) = x(0)\, x(N). \tag{5}$$

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

$$x(N) = R(N)/x(0). \tag{6}$$

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 $R(-N)$. 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 $|X(\omega)|^2$. Since $|X(\omega)|^2$ is an even function, it can be easily seen that $2N$ uniformly spaced samples of $|X(\omega)|^2$ over the frequency interval $[0, 2\pi]$ are equivalent to $N+1$ samples in the interval $[0, \pi]$. This is consistent with the fact that $R(n)$ has $N+1$ unknown samples. More generally, it can also be shown that $R(n)$ can be obtained even if the $N+1$ samples of $|X(\omega)|^2$ in $[0, \pi]$ are not uniformly sampled.

C. Multiple Samples Extrapolation

This section presents a theorem on the extrapolation 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 $0 \leqslant n \leqslant N$. As indicated earlier, 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 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 1. In fact, Theorem 1 can be derived as a corollary of Theorem 2. However, we chose not to do this in order to emphasize the simplicity of the direct proof of Theorem 1.

Theorem 2: For $N > 0$ let $x(n)$ be a sequence that is zero outside the interval $0 \leqslant n \leqslant N$. Suppose $x(0)$ is nonzero. Then, $|X(\omega)|$ and the P samples of $x(n)$ in the interval $0 \leqslant n < P$ uniquely specify the entire sequence $x(n)$ *if and only if* $P \geqslant \lceil M/2 \rceil$ (where $M = N+1$ and $\lceil \alpha \rceil$ is the smallest integer greater or equal to α).

Proof: Throughout this proof, the samples of $x(n)$ for $0 \leqslant n < P$ will be referred to as the initial P samples of $x(n)$.

We first show that the unknown samples of $x(n)$ are uniquely specified when $P = \lceil M/2 \rceil$. Clearly, if uniqueness holds for $P = \lceil M/2 \rceil$, uniqueness must also hold for $P > \lceil M/2 \rceil$. From $|X(\omega)|$ the autocorrelation $R(n)$ of $x(n)$ can be obtained:

$$R(n) = x(n) * x(-n) = \sum_{m=0}^{M-1-n} x(m)\, x(n+m). \tag{7}$$

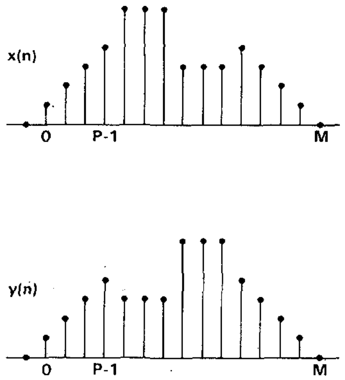


Fig. 1. Counterexample for the proof of Theorem 2.

Consider the case where $M = N+1$ is even. From (7), $M/2$ linear equations are obtained in $M/2$ unknowns, $x(M/2)$, $x((M/2)+1), \cdots, x(M-1)$. In matrix form these equations are

$$\begin{bmatrix} x(0) & & & \\ x(1) & x(0) & & \\ x(2) & x(1) & & \\ \vdots & \vdots & & \\ x((M/2)-1) & x((M/2)-2) & \cdots & x(0) \end{bmatrix} \cdot \begin{bmatrix} x(M-1) \\ x(M-2) \\ \vdots \\ x(M/2) \end{bmatrix} = \begin{bmatrix} R(M-1) \\ R(M-2) \\ \vdots \\ R(M/2) \end{bmatrix}. \tag{8}$$

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 $x(n)$, $n = M/2, (M/2)+1, \cdots, M-1$. For M odd, the $(M-1)/2$ unknowns $x((M+1)/2)$, $x(((M+1)/2)+1), \cdots, x(M-1)$ are solved for, through a set of equations similar to (8).

We now provide counterexamples to show that if $P < \lceil M/2 \rceil$, then $x(n)$ cannot in general be uniquely specified by $|X(\omega)|$ and the initial P samples. In particular, it is easily seen that the condition $P < \lceil M/2 \rceil$ is equivalent to the following two conditions:

a) $P < \lfloor M/2 \rfloor$

b) $P = \lfloor M/2 \rfloor$, M odd

where $\lfloor \alpha \rfloor$ is the largest integer smaller or equal to α. We will now construct counterexamples for each of these two cases.

For $P < \lfloor M/2 \rfloor$ consider any sequence $x(n)$ such that $x(n) = x(N-n)$ for $n = 0, 1, \cdots, P-1$, and $x(n) \neq x(N-n)$ for $n = P, P+1, \cdots, M-P-1$ (see Fig. 1). Then, the sequences $x(n)$ and $y(n) = x(N-n)$ have the same samples for $n = 0, 1, \cdots, P-1$. Furthermore, since $y(n)$ is a time-reversed and shifted version of $x(n)$, the two sequences have the same Fourier transform magnitude [8]. Since $x(n) \neq x(N-n)$ for $n = P, P+1, \cdots, N-P$, $y(n)$ and $x(n)$ are distinct. Thus, the initial

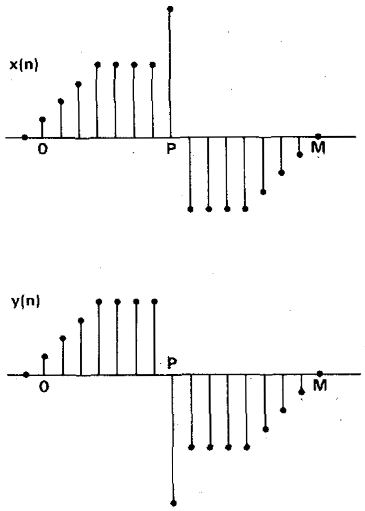


Fig. 2. Counterexample for the proof of Theorem 2.

P samples and $|X(\omega)|$ are not sufficient to uniquely represent $x(n)$.

For M odd and $P = \lfloor M/2 \rfloor$ we consider any sequence $x(n)$ such that $x(n) = -x(N-n)$ for $n = 0, 1, \cdots, P-1$ and $x(P) \neq 0$ (see Fig. 2). Then the sequences $x(n)$ and $y(n) = -x(N-n)$ have the same samples for $n = 0, 1, \cdots, P-1$ and $y(P) = -x(P)$. Furthermore, $|Y(\omega)| = |X(\omega)|$ and $y(n) \neq x(n)$. This completes the proof of Theorem 2.

III. Signal Representation with Short-Time Fourier Transform Magnitude

In this section, we address the problem of uniquely representing a signal by its STFT magnitude. That is, we develop conditions under which distinct sequences cannot have the same STFT magnitude. In deriving these conditions, we assume that the analysis window is a *known* finite-length sequence. This permits us to use the extrapolation theorems of Section II for developing conditions which ensure unique correspondence between a signal and its STFT 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 III-A. In Section III-B we present various conditions for unique signal representation with the STFT magnitude. These conditions concern primarily 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 STFT magnitude. However, the conditions we develop are broad enough to be of significant practical interest, as illustrated in later sections. In Section III-C, we demonstrate that the uniqueness conditions can be easily extended to the STFT magnitude of multidimensional signals.

A. Uniqueness Problems

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 STFT magnitude, $S_w(nL, \omega)$. In expression (3) 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 STFT magnitude. This ambiguity may be resolved, for example, by knowing the sign of some nonzero sample of $x(n)$.

In the case of a finite-length analysis window, a gap of zero samples between two nonzero 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. Suppose that the gap of zeros between $x_1(n)$ and $x_2(n)$ is large enough so that there is no short-time section which includes nonzero contribution from $x_1(n)$ as well as $x_2(n)$. Clearly, in such a situation, the STFT magnitude of $x(n)$ is the sum of the STFT magnitudes of $x_1(n)$ and $x_2(n)$. However, we previously saw that a signal and its negative have the same STFT magnitude. It follows that $x(n)$ has the same STFT magnitude as the signals obtained from the differences $x_1(n) - x_2(n)$ and $x_2(n) - x_1(n)$. 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 section include a restriction on the length of zero gaps between nonzero portions of the signal.

In Section III-B we will see that $S_w(nL, \omega)$ with $L = 1$ uniquely specifies a one-sided signal $x(n)$ within a sign factor under conditions whose only restriction on $x(n)$ is a limit on the size of any zero gaps. On the other hand, with $L > 1$ it can be shown that there are sequences that have no zero gaps and are not specified even up to a sign factor by $S_w(nL, \omega)$ [10]. For example, consider a one-sided signal $x(n)$, which for $n \geqslant 0$ is periodic with period L and is zero for $n < 0$. In addition, let the analysis window $w(n)$ be symmetric and its extent be over a region $0 \leqslant n < mL$ for some positive integer $m > 1$. In this case, it is easily seen that if each period of $x(n)$ is replaced by a time-reversed version which occupies the same portion of the time axis, the resulting sequence has the same $S_w(nL, \omega)$ as $x(n)$. Generalizations of this example have been developed for aperiodic signals as well [10].

We have thus 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 zero gap restriction. In Section III-B-2, knowledge of the L initial samples of a one-sided signal $x(n)$ is found to be sufficient for this purpose. This condition arises naturally from the extrapolation approach used in deriving the various results in this paper.

B. Uniqueness Conditions

In this section, we present various conditions and their derivations for uniquely representing a signal with its STFT magnitude. The analysis window 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 STFT 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 STFT.

1) Maximum Analysis Window Overlap: The STFT magnitude $S_w(nL, \omega)$, defined in (3), 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) is defined for positive integer values of L, it is clear that with $L = 1$, adjacent analysis window positions have maximum overlap and the STFT magnitude is given 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, we will use Theorem 1 on single sample extrapolation of finite-length sequences. Although the theorem is stated for $x(n)$ in the interval $0 \leqslant n \leqslant N$, 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 nonzero 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'. Furthermore, we restrict $w(n)$ to be nonzero over its finite length N_w. This simplifies the restrictions imposed on $x(n)$ for avoiding the zero gap ambiguities discussed in Section III-A.

Conditions 1–For Representing $x(n)$ Uniquely with $S_w(n, \omega)$:

$w(n)$: a) Known sequence of finite length $N_w > 1$
b) No zeros within length N_w

$x(n)$: a) One-sided
b) At most $N_w - 2$ consecutive zero samples between any two nonzero samples
c) Sign of first nonzero sample known.

To show that $S_w(n, \omega)$ uniquely specifies the signal $x(n)$ under Conditions 1, let us consider the case when the analysis window $w(n)$ is restricted to the interval $0 \leqslant n < N_w$. 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. Furthermore, we will consider only the case with $x(n)$ right-sided. The case with $x(n)$ left-sided can be proved analogously.

Let us consider all sequences $y(n)$ which have the same short-time spectral magnitude, $S_w(n, \omega)$, as $x(n)$. Clearly, $y(n)$ must satisfy

$$\left| \sum_m y(m)\, w(n - m)\, e^{-j\omega m} \right| = S_w(n, \omega). \tag{9}$$

We will show that under Conditions 1, $y(n)$ must equal $x(n)$. Let n' be the smallest value of n such that $S_w(n, \omega)$ is nonzero. Then from (9), it follows that $y(n) = 0$ for $n < n'$, and $y(n')$ is nonzero. In particular

$$S_w(n', \omega) = |w(0)\, y(n')| \quad \text{for all} \quad \omega. \tag{10}$$

We then have

$$y(n') = \pm \frac{S_w(n', 0)}{w(0)}. \tag{11}$$

The sign ambiguity in this equation can be resolved since Conditions 1 specify the sign of the first nonzero sample. Thus, (11) has a unique solution for $y(n')$. Since by assumption $x(n)$ satisfies (11), we have $y(n') = x(n')$.

Our next step is to use Theorem 1 to solve for $y(n' + 1)$. Since $N_w > 1$, the short-time section $f_{n'+1}(m) = y(m)\, w(n' + 1 - m)$ has zero samples outside the interval $n' \leqslant m \leqslant n' + 1$ and all its samples have been uniquely solved for except at $m = n' + 1$ where it equals $y(n' + 1)\, w(0)$. The spectral magnitude $S_w(n' + 1, \omega)$ of this section is known. Thus, applying Theorem 1 with $N = 1$, $y(n' + 1)\, w(0)$ can be obtained uniquely. Since $w(0)$ was assumed known and nonzero, we conclude that $y(n' + 1)$ has a unique solution. Since $x(n' + 1)$ is a known solution, $y(n' + 1) = x(n' + 1)$.

We have now shown that $y(n) = x(n)$ is the only solution to (9) up to and including $n = n' + 1$. We will next show that for any $n'' > n'$, $y(n'')$ has a unique solution, $y(n'') = x(n'')$, provided $y(n)$ has unique values $y(n) = x(n)$ for $n < n''$. By induction, we can then conclude that (9) has a unique solution $y(n) = x(n)$ for all n.

Consider the short-time section, $f_{n''}(m) = y(m)\, w(n'' - m)$. We assume that $y(n)$ has a unique value $y(n) = x(n)$ for $n < n''$ and we wish to show that $y(n'')$ has a unique value, $y(n'') = x(n'')$. Theorem 1 tells us that provided the $N_w - 1$ samples preceding $y(n'')$ are not all zero, $S_w(n'', \omega)$ uniquely determines $y(n'')$. Conditions 1 ensure that the $N_w - 1$ samples preceding $y(n'')$ are not all zero. We conclude that $y(n'')$ has a unique value, $y(n'') = x(n'')$.

Implicit in the above discussion is a recursive algorithm for determining $x(n)$ for $n > n'$. This procedure can be 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

$$r_n(m) = \sum_{k = n - (N_w - 1)}^{n} x(k)\, w(n - k)\, x(k - m)\, w(n - (k - m)). \tag{12}$$

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

$$x(n) = \frac{r_n(m) - \displaystyle\sum_{k = n - (N_w - 1)}^{n - 1} w(n - k)\, w(n - (k - m))\, x(k)\, x(k - m)}{w(0)\, w(m)\, x(n - m)}. \tag{13}$$

This is a valid equation only for values of m for which $w(m)\, x(n - m)$ is nonzero. Since $w(m)$ is nonzero only for $0 \leqslant m < N_w$, we require that $x(n - m)$ be nonzero 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 nonzero samples. This is consistent with our observation in Section III-A that there should not be a zero gap which gives a zero short-time section. Since Conditions 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.

From Section III-A, we know that $-x(n)$ has the same $S_w(nL, \omega)$ as $x(n)$. It follows that under Conditions 1, $-x(n)$ can be uniquely specified with $S_w(n, \omega)$. However, the only difference between $x(n)$ and $-x(n)$ in our proof above is that different signs are selected for $x(n')$ in (11). It follows that without the *a priori* sign knowledge in Conditions 1, $x(n)$ can be uniquely specified up to a sign ambiguity from $S_w(n, \omega)$. Consequently, the sign of *any* nonzero sample of $x(n)$ would be sufficient to guarantee unique specification of $x(n)$.

2) Partial Analysis Window Overlap: We will now develop a set of conditions that are sufficient for uniquely specifying a signal with its STFT 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 III-A, even if we do not allow any zero samples within $x(n)$, 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 nonzero sample.

Conditions 2–For Representing $x(n)$ Uniquely with $S_w(nL, \omega)$:

L: a) $1 < L \leqslant \lfloor N_w/2 \rfloor$

$w(n)$: 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 nonzero sample.

In the above conditions $\lfloor \alpha \rfloor$ denotes the largest integer smaller than or equal to α. The derivation of these conditions relies on Theorem 2.

Let us consider all sequences $y(n)$ which have the STFT magnitude $S_w(nL, \omega)$ of $x(n)$. Clearly, $y(n)$ must satisfy

$$\left| \sum_m y(m)\, w(nL - m)\, e^{-j\omega m} \right| = S_w(nL, \omega). \tag{14}$$

If $S_w(nL, \omega)$ is a unique representation of $x(n)$, then (14) should have a unique solution for $y(n)$, i.e., $y(n) = x(n)$. We will now show that under Conditions 2 there is indeed a unique solution.

Let n' be the smallest n for which $x(n) \neq 0$. Without loss of generality, assume $1 \leqslant n' \leqslant L$. Let $x_L(n)$ denote a sequence which equals $x(n)$ for $n' \leqslant n < n' + L$ and is zero otherwise. Thus, $x_L(n)$ represents the L known initial samples required by Conditions 2. Without loss of generality we assume that $w(n)$ occupies the region $0 \leqslant n < N_w$. Since $x_L(n)$ is given, it follows that $y(n)$ under the analysis window $w(L - n)$ is the same as $x(n)$. Our next objective is to solve for $y(n)$ over the duration of $w(2L - n)$.

In order to solve for $y(n)$ under $w(2L - n)$, consider the sequence $y_2(n) = y(n)\, w(2L - n)$. Since $L \leqslant \lfloor N_w/2 \rfloor$, knowledge of $x_L(n)$ assures that at least L samples of $y_2(n)$ beginning at $n = n'$ are the same as the corresponding samples of $x(n)\, w(2L - n)$. Furthermore, the length of $y_2(n)$ is $2L - n' + 1$ and $L \geqslant \lceil (2L - n' + 1)/2 \rceil$. Therefore, applying Theorem 2, $y_2(n)$ must have a unique value. Since $x(n)\, w(2L - n)$ is a known solution of (14), it follows that $y_2(n)$ must equal $x(n)\, w(2L - n)$. Since $w(2L - n)$ is nonzero over its duration, it follows that $y(n) = x(n)$ in the duration of $w(2L - n)$.

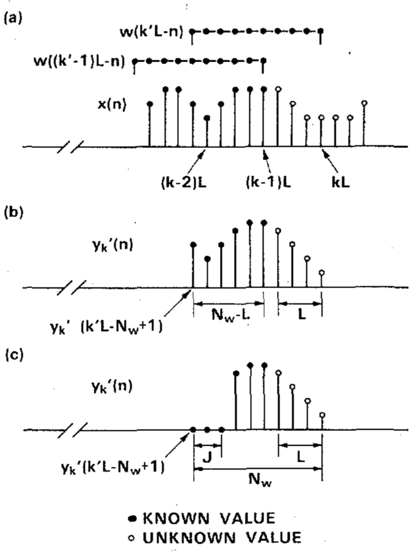


Fig. 3. Sequences for proof of Conditions 2.

We have now shown that $y(n) = x(n)$ is the only solution to (14) for $n \leqslant 2L$. We will next show that if $y(n) = x(n)$ is a unique solution up to $n = (k' - 1)L$, then $y(n) = x(n)$ is a unique solution up to $n = k'L$. By induction, we shall conclude that (14) has a unique solution $y(n) = x(n)$ for all n, under Conditions 2.

Consider the short-time section $y_{k'}(n) = x(n)\, w(k'L - n)$ for a particular $k = k'$. Suppose that $y(n) = x(n)$ is a unique solution to (14) for $n \leqslant (k' - 1)L$. Then beginning at $n = k'L - N_w + 1$ (see Fig. 3), $N_w - L$ consecutive samples of $y_{k'}(n)$ are also uniquely specified. Our objective now is to show that the last L samples of $y_{k'}(n)$ also have a unique solution. Clearly, the ability to do so depends on the value of L.

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

Suppose $1 < L \leqslant \lfloor N_w/2 \rfloor$. Furthermore, suppose that the initial value $y_{k'}(k'L - N_w + 1)$ is nonzero. The $N_w - L$ values of $y_{k'}(n)$ starting from $n = k'L - N_w + 1$ are given uniquely. Since $N - L \geqslant \lceil N_w/2 \rceil$ and $S_w(k'L, \omega)$ is known, $y_{k'}(n)$ is uniquely determined for all n by using Theorem 2. Now consider the cases when the first nonzero 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 $n = k'L - N_w + 1$ [see Fig. 3(c)]. Let us find the largest J for which the L unspecified samples of $y_{k'}(n)$ can be uniquely determined. Theorem 2 requires at least L known samples preceding the L unknown samples. Thus, the maximum allowable value of J is $N_w - [L + L] = N_w - 2L$. This is consistent with Conditions 2.

C. Multidimensional Extension

This section extends signal representation with STFT magnitude to multidimensional discrete-time signals. Since the ex-

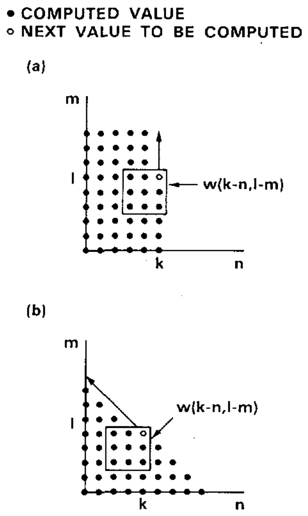


Fig. 4. Two-dimensional reconstruction procedures.

tension is conceptually straightforward but notationally cumbersome, it will be presented here only for the STFT magnitude with maximum analysis window overlap and for two-dimensional signals with finite support. For a two-dimensional signal $x(m, n)$, the STFT magnitude is given by

$$S_w(m, n; \omega, \nu) = \left| \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} x(m_1, m_2) \cdot w(n - m_1, m - m_2)\, e^{-j\omega m_1}\, e^{-j\nu m_2} \right| \quad (15)$$

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

Let $[x(m, n)]_N$ represent the class of two-dimensional signals whose finite regions of support contain no blocks of zeros larger than $(N - 2) \times (N - 2)$. 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 unique representation.

Conditions 3–For Representing $x(m, n)$ Uniquely with $S_w(m, n; \omega, \nu)$:

$w(m, n)$: a) Known and nonzero over its $N \times N$ rectangular support

$x(m, n)$: a) Belongs to $[x(m, n)]_N$
b) Sign of one nonzero sample known.

The derivation of these conditions is analogous to those used for one-dimensional signals. In particular, sequential procedures can be easily designed in a manner similar to the sequential extrapolation procedures based on the theorems in Section II. One such procedure for obtaining $x(m, n)$ proceeds along successive columns (rows). Suppose, in particular, that $x(m, n)$ has been computed up to the $(k - 1)$th column and $(r - 1)$th row [see Fig. 4(a)]. 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 Fig. 4(b).

IV. Signal Reconstruction from Unmodified Short-Time Fourier Transform Magnitude

We have established a number of conditions under which a signal is uniquely represented by its STFT magnitude. However, for such a signal representation to be practical, we need techniques that *reconstruct* a signal from its STFT magnitude. In Section III, we implicitly introduced one such technique while developing conditions for unique signal correspondence with the STFT magnitude. That technique belongs to a more general class of techniques 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 Fourier transform magnitude. Two theorems were presented in Section II for such extrapolation and were used in demonstrating Conditions 1 and 2. However, in those cases only a portion of the autocorrelation of each short-time section was used to perform the extrapolation. In Sections IV-B and IV-C, we consider techniques which use more samples of the autocorrelation of each short-time section. In particular, we develop techniques which require the extrapolated short-time section to match the autocorrelation using various error criteria. This is particularly useful when the known information is not exact. For example, in Section IV-D we will see that the extrapolation techniques of Sections IV-B and IV-C are less sensitive to roundoff errors when compared to the extrapolation technique which follows straightforwardly from the theorems of Section II. Furthermore, in Section V, we will see that these techniques give better signal reconstructions when the STFT magnitude is purposely modified for accomplishing signal processing tasks such as noise reduction and time-scale modification of speech. Finally, Section IV-E discusses an alternative reconstruction approach that is referred to as *simultaneous extrapolation.*

A. The Sequential Extrapolation Approach

For reconstruction of $x(n)$ from $S_w(nL, \omega)$, 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 $0 \leqq n < N_w$. The signal $x(n)$ has no more than $N_w - 2L$ consecutive zeros separating any two nonzero samples. 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 $0 \leqq n < L$ are known. Note that if $L = 1$, just the sign of $x(0)$ would be sufficient according to Conditions 1.

The sequential extrapolation approach to signal reconstruction from STFT magnitude is illustrated in Fig. 5. 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 Fourier transform magnitude and its known samples 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.

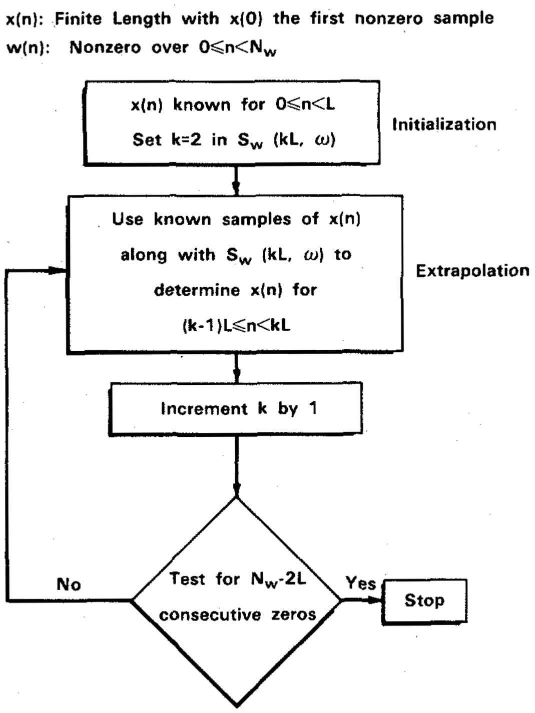

x(n): Finite Length with x(0) the first nonzero sample
w(n): Nonzero over 0≤n<N_w
x(n) known for 0≤n<L
Set k=2 in S_w (kL, ω)
Initialization
Use known samples of x(n) along with S_w (kL, ω) to determine x(n) for (k-1)L≤n<kL
Extrapolation
Increment k by 1
Test for N_w-2L consecutive zeros
No
Yes
Stop

Fig. 5. Sequential extrapolation approach.

The extrapolation procedures we are concerned with in this paper generally require certain samples of the autocorrelation function of each short-time section. As shown in the proof of Theorem 1, this autocorrelation function can be obtained by applying a long enough discrete Fourier transform (DFT) to the Fourier transform magnitude of the appropriate short-time section. The reconstruction procedures yield the unknown samples of the short-time section $f_n(m) = x(m)\, w(nL - m)$. Since the analysis window is nonzero over the length of each short-time section, dividing the short-time section by the analysis window yields the samples of the signal $x(n)$ which are then used in the extrapolation of the next short-time section $f_{n+1}(m)$. The reconstruction stops when a short-time section is encountered for which the known samples are not sufficient to complete the extrapolation. For finite-length signals, the reconstruction stops only after all the nonzero short-time sections have been extrapolated.

B. Least Squares Sequential Extrapolation

In this section we develop a least squares approach for the sequential extrapolation step of the signal reconstruction procedure of the previous section (see Fig. 5). The major idea here is to use more information from the STFT magnitude than is strictly necessary to reconstruct the signal. This makes the reconstruction algorithm more robust to errors in the short-time Fourier transform magnitude as will be seen in Section IV-D and Section V.

Let $f(n)$ be the short-time section being extrapolated. For simplifying notation, we assume that $f(0)$ is the first nonzero sample of $f(n)$. However, the technique developed here is not affected by the particular location of the first nonzero sample. Assume that the analysis window is N_w points long and thus $f(n)$ is known to be zero for $n \geqq N_w$. In the sequential extrapolation approach, outlined in Fig. 5, the known samples of $f(n)$ are in the range $0 \leqq n < M$ where $M \geqq \lceil N_w/2 \rceil$. The problem is to extrapolate the unknown samples of $f(n)$ in the range $M \leqq n < N_w$. For this we use a least squares algorithm that minimizes

$$E = \sum_{m=-\infty}^{\infty} (r(m) - s(m))^2 \tag{16}$$

where $r(m)$ is the autocorrelation function obtained by taking the inverse Fourier transform of the squared Fourier transform magnitude of $f(n)$. The function $s(m)$ represents the inverse Fourier transform of the squared Fourier transform magnitude of the reconstructed $f(n)$. By Parseval's theorem [8], minimizing the above expression is equivalent to minimizing the integral over the squared difference between the squared Fourier transform magnitude of $f(n)$ and the squared Fourier transform 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 (16) is equivalent to minimizing

$$E = 2 \sum_{m=1}^{N_w - 1} (r(m) - s(m))^2 + (r(0) - s(0))^2. \tag{17}$$

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:

$$f^3(N_w - 1) - (r(0) - 2t(0))\, f(N_w - 1) - \sum_{m=1}^{N_w - 1} (r(m) - t(m))\, f(N_w - 1 - m) = 0 \tag{18}$$

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. When there are three real solutions, the unique solution (guaranteed by the theory) is found by explicitly checking for the root that minimizes E.

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 (17). If there are L unknowns and we neglect the terms for $0 \leqq m < L$, 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 $f(N_w - 1)$:

$$f(N_w - 1) = \frac{\displaystyle\sum_{m=1}^{N_w - 1} (r(m) - t(m))\, f(N_w - 1 - m)}{\displaystyle\sum_{m=1}^{N_w - 1} f^2(N_w - 1 - m)} \tag{19}$$

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

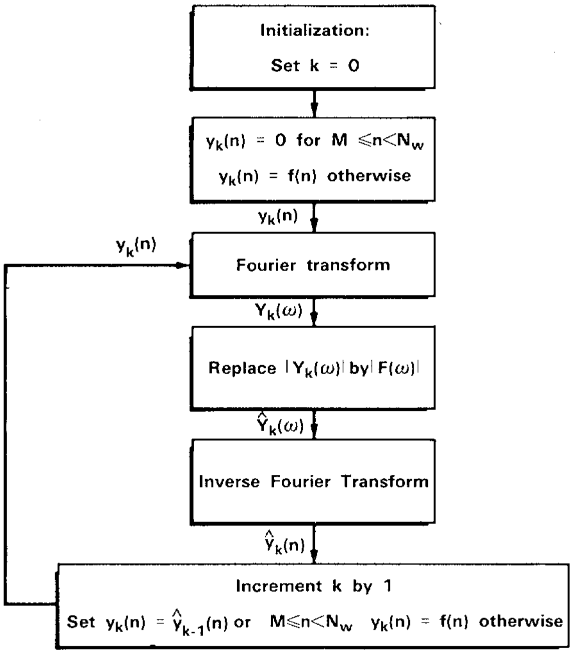

Initialization:
Set k = 0
$y_k(n) = 0$ for $M \leq n < N_w$
$y_k(n) = f(n)$ otherwise
$y_k(n)$
$y_k(n)$
Fourier transform
$Y_k(\omega)$
Replace $|Y_k(\omega)|$ by $|F(\omega)|$
$\hat{Y}_k(\omega)$
Inverse Fourier Transform
$\hat{y}_k(n)$
Increment k by 1
Set $y_k(n) = \hat{y}_{k-1}(n)$ or $M \leq n < N_w$ $y_k(n) = f(n)$ otherwise

Fig. 6. Iterative extrapolation.

C. 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 sequential extrapolation step (see Fig. 5) in signal reconstruction from STFT 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 sections, this algorithm proves to be very robust to errors in the STFT magnitude information.

Following the notation of Section IV-B, let $f(n)$ be the short-time section being extrapolated. For simplifying notation, we assume that $f(0)$ is the first nonzero sample of $f(n)$. Assuming the analysis window is N_w samples long, $f(n)$ is known to be zero for $n \geqq N_w$. In the sequential extrapolation approach outlined in Fig. 5, the known samples of $f(n)$ are in the range $0 \leqq n < M$ where $M \geqq [N_w/2]$. Fig. 6 presents an iterative technique that alternates between the time and frequency domains, imposing the known constraints in each domain. The constraints imposed in the time domain are all the known samples of $f(n)$ outside the region $M \leqq n < N_w$. 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 desired signal for the unknown samples of $f(n)$ in the region $M \leqq n < N_w$. This algorithm is similar in style to other iterative algorithms based on the known Fourier transform magnitude and certain time-domain information [11].

The problem of determining whether the iterative procedure outlined in Fig. 6 converges has not been addressed in this paper. However, we have empirically observed that the procedure converges to the desired signal in many cases. In other instances, however, the procedure appears to converge but not to the samples we seek. In Section IV-D we will see that for signal reconstruction from STFT magnitude, the failure to converge to the desired signal in some of the short-time sections leads to a reconstructed signal different from the original. On the other hand, for speech signals, the reconstruction retains the quality[1] of the original.

D. Reconstruction Examples

This section presents results of experiments conducted on speech signals to test the reconstruction algorithms discussed above. In particular, we have tested the algorithms on the STFT magnitude of speech waveforms such as the one shown in Fig. 7. This waveform corresponds to the sentence "The bowl dropped from his hand," spoken by a female speaker. The processing was carried out on a PDP 11/50 with floating point arithmetic. For this processing, the waveform is sampled at a 10 kHz rate and quantized with 12 bits/sample.

In the first experiment, the goal was to reconstruct the signal from $S_w(n, \omega)$ using the direct sequential extrapolation approach based on the proof of Theorem 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. The direct approach was applied with rectangular as well as Hamming analysis windows of various lengths up to 128 points. Using double position (64 bits) floating point computation, the reconstruction was exact to within the 12 bit precision of the original speech signal of Fig. 7. For the case of a rectangular window of 128 points, the reconstruction from $S_w(n, \omega)$ is shown in Fig. 8. Signal reconstruction was also exact for the cases when the analysis window spacing L was slightly larger than unity. In these cases, we applied the direct sequential extrapolation procedure based on the proof of Theorem 2 of Section II. However, when the analysis window overlap was greater than four, the reconstructed speech was not intelligible. The failure appears to occur due to roundoff errors in the algorithm implementation. We have observed that once such an error is made at any point in a direct reconstruction, the remainder of the reconstruction is usually bad enough to render the reconstructed sentence unintelligible.

We next implemented signal reconstruction using the linear version of the sequential least squares technique of Section IV-B. The analysis window was a 128 point rectangular window. With double precision, the reconstruction was exact for small values of L. However, as L increased, the reconstruction was not exact, due to roundoff errors, but retained the speech quality[1] of the original up to around $L = 30$. As L approaches $N_w/2 = 64$, there are not enough autocorrelation coefficients to make the computation robust to roundoff errors. Consequently, for such large values of L (e.g., above 40) the reconstructions degraded rapidly and became unintelligible.

We applied the sequential iterative algorithm to reconstruct speech from STFT magnitudes with 128 point analysis windows. This algorithm generally does not reconstruct the signal exactly. However, for large values of L (e.g., above 40) the reconstructed signal retains the speech quality of the

[1] Speech quality comparisons in this paper are based on informal listening by the authors.

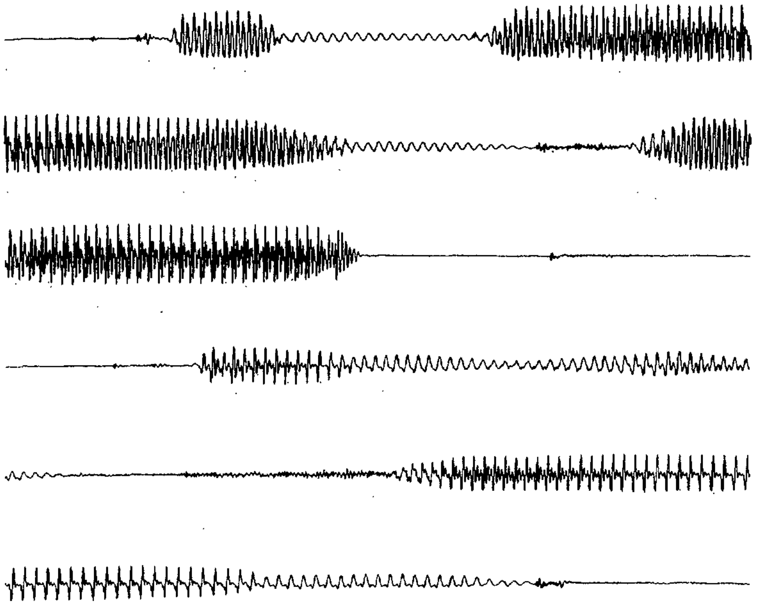


Fig. 7. Test speech waveform, "The bowl dropped from his hand."

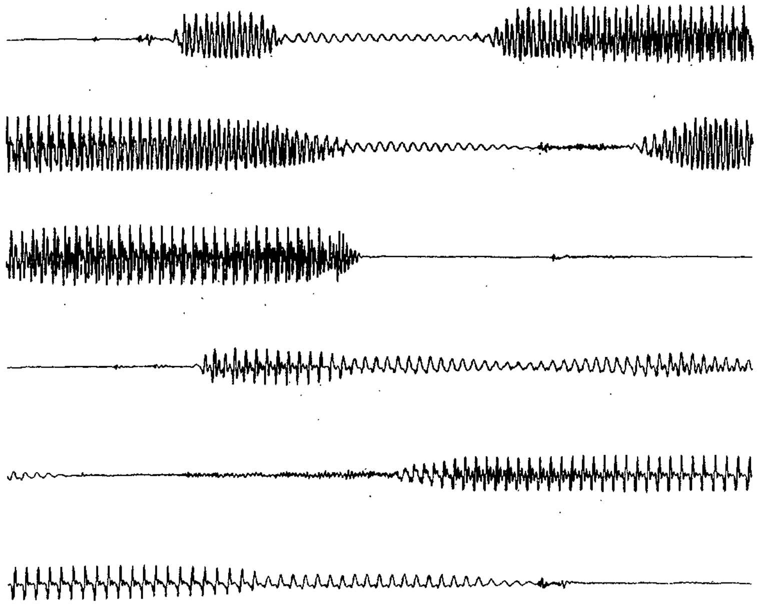


Fig. 8. Sequential reconstruction of waveform in Fig. 7 based on Theorem 1.

original. For example, Fig. 9 shows the reconstruction of the speech in Fig. 7 using the iterative reconstruction algorithm. The analysis window is a rectangular window of 128 points and the window spacing L is 64.

Finally, we tested the sensitivity of these algorithms to errors in the L initial samples which are assumed known. The behavior of these algorithms in the presence of such errors was similar to the one observed for roundoff errors. In particular, regardless of errors in the initial samples, the iterative technique for a 128 point Hamming window produced speech of the same quality.

E. Simultaneous Extrapolation Approach

The emphasis in this paper is on the sequential extrapolation algorithms of the previous sections for signal reconstruction from STFT magnitude. However, other approaches can be designed for reconstructing a signal from its STFT 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 section. However, in case of errors or purposeful modifications in the STFT magnitude, we have seen previously that it is useful to incorporate extra information in the reconstruction procedures. For example, the

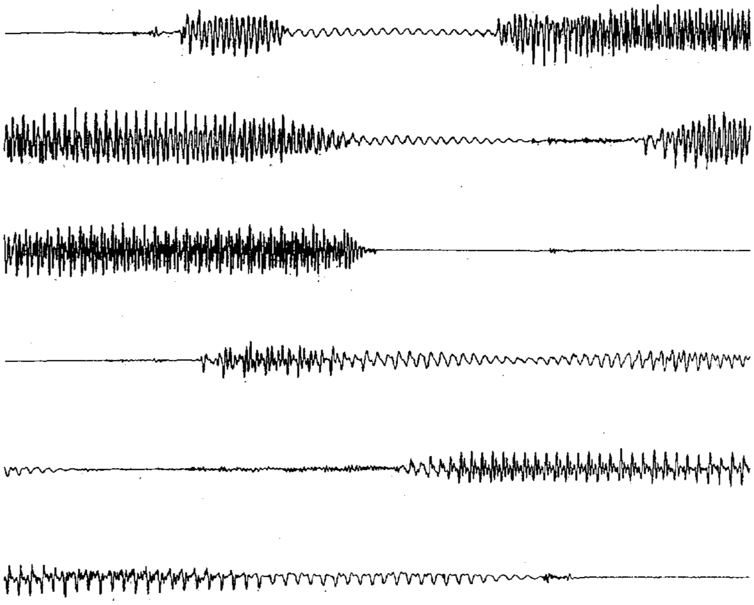


Fig. 9. Sequential iterative reconstruction of waveform in Fig. 7.

least squares and iterative techniques of the previous section used much more of the autocorrelation function of each short-time section than the techniques based on the proofs of the theorems in Section II. In the simultaneous extrapolation approach, we 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 discussing an extension to the least squares technique of Section IV-B. The problem is to reconstruct a finite-length signal $x(n)$ from its STFT magnitude, $S_w(nL, \omega)$, under the conditions developed in Section III. In Section IV-B 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 IV-B 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 preceding in time. Clearly, such a solution neglects the structure of the STFT that is contained in the overlap of any particular short-time section with the short-time section that follows it. This structure is important to exploit when there are errors in the STFT magnitude. In fact, the structure of the STFT 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.

V. Signal Estimation from Modified Short-Time Fourier Transform Magnitude

A number of signal processing applications require signal estimation from a modified STFT. It is of interest to determine whether such signal estimates can be obtained from only the magnitude of the modified STFT. In this section, we will consider two types of STFT modification which are used for time-scale modification and noise reduction in speech. In each case, we have found that the iterative reconstruction algorithm of Section IV yields signal estimates with the desired characteristics. Furthermore, this magnitude-only approach to STFT signal processing has potential for further improvements as simultaneous extrapolation techniques (see Section IV) are explored in the future.

Since the modified STFT magnitude is generally not a valid STFT magnitude [3], [4], any algorithm that relies critically on the validity of the STFT magnitude performs poorly. This is the case, for example, in the sequential extrapolation algorithms that use the extrapolation techniques in the proofs of Theorems 1 and 2 of Section II. 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 Fourier transform magnitude is unmodified, the remaining portion of the section's autocorrelation is also consistent with the extrapolation. However, if the Fourier transform 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 section were designed for this purpose. Therefore, we will use the same techniques for signal estimation from modified STFT magnitude.

A. Time-Scale Modification

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

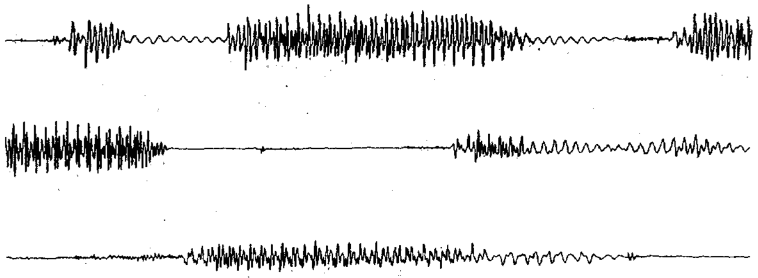


Fig. 10. Time-scale compression of 2:1 of waveform in Fig. 7.

preserving the instantaneous frequency locations while changing their rate of change in time. For such processing of speech by a factor of β, a modification [7] that has been applied to the STFT, $X_w(nL, \omega)$, is given by

$$Y_w(nL, \omega) = X_w(\beta nL, \omega)$$

where $Y_w(nL, \omega)$ is the modified STFT. Our goal is to estimate the time-scale modified signal from only the magnitude of $Y_w(nL, \omega)$. We have implemented this strategy for 2:1 time compression in sentences such as that shown in Fig. 7. For signal estimation from the magnitude of $Y_w(nL, \omega)$ we found the iterative technique to be the best among the techniques of Section IV. We used a 128 point Hamming window with window spacing $L = 32$. The resulting time-compressed waveform corresponding to Fig. 7 is shown in Fig. 10. Clearly, the duration of the waveform has been cut by half. Furthermore, the pitch of the various segments is preserved even though segment duration is shortened. Informal listening indicates that the processed speech retained its natural quality and speaker-dependent features and was free from artifacts such as "burbles" and reverberation. However, there was a small amount of background "crackle," which we anticipate will be eliminated with better reconstruction techniques based on the simultaneous extrapolation ideas of Section IV-E.

B. Noise Reduction

A number of STFT processing techniques have been developed over the years for the reduction of additive noise in speech [9] and image [2] signals. 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.

The STFT implementation of the standard spectral subtraction technique is based on the interpretation of the STFT magnitude square as a time-varying power spectrum. If $x(n)$ is the signal corrupted by additive random noise $e(n)$ and $S_w(nL, \omega)$ is the STFT magnitude of $x(n)$, the spectral subtraction procedure yields the following function:

$$M_w^2(nL, \omega) = \begin{cases} S_w^2(nL, \omega) - \alpha P_e(\omega) & \text{if } S_w^2(nL, \omega) > \alpha P_e(\omega) \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where the parameter α serves as a control for the degree of noise smoothing to be achieved and $P_e(\omega)$ represents the power spectrum of the background noise. In practice, it has been found that values of α between 2 and 3 produce acceptable results [2], [9]. We can obtain a signal estimate directly from the modified STFT magnitude, $M_w(nL, \omega)$. For such signal estimation algorithms, we require *a priori* knowledge of L consecutive samples of $x(n)$, starting from the first nonzero sample. Our approach is to use some reasonable estimate for those samples. For example, one approach is to use the corresponding L samples of the noisy signal $x(n)$.

In our experiments with magnitude-only short-time spectral subtraction, we have applied the sequential iterative technique of Section IV for the signal estimation from $M_w(nL, \omega)$. We selected this particular technique because of its simple implementation requirements. Furthermore, it performs well compared to the other sequential reconstruction techniques that we have tested on speech. We used a 128 point Hamming window and a window spacing of $L = 64$. Speech sentences were corrupted by the addition of stationary white noise with a variety of signal-to-noise ratios between 0 dB and 20 dB. We found that for signal-to-noise ratios above 10 dB, the signal estimates from the modified STFT magnitude had a reduced noise level while retaining their natural speech quality and speaker dependent features. The only processing artifact was the presence of short tone-bursts of varying frequency in the background. This artifact has always been associated with short-time spectral subtraction [9].

VI. Summary

In this paper, we have shown that large classes of signals are uniquely representable with the STFT 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 STFT magnitude. These algorithms include some that are designed to yield reasonable signal estimates from a processed STFT magnitude which does not correspond to the STFT magnitude of any signal. To illustrate the practical usefulness of the results in this paper, we considered the problems of time-scale modification and noise reduction in speech.

In this paper, we have presented only the results that we considered most important. There exist additional results, such as other conditions under which a signal is uniquely specified by its STFT magnitude and alternative proofs to the theoretical results presented in this paper. These additional results and more extensive discussions can be found in [10].

Acknowledgment

We would like to thank Prof. A. V. Oppenheim for invaluable discussons that shaped the direction of this research and for his comments on the manuscript at various stages. We would also like to thank the reviewers for their careful scrutiny of a lengthy paper.

References

[1] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[2] J. S. Lim, "Image restoration by short-space spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 191-197, Apr. 1980.

[3] C. J. Weinstein, "Short-time Fourier analysis and its inverse," S.M. thesis, M.I.T., Cambridge, MA, 1966.

[4] M. R. Portnoff, "Representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 55-69, Feb. 1980.

[5] J. B. Allen, "Short-term spectral analysis and synthesis and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, no. 3, pp. 235-238, June 1977.

[6] R. A. Altes, "Detection estimation and classification with spectrograms," *J. Acoust. Soc. Amer.*, vol. 67, pp. 1232-1246, Apr. 1980.

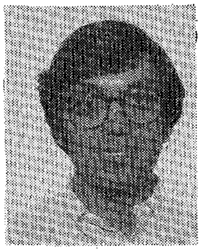
[7] M. R. Portnoff, "Time-scale modification of speech based on short-time Fourier analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 374-390, June 1981.

[8] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[9] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586-1604, Dec. 1979.

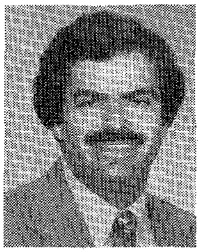
[10] S. H. Nawab, "Signal estimation from short-time spectral magnitude," Ph.D. dissertation, M.I.T., Cambridge, MA, 1982.

[11] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237-246, Nov. 1972.



S. Hamid Nawab was born on January 13, 1955 in Swindon, England. After early schooling mostly in Pakistan, he received the S.B., S.M., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1977, 1979, and 1982, respectively.

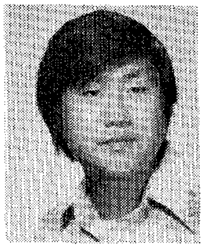
During his graduate studies he was a Research Assistant at the M.I.T. Research Laboratory of Electronics (1977-1979) and at the M.I.T. Lincoln Laboratory, Lexington, MA (1979-1982). His graduate research included work on fast DFT algorithms, image enhancement, and signal reconstruction from partial transform information. In June 1982 he was appointed Staff Member at M.I.T. Lincoln Laboratory, where his research is concerned with the evaluation and design of signal processing systems for acoustic wave analysis.



Thomas F. Quatieri (S'73-M'79) was born in Somerville, MA, on January 31, 1952. He received the B.S. degree (summa cum laude) from Tufts University, Medford, MA, in 1973, and the S.M., E.E., and Sc.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1975, 1977, and 1979, respectively.

From 1973 to 1975 he was a Teaching Assistant and from 1975 to 1979 a Research Assistant in the area of digital signal processing, both with the Department of Electrical Engineering of M.I.T. His research for the Master's degree involved the design of two-dimensional digital filters and for the Sc.D. involved phase estimation with application to speech analysis-synthesis. He is presently a Research Staff Member at the M.I.T. Lincoln Laboratory, Lexington, MA, where he is working on problems in digital signal processing with applications to speech and image processing.

Dr. Quatieri is the recipient of the 1982 Paper Award of the IEEE Acoustics, Speech, and Signal Processing Society, and is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.



Jae S. Lim (S'76-M'78) was born on December 2, 1950. He received the S.B., S.M., E.E., and Sc.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1974, 1975, 1978, and 1978, respectively.

He joined the M.I.T. faculty in 1978 as an Assistant Professor and is currently Associate Professor in the Department of Electrical Engineering and Computer Science. His research interests include digital signal processing and its applications to image and speech processing. He has contributed more than 50 articles to journals and conference proceedings, and is the editor of a reprint book, *Speech Enhancement* (Englewood Cliffs, NJ: Prentice-Hall, 1982).

Dr. Lim is the winner of two prize paper awards, one from the Boston Chapter of the Acoustical Society of America in December 1976, and one from the IEEE Acoustics, Speech, and Signal Processing Society in April 1979. He is a member of Eta Kappa Nu, Sigma Xi, and the IEEE Technical Committee on Digital Signal Processing.