Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis

MICHAEL R. PORTNOFF, MEMBER, IEEE

Abstract—This paper develops a representation for discrete-time signals and systems based on short-time Fourier analysis. The short-time Fourier transform and the time-varying frequency response are reviewed as representations for signals and linear time-varying systems. The problems of representing a signal by its short-time Fourier transform and synthesizing a signal from its transform are considered. A new synthesis equation is introduced that is sufficiently general to describe apparently different synthesis methods reported in the literature. It is shown that a class of linear-filtering problems can be represented as the product of the time-varying frequency response of the filter multiplied by the short-time Fourier transform of the input signal. The representation of a signal by samples of its short-time Fourier transform is applied to the linear filtering problem. This representation is of practical significance because there exists a computationally efficient algorithm for implementing such systems. Finally, the methods of fast convolution are considered as special cases of this representation.

INTRODUCTION

THE Fourier transform plays a fundamental role in the analysis of signals and linear time-invariant systems. The efficacy of the Fourier transform is a result of its providing a unique representation for signals in terms of the eigenfunctions of linear time-invariant systems, namely the complex exponentials. The essentials of this representation are summarized by the following well-known results from the theory of linear time-invariant systems (see, for example, [1]-[3]).

If $t(n)$ denotes the unit-sample (impulse) response of a linear time-invariant system, then the response $y(n)$ of the system to the input $x(n)$ is given by the convolution sum

$$y(n) = \sum_{m=-\infty}^{\infty} t(n-m)\,x(m) = \sum_{m=-\infty}^{\infty} t(m)\,x(n-m). \tag{1}$$

If $x(n)$ is the complex exponential $\exp[j\omega n]$ then (1) gives

$$\begin{aligned} y(n) &= \sum_{m=-\infty}^{\infty} t(m) \exp[j\omega(n-m)] \\ &= \left\{ \sum_{m=-\infty}^{\infty} t(m) \exp[-j\omega m] \right\} \exp[j\omega n] \end{aligned}$$

or,

$$y(n) = T(\omega) \exp[j\omega n] \tag{2}$$

where $T(\omega)$, the frequency response of the system, is the Fourier transform of the unit-sample response given by

$$T(\omega) = \sum_{m=-\infty}^{\infty} t(m) \exp[-j\omega m]. \tag{3}$$

Suppose $x(n)$ is now a general signal that can be expressed as the Fourier integral

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) \exp[j\omega n]\, d\omega \tag{4}$$

where $X(\omega)$ denotes the Fourier transform of $x(n)$ given by

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n) \exp[-j\omega n]. \tag{5}$$

Equations (4) and (5) provide a unique correspondence between $x(n)$ and $X(\omega)$ and either one is an equally valid representation of the signal. The Fourier transform, however, is a particularly convenient representation for a signal to be processed by a linear time-invariant system because the basis functions of the Fourier transform are the eigenfunctions of linear time-invariant systems. Specifically, since the Fourier integral (4) is, in essence, a linear combination of complex exponentials, and since the system $t(n)$ is linear, the response of $t(n)$ to the input (4) is

$$y(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} T(\omega)\, X(\omega) \exp[j\omega n]\, d\omega \tag{6}$$

(assuming, of course, the sum (3) converges), and the product

$$Y(\omega) = T(\omega)\, X(\omega) \tag{7}$$

is just the Fourier transform of the response $y(n)$. Thus, the Fourier transform maps the convolution in the time domain to multiplication in the frequency domain. Furthermore, in addition to the Fourier transform being a powerful analytical technique, the property that its basis functions are the eigenfunctions of linear time-varying systems leads to a great deal of intuition, invaluable for solving signal processing problems.

The Fourier transform representation has several practical and conceptual limitations because it represents, for each frequency ω, the global (in time) characteristics of the signal. Consequently, the Fourier transform has the practical limita-

Manuscript received November 28, 1978; revised June 5, 1979.
The author was with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139. He is now with the Engineering Research Division, University of California, Lawrence Livermore Laboratory, CA 94550.

0096-3518/80/0200-0055$00.75

tion that the entire signal must be known in order to obtain its transform. Moreover, the Fourier transform does not provide an adequate representation for linear time-varying systems, nor does it always provide an intuitively meaningful representation for the output of such systems.

This paper is divided into two parts. Part I formulates a time-frequency representation for signals and linear time-varying systems that characterizes their local behavior in terms of complex exponentials. The representation of linear time-varying systems is based on the time-varying frequency response [4]-[7], which is a generalization of the frequency response (3) for linear time-invariant systems. The time-frequency representation for signals is based on the short-time Fourier transform [8]-[21], which is a formal representation for the output of a filter-bank spectrum analyzer or, equivalently, the usual Fourier transform of the signal viewed through a sliding time window. The results derived here are of interest both as a theoretical representation for time-varying systems and signals and because they provide techniques applicable to a variety of signal processing problems.

On a digital processor, in order to realize a signal processing algorithm based on the time-frequency representation discussed in Part I, the short-time Fourier transform and time-varying frequency response must be represented by a finite number of frequency samples. Moreover, to make the amount of computation tractable, the short-time Fourier transform must be decimated in time (down sampled) as well.

Part II of this paper extends the results of Part I to a sampled transform representation based on the discrete (sampled) short-time Fourier transform and the discrete time-varying frequency response. The development focuses on three problems. The first is the representation of a sequence in terms of samples of its short-time Fourier transform and the resynthesis of the original sequence without distortion. Of special interest is the problem of formulating such a representation with no redundancy, i.e., so that there is, on the average, one sample of the transform representation for each sample of the original signal. The second problem is the efficient implementation of the discrete short-time Fourier analysis and synthesis formulas based on the fast Fourier transform (FFT) algorithm. Because the short-time Fourier analysis and synthesis formulas do not have the form of discrete Fourier transforms (DFT) they cannot be computed directly with the FFT algorithm. The third, and final, problem considered here is the implementation of linear time-varying filtering as the product of the discrete short-time Fourier transform of the input signal multiplied by samples of the time-varying frequency response of the filter. For a class of linear time-varying filters, determined by the short-time Fourier analysis and synthesis filters, such an implementation is possible, and for linear time-invariant filters, this implementation reduces to the conventional overlap-save or overlap-add method of fast convolution [22], depending on the particular choice of the analysis and synthesis filters.

PART I

I. THE PARTIAL FOURIER TRANSFORM AND ITS INVERSE

The time-frequency representation presented in this paper represents one-dimensional signals by two-dimensional signals. Because Fourier transforms with respect to one, or the other, or both, of the indices of such two-dimensional sequences frequently arise in this context, the definition and notation for such transforms will now be formalized. Let $f(n, m)$ denote a two-dimensional discrete-time sequence. The *complete* Fourier transform of $f(n, m)$, denoted $F(\psi, \omega)$, is defined as the (usual) two-dimensional Fourier transform

$$F(\psi, \omega) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} f(n, m) \exp\left[-j(\psi n + \omega m)\right], \tag{8}$$

with inverse transform

$$f(n, m) = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} F(\psi, \omega) \cdot \exp\left[j(\psi n + \omega m)\right] d\psi\, d\omega. \tag{9}$$

The *partial* Fourier transform of $f(n, m)$, with respect to its first argument, denoted $F_1(\psi, m)$ is defined as the one-dimensional Fourier transform of $f(n, m)$ over n, that is,

$$F_1(\psi, m) = \sum_{n=-\infty}^{\infty} f(n, m) \exp\left[-j\psi n\right] \tag{10}$$

(where the subscript 1 is used to indicate that the first argument is the transform variable). Moreover, the inverse partial transform is given by

$$f(n, m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_1(\psi, m) \exp\left[j\psi n\right] d\psi. \tag{11}$$

Similarly, the partial Fourier transform $F_2(n, \omega)$ is defined as the one-dimensional Fourier transform of $f(n, m)$ with respect to its second argument, i.e.,

$$F_2(n, \omega) = \sum_{m=-\infty}^{\infty} f(n, m) \exp\left[-j\omega m\right] \tag{12}$$

and

$$f(n, m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_2(n, \omega) \exp\left[j\omega m\right] d\omega. \tag{13}$$

Finally, the *complete* Fourier transform $F(\psi, \omega)$ can be obtained by *successive partial* Fourier transforms with respect to each of the independent arguments of $f(n, m)$, that is,

$$F(\psi, \omega) = \sum_{m=-\infty}^{\infty} F_1(\psi, m) \exp\left[-j\omega m\right] \tag{14}$$

$$= \sum_{n=-\infty}^{\infty} F_2(n, \omega) \exp\left[-j\psi n\right]. \tag{15}$$

II. THE TIME-VARYING FREQUENCY RESPONSE

The input-output behavior of a linear time-varying system can be characterized in the time domain by a *weighting pattern*, or *Green's function*, $g(n, m)$, which represents the response of the system at time n to a unit sample applied at time m. Equivalently, the same system can be described by a *time-varying unit-sample response* $t(n, m)$ defined as the response of the system at time n to a unit sample applied m samples earlier, i.e, at time $(n - m)$. Furthermore, the time-varying unit-sample response $t(n, m)$ and the Green's function $g(n, m)$ are related by

$$t(n, m) = g(n, n - m) \tag{16}$$

or, equivalently,

$$g(n, m) = t(n, n - m). \tag{17}$$

If $y(n)$ is the response of a system to the input $x(n)$, then $y(n)$ is given by the superposition sum

$$y(n) = \sum_{m=-\infty}^{\infty} g(n, m)\, x(m) \tag{18}$$

or

$$y(n) = \sum_{m=-\infty}^{\infty} t(n, m)\, x(n - m) = \sum_{m=-\infty}^{\infty} t(n, n - m)\, x(m). \tag{19}$$

If the system represented by $g(n, m)$ is time-invariant, then $g(n, m)$ depends only on the difference $(n - m)$ corresponding to the number of samples between the application of the unit sample and the observation of the output; thus,

$$g(n, m) = g(n - m). \tag{20}$$

From the relation (16), the time-varying unit-sample response $t(n, m)$ for a linear time-invariant system becomes

$$\begin{aligned} t(n, m) &= g(n, n - m) = g(n - (n - m)) = g(m) \\ &= t(m) \end{aligned} \tag{21}$$

and corresponds to the ordinary unit-sample response of such a system. Conversely, if $t(n, m)$ is independent of n, then the system represented by $t(n, m)$ is time-invariant. Moreover, if $t(n, m)$ is a "slowly-varying" function of n, then the system represented by $t(n, m)$ will be said to be slowly time-varying. The notion of such a slowly time-varying system is, in general, imprecise and must be considered in the context of a particular set of assumptions about the system or the signals to be processed by the system.

Because the time-varying unit-sample response is a characterization of a system relative to a sliding time frame, and because it is slowly varying function of n for slowly varying systems, the time-varying unit-sample response is more convenient than the weighting pattern in the context of short-time analysis. For the remainder of this paper, therefore, the time-varying unit-sample response will be employed exclusively and referred to, simply, as the "unit-sample response."

If the input $x(n)$ to a linear time-varying system with unit-sample response $t(n, m)$ is the complex exponential $\exp[j\omega n]$, then the resulting output is

$$\begin{aligned} y(n) &= \sum_{m=-\infty}^{\infty} t(n, m)\, x(n - m) \\ &= \sum_{m=-\infty}^{\infty} t(n, m) \exp[j\omega(n - m)] \\ &= \sum_{m=-\infty}^{\infty} t(n, m) \exp[-j\omega m] \exp[j\omega n], \end{aligned}$$

or

$$y(n) = T_2(n, \omega) \exp[j\omega n] \tag{22}$$

where

$$T_2(n, \omega) = \sum_{m=-\infty}^{\infty} t(n, m) \exp[-j\omega m]. \tag{23}$$

$T_2(n, \omega)$, the partial Fourier transform of $t(n, m)$ with respect to m, is interpreted according to (22) as the *time-varying frequency response* of the system with unit-sample response $t(n, m)$. For simplicity, $T_2(n, \omega)$ will often be referred to, simply, as the "frequency response" of $t(n, m)$.

If $X(\omega)$ is the Fourier transform of an arbitrary input $x(n)$, then the response $y(n)$ of $t(n, m)$ can be expressed as the inverse partial Fourier transform of the product of $X(\omega)$ with $T_2(n, \omega)$, that is,

$$\begin{aligned} y(n) &= \sum_{m=-\infty}^{\infty} t(n, m)\, x(n - m) \\ &= \sum_{m=-\infty}^{\infty} t(n, m) \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) \exp[j\omega(n - m)]\, d\omega \right\} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \sum_{m=-\infty}^{\infty} t(n, m) \exp[-j\omega m] \right\} \\ &\quad \cdot X(\omega) \exp[j\omega n]\, d\omega \end{aligned}$$

$$y(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} T_2(n, \omega)\, X(\omega) \exp[j\omega n]\, d\omega. \tag{24}$$

Equation (24) is the generalization, for linear time-varying systems, of (6) for linear time-invariant systems. In contrast to the case of linear time-invariant systems, however, it is not generally true that the time-varying frequency response of the cascade combination of linear time-varying systems is equal to the product of the corresponding individual time-varying frequency responses. In fact, there exists no such scalar-valued function with this property because the input-output behavior of such a cascade combination of systems depends on the order in which the systems are cascaded.

For the case of linear time-invariant systems, $t(n, m)$ and, hence, $T_2(n, \omega)$ are independent of n. Thus, $t(n, m) = t(m)$ and $T_2(n, \omega) = T(\omega)$ are the ordinary unit-sample response and frequency response for such a system.

III. Short-Time Fourier Analysis and Synthesis

The usual short-time Fourier transform representation for a discrete-time signal $x(n)$ is given by the pair of equations [10]

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X_2(n, \omega) \exp[j\omega n]\, d\omega \tag{25}$$

$$X_2(n, \omega) = \sum_{m=-\infty}^{\infty} h(n - m)\, x(m) \exp[-j\omega m] \tag{26}$$

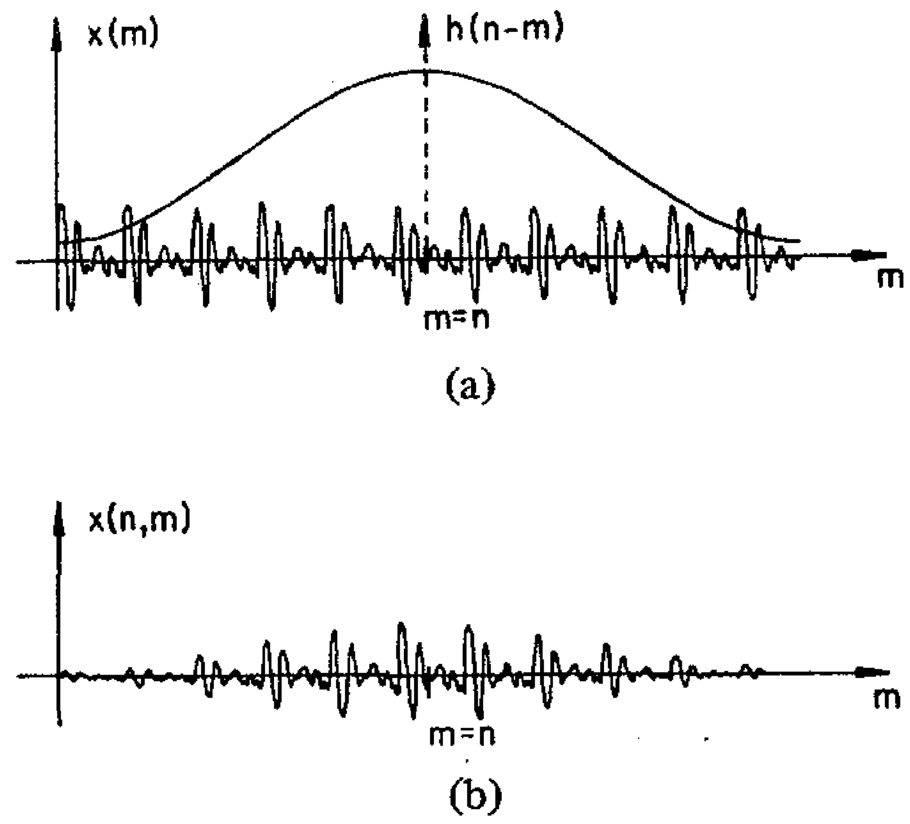


Fig. 1. (a) Time-reversed and shifted analysis window $h(n-m)$ superimposed on data $x(m)$. (b) Short-time sequence $x(n,m) = h(n-m)x(m)$ for a particular value of n.

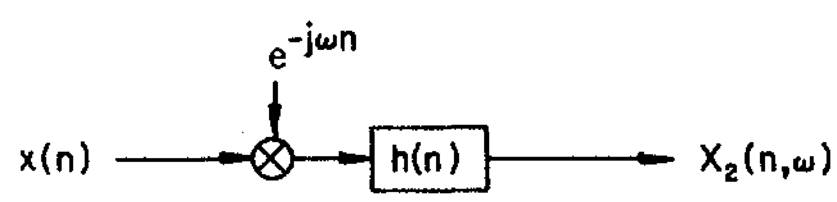


Fig. 2. Short-time Fourier transform as output of a demodulator followed by an analysis filter.

where, using the notation of Section I, $X_2(n, \omega)$ denotes the short-time Fourier transform of $x(n)$. $h(n)$ is referred to as the *analysis window* and is generally chosen to have the property that it is, in some sense, narrow in time, or frequency, or both, and is normalized such that $h(0) = 1$. Equation (25) is similar in form to the ordinary Fourier synthesis relation (4) except that $X_2(n, \omega)$ is now a function of the time index n and represents only the local behavior of $x(m)$ as viewed through the sliding window $h(n - m)$. Referring to Fig. 1, $X_2(n, \omega)$ can be interpreted for each value of n as the partial Fourier transform, with respect to m,

$$X_2(n, \omega) = \sum_{m=-\infty}^{\infty} x(n, m) \exp[-j\omega m] \tag{27}$$

of the "short-time function"

$$x(n, m) = h(n - m)\, x(m). \tag{28}$$

Equivalently, by considering (26) as the convolution

$$X_2(n, \omega) = h(n) *_n x(n) \exp[-j\omega n] \tag{29}$$

where $*_n$ denotes the convolution operator with respect to n, $X_2(n, \omega)$ can be interpreted as the output of a linear time-invariant filter $h(n)$, excited by the demodulated (frequency-shifted) signal $x(n) \exp[-j\omega n]$, as shown in Fig. 2. For this reason, $h(n)$ is also referred to as the *analysis filter*.

Because $X_2(n, \omega)$ is a function of the continuous variable ω for every value of n, the short-time Fourier transform contains redundant information about the signal, depending upon the particular analysis window used in (26). Furthermore, (26) imposes a structure on $X_2(n, \omega)$ so that not all functions of n and ω are valid short-time Fourier transforms.[1]

[1] Alternative viewpoints on the structure of the short-time Fourier transform are given in [18]-[21].

To illustrate the structure imposed on $X_2(n, \omega)$ by (26), observe that the inverse Fourier transform of $X_2(n, \omega)$ with respect to ω is the short-time function $x(n, m)$, which factors as the product of the signal multiplied by the shifted window, i.e.,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} X_2(n, \omega) \exp[j\omega m]\, d\omega = x(n, m)$$

$$= h(n - m)\, x(m). \tag{30}$$

Thus, not only can the signal $x(n)$ be recovered from the short-time Fourier transform by evaluating (30) for $n = m$, but the analysis window $h(n)$ can also be recovered, to within the multiplicative constant $x(0)$, by evaluating (30) for $m = 0$. In addition to the convolutional structure of (26), $X_2(n, \omega)$ also exhibits a convolutional structure when expressed in terms of the Fourier transforms of $h(n)$ and $x(n)$. Replacing $h(n - m)$ in (26) by its Fourier integral representation and simplifying gives $X_2(n, \omega)$ as the frequency domain convolution

$$X_2(n, \omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega + \psi) H(\psi) \exp[j\psi n]\, d\psi \tag{31a}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\psi) H(\psi - \omega) \exp[j(\psi - \omega) n]\, d\psi \tag{31b}$$

$$= \frac{1}{2\pi} X(\omega) *_\omega H(-\omega) \exp[-j\omega n]. \tag{31c}$$

Furthermore, the partial Fourier transform of $X_2(n, \omega)$ with respect to n is obtained by inspection of (31a) as

$$X(\psi, \omega) = H(\psi)\, X(\omega + \psi) \tag{32}$$

and also factors as the product of a function which depends only on the window and a function that depends only on the signal. Finally, $X(\psi, \omega)$ is recognized as the two-dimensional Fourier transform of the short-time function $x(n, m)$, that is,

$$X(\psi, \omega) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(n - m)\, x(m) \exp[-j(\psi n + \omega m)]$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} x(n, m) \exp[-j(\psi n + \omega m)]. \tag{33}$$

As a result of the mathematical structure of $X_2(n, \omega)$, (25) is not the only means for synthesizing $x(n)$ from $X_2(n, \omega)$. Equation (25) corresponds to inverse transforming $X_2(n, \omega)$ with respect to ω to obtain the short-time function (28), which is then evaluated at $m = n$ to give

$$x(n, m)\big|_{m=n} = h(0)\, x(n) = x(n) \quad \text{for} \quad h(0) = 1. \tag{34}$$

Alternatively, $x(m)$ could be obtained by again inverse transforming $X_2(n, \omega)$ to get $x(n, m) = h(n - m)\, x(m)$, but now, fixing $n = n_o$ and dividing by the shifted window $h(n_o - m)$, i.e.,

$$x(m) = (1/2\pi h(n_o - m)) \int_{-\pi}^{\pi} X_2(n_o, \omega) \exp[j\omega m]\, d\omega \quad (35)$$

$$= x(n_o, m)/h(n_o - m) = [h(n_o - m)x(m)]/h(n_o - m)$$

$$= x(m).$$

Clearly, for the particular value n_o, (35) is useful only for obtaining values of $x(m)$ where $h(n_o - m) \neq 0$. Another method of short-time Fourier synthesis [14]-[16] can be derived by evaluating $X(\psi, \omega)$, given by (32), for $\psi = 0$ to obtain

$$X(\omega) = X(0, \omega)/H(0)$$

$$= \frac{1}{H(0)} \sum_{n=-\infty}^{\infty} X_2(n, \omega). \quad (36)$$

Since $x(n)$ is the inverse Fourier transform of $X(\omega)$:

$$x(n) = (1/2\pi H(0)) \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} X_2(r, \omega) \exp[j\omega n]\, d\omega. \quad (37)$$

From the definition (26) of the short-time Fourier transform, its value at a particular time sample $n = n_o$ represents information not only about $x(n_o)$, but about all values of $x(n)$ "viewed" through the sliding (time) window $h(n_o - n)$. Similarly, from (31b), the short-time Fourier transform evaluated at a particular frequency $\omega = \omega_o$ contains information about all values of $X(\omega)$ viewed through the sliding (frequency) window $H(\omega_0 - \omega)$. Thus, the values of $X_2(n, \omega)$ are locally correlated in time and frequency. The synthesis formula (25) corresponds to the inverse partial Fourier transform of $X_2(n, \omega)$ with respect to ω, evaluated for $n = m$: for a particular value $n = n_o$, $x(n_o)$ is computed solely from the values of $X_2(n_o, \omega)$, ignoring the local correlation of the values of $X_2(n, \omega)$ in time. The synthesis formula (35) corresponds to the inverse Fourier transform of $X_2(n, \omega)$ evaluated at $n = n_o$, so that for all values of n, the values of $x(n)$ are computed from $X_2(n, \omega)$ evaluated only at the particular sample $n = n_o$. The synthesis formula (37) corresponds to the inverse Fourier transform of

$$X(0, \omega) = \sum_{n=-\infty}^{\infty} X_2(n, \omega).$$

Although this synthesis procedure utilizes information from adjacent time samples of $X_2(n, \omega)$, it simply sums over all n, giving equal weight to each value.

All of these synthesis formulas can be viewed in a more general framework by exploiting the local correlation of the values of $X_2(n, \omega)$. Because the form of the resulting synthesis equation is somewhat more complicated than the conventional synthesis equation (25) and relies on the introduction of a *synthesis window* (or, equivalently, a *synthesis filter*), the new synthesis equation will be motivated by a heuristic argument and the synthesis equations (25), (35), and (37) will be shown to be special cases. The general synthesis equation will then be proved for a general set of analysis and synthesis filter pairs, and equivalent interpretations will be offered in terms of the summation of the outputs of a filter bank and in terms of a weighted projection of a two-dimensional short-time sequence.

In order to exploit the local correlation of $X_2(n, \omega)$, a *synthesis window*, denoted $F_1(\omega, n)$, is introduced, and a new synthesis equation is formulated by replacing $X_2(n, \omega)$ in the conventional synthesis formula (25) by the moving average

$$X_2'(n, \omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} F_1(\omega - \varphi, n - r) X_2(r, \varphi)\, d\varphi \quad (38)$$

to obtain

$$x(n) = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} F_1(\omega - \varphi, n - r) \cdot X_2(r, \varphi) \exp[j\omega n]\, d\varphi\, d\omega \quad (39)$$

where $F_1(\omega, n)$ depends on the analysis window and remains to be determined. Equation (39) can be simplified by performing the integration with respect to ω to obtain

$$x(n) = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} \left\{ \int_{-\pi}^{\pi} F_1(\omega - \varphi, n - r) \cdot \exp[j(\omega - \varphi)n]\, d\omega \right\} X_2(r, \varphi) \exp[j\varphi n]\, d\varphi$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} f(n, n - r) X_2(r, \varphi) \exp[j\varphi n]\, d\varphi$$

or, replacing φ with ω,

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} f(n, n - r) X_2(r, \omega) \exp[j\omega n]\, d\omega. \quad (40)$$

The three synthesis formulas (25), (35), and (37) now become the special cases of (40):

$$f(n, m) = \delta(m) \quad (41a)$$

$$f(n, m) = \delta(n - m - n_o)/h(-m) \quad (41b)$$

and

$$f(n, m) = 1/H(0), \quad (41c)$$

respectively, where $\delta(n)$ denotes the unit sample.

To derive the general relationship between $f(n, m)$ and $h(n)$ so that (40) synthesizes $x(n)$ from $X_2(n, \omega)$, interchange the order of integration and summation in (40), and recognize the integral as the inverse partial Fourier transform of $X_2(r, \omega)$, which is just the short-time function $x(r, n) = h(r - n)x(n)$. Equation (40), therefore, reduces to

$$x(n) = \sum_{r=-\infty}^{\infty} f(n, n - r) h(r - n) x(n)$$

$$= \left\{ \sum_{m=-\infty}^{\infty} f(n, -m) h(m) \right\} x(n)$$

$$= x(n) \quad (42)$$

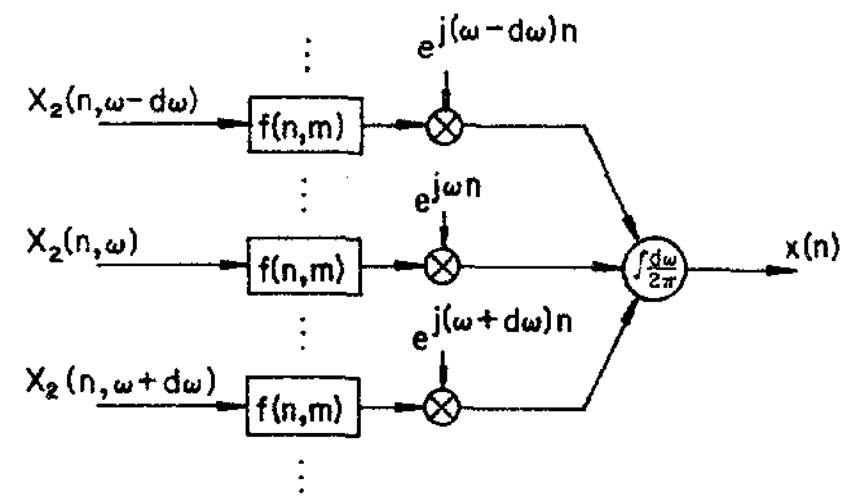


Fig. 3. Synthesis of a time sequence as a combination of filter bank outputs.

if and only if

$$\sum_{m=-\infty}^{\infty} f(n,-m)\,h(m)=1 \qquad \text{for all } n, \tag{43}$$

or equivalently, if and only if

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} F_2(n,\omega)\,H(\omega)\,d\omega=1 \qquad \text{for all } n. \tag{44}$$

The synthesis procedure, implied by (40), has two interpretations depending on the interpretation of the short-time Fourier transform. If $X_2(n,\omega)$ is interpreted as a set (indexed by ω) of sequences in n, then the synthesis procedure corresponds to filtering $X_2(n,\omega)$ with the linear time-varying filter $f(n,m)$ to obtain the set (also indexed by ω) of sequences in n

$$\overline{X}_2(n,\omega)=\sum_{r=-\infty}^{\infty} f(n,n-r)\,X_2(r,\omega). \tag{45}$$

$x(n)$ is obtained by modulating $\overline{X}_2(n,\omega)$ by $\exp[j\omega n]$ for every value of ω and integrating over ω as illustrated in Fig. 3.[2] Consequently, $f(n,m)$ will also be referred to as the *synthesis filter*. Alternatively, if $X_2(r,\omega)$ is interpreted as a set (indexed by r) of Fourier transforms, then (40) can be viewed as inverse Fourier transforming $X_2(r,\omega)$ to obtain the set (also indexed by r) of short-time sequences in n

$$\xi(r,n)=f(n,n-r)\,x(r,n) \tag{46}$$

where $\xi(r,n)$ corresponds to $x(r,n)$ weighted by the time-varying shifted window $f(n,n-r)$. For each value of n, $x(n)$ is obtained, as illustrated in Fig. 4, by projecting (summing) $\xi(r,n)$ in r:[3]

$$x(n)=\sum_{r=-\infty}^{\infty} f(n,n-r)\,x(r,n) \tag{47a}$$

$$=\sum_{r=-\infty}^{\infty} \xi(r,n). \tag{47b}$$

A more general theory of short-time Fourier analysis and synthesis can be formulated by also allowing the analysis window to vary as a function of time. Such a formulation is appropriate for adaptive-processing schemes when the particular choice of analysis window depends on the data to be analyzed [23], [24]. In addition, both the analysis and synthesis windows can be allowed to be functions of frequency for such techniques as constant-Q analysis [25], [26]. These formulations are more general than required for many applications and will not be pursued here. Furthermore, for the remainder of this paper, only time-invariant synthesis filters $f(n,m)=f(m)$ will be considered. Thus, the synthesis formula (40) reduces to

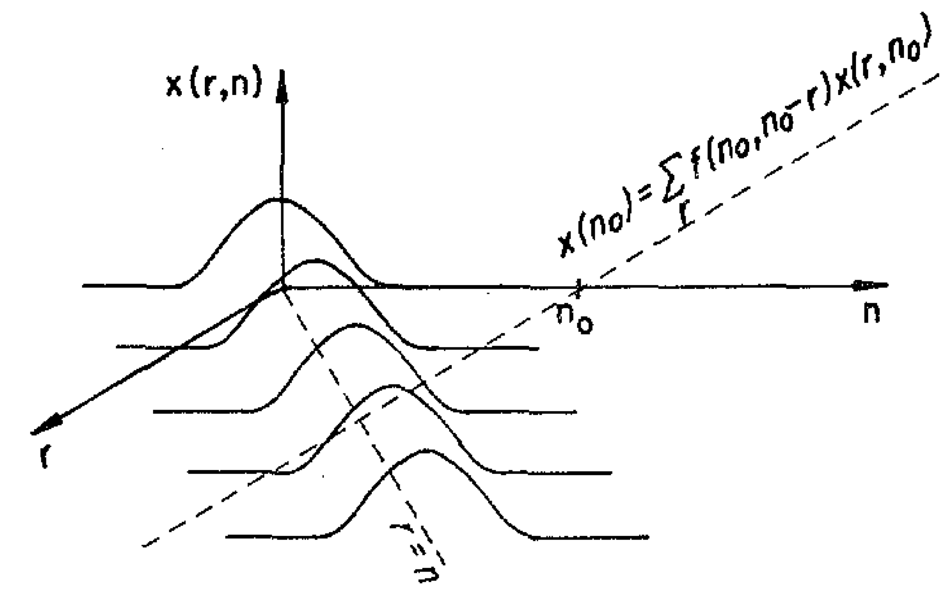


Fig. 4. Synthesis of a time sequence as a weighted projection of short-time sequences.

$$x(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{r=-\infty}^{\infty} f(n-r)\,X_2(r,\omega)\exp[j\omega n]\,d\omega \tag{48}$$

where the analysis and synthesis filters satisfy the condition

$$\sum_{m=-\infty}^{\infty} f(-m)\,h(m)=\frac{1}{2\pi}\int_{-\pi}^{\pi} F(\omega)\,H(\omega)\,d\omega=1. \tag{49}$$

Condition (49) is the time-invariant form of the conditions (43) and (44) and can be satisfied simply by appropriate normalization of the analysis and synthesis filters, provided that the summation and integral in (49) do not vanish.

IV. Linear Filtering Based on Short-Time Fourier Analysis

In view of the property of the Fourier transform that maps convolution in one domain to multiplication in the other domain and the generalization of this property to (24) for linear time-varying systems, it is natural to ask whether an analogous property exists for short-time Fourier transforms. Such a property does, indeed, exist, but applies only to a restricted class of linear time-varying systems determined by the filters used in the short-time Fourier analysis and synthesis. Specifically, this property is the following. Let $X_2(n,\omega)$ denote the short-time Fourier transform of $x(n)$ and let $T_2(n,\omega)$ denote the frequency response of an arbitrary linear time-varying system $t(n,m)$. Further, define the *modified* short-time Fourier transform $Y_2(n,\omega)$ as the product

$$Y_2(n,\omega)=T_2(n,\omega)\,X_2(n,\omega), \tag{50}$$

noting that, in general, $Y_2(n,\omega)$ is not a valid short-time Fourier transform in the sense that it cannot be expressed in the form of (26). If $y(n)$ is synthesized from $Y_2(n,\omega)$ according to the formula

$$y(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{r=-\infty}^{\infty} f(n-r)\,Y_2(r,\omega)\exp[j\omega n]\,d\omega \tag{51}$$

[2] This interpretation corresponds to the "filter bank sum" method [16].

[3] This interpretation corresponds to the "overlap-add" method [16].

then

$$y(n) = \sum_{m=-\infty}^{\infty} \bar{t}(n, m)\, x(n - m) \tag{52}$$

and corresponds to the response, to $x(n)$, of the linear time-varying system with the unit-sample response $\bar{t}(n, m)$ given by

$$\begin{aligned}\bar{t}(n, m) &= \sum_{r=-\infty}^{\infty} f(r)\, h(m - r)\, t(n - r, m)\\ &= \{f(n)\, h(m - n)\} *_n t(n, m).\end{aligned} \tag{53}$$

Furthermore, the modified system given by (53) is conveniently characterized as the product of the partial Fourier transforms

$$\bar{T}_1(\psi, m) = F_2(m, \psi)\, T_1(\psi, m) \tag{54}$$

where

$$F_2(m, \psi) = \sum_{r=-\infty}^{\infty} f(r)\, h(m - r) \exp[-j\psi r]$$

denotes the short-time Fourier transform of the synthesis filter $f(n)$ [not to be confused with the partial Fourier transform of the now discarded time-varying synthesis filter, $f(n, m)$].

The proof of (52) and (53) follows from substituting (50) into (51) to obtain

$$\begin{aligned}y(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} f(n - r)\, T_2(r, \omega)\\ &\quad \cdot X_2(r, \omega) \exp[j\omega n]\, d\omega\\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} f(n - r)\, T_2(r, \omega)\\ &\quad \cdot \sum_{m=-\infty}^{\infty} h(r - m)\, x(m) \exp[-j\omega m] \exp[j\omega n]\, d\omega\\ &= \sum_{m=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} f(n - r)\, h(r - m)\\ &\quad \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} T_2(r, \omega) \exp[j\omega(n - m)]\, d\omega\, x(m)\\ &= \sum_{m=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} f(n - r)\, h(r - m)\, t(r, n - m)\, x(m).\end{aligned}$$

Letting $m' = n - m$ and $r' = n - r$ gives

$$y(n) = \sum_{m'=-\infty}^{\infty} \sum_{r'=-\infty}^{\infty} f(r')\, h(m' - r')\, t(n - r', m')\, x(n - m')$$

from which (52) and (53) follow.

Thus, for a particular pair of analysis and synthesis filters, linear filtering implemented as the product of a time-varying frequency response multiplied by the short-time Fourier transform of the input sequence is restricted to the class of filters with unit-sample responses of the form $\bar{t}(n, m)$ specified by (53), with $t(n, m)$ arbitrary. The significance of this condition, a limitation on the simultaneous time and frequency variation of the filter $\bar{t}(n, m)$, determined by the short-time Fourier analysis and synthesis filters, is illustrated by considering three special cases.

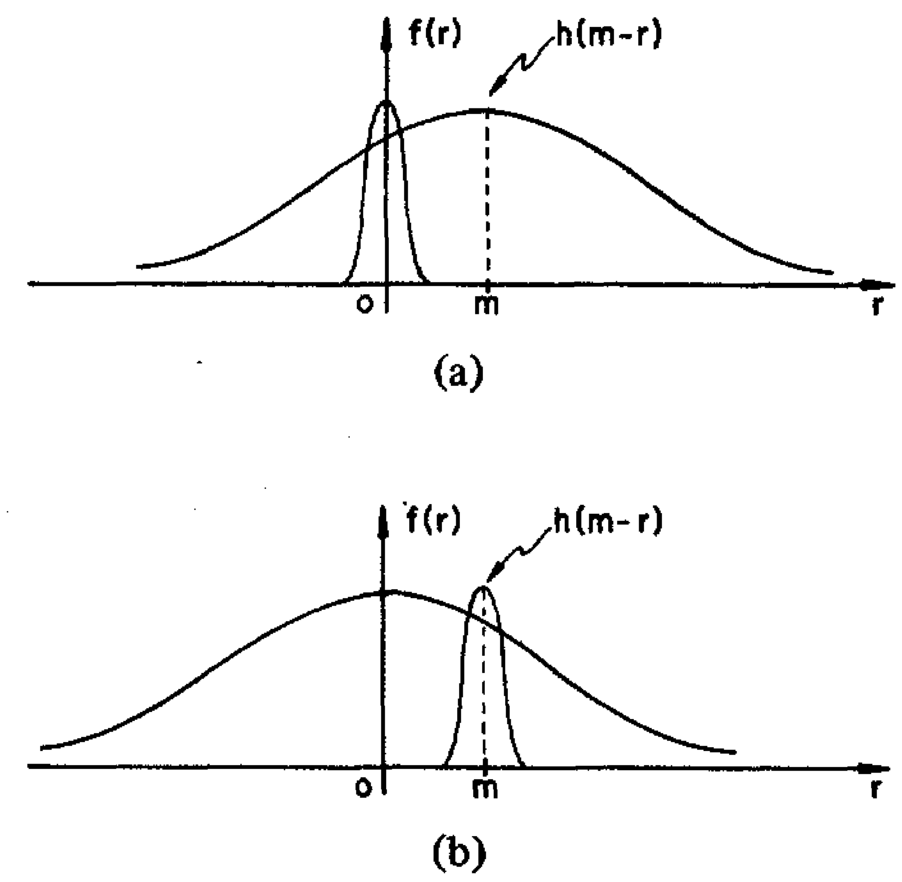


Fig. 5. (a) Example of analysis and synthesis filters for which $f(r)\, h(m - r) \approx f(r)\, h(m)$. (b) Example of analysis and synthesis filters for which $f(r)\, h(m - r) \approx f(m)\, h(m - r)$.

First, a linear time-varying filter $\bar{t}(n, m)$ with arbitrary time variation in n can be implemented for a given analysis window $h(n)$ by designing the synthesis window $f(n)$ to be of much shorter duration than $h(n)$, as illustrated in Fig. 5(a). In this case, $h(n)$ is approximately constant over the duration of $f(n)$, so that $f(r)\, h(m - r) \approx f(r)\, h(m)$, and $\bar{t}(n, m)$ becomes

$$\begin{aligned}\bar{t}(n, m) &= \sum_{r=-\infty}^{\infty} f(r)\, h(m - r)\, t(n - r, m)\\ &\approx \sum_{r=-\infty}^{\infty} f(r)\, h(m)\, t(n - r, m)\\ &= \sum_{r=-\infty}^{\infty} f(n - r)\, t(r, m)\, h(m)\end{aligned}$$

$$\bar{t}(n, m) \approx f(n) *_n t(n, m)\, h(m). \tag{55}$$

Thus, the unit-sample response $t(n, m)$ is windowed by $h(m)$ in the m direction and smoothed by $f(n)$ in the n direction. Equivalently, the time-varying frequency response $T_2(n, \omega)$ is smoothed by $H(\omega)$ in ω and by $f(n)$ in n. In the limit, $f(n)$ can be chosen as $f(n) = \delta(n)$ so that $\bar{t}(n, m)$ becomes (exactly)

$$\bar{t}(n, m) = t(n, m)\, h(m). \tag{56}$$

The second case is complementary to the first. Namely, a linear time-varying filter with arbitrary frequency variation in ω can be implemented by designing the synthesis window $f(n)$ to be of much greater duration than $h(n)$, as illustrated in Fig. 5(b). Here, $f(n)$ is approximately constant over the duration of $h(n)$, so that $f(r)\, h(m - r) \approx f(m)\, h(m - r)$, and $\bar{t}(n, m)$ becomes

$$\bar{t}(n, m) = \sum_{r=-\infty}^{\infty} f(r)\, h(m - r)\, t(n - r, m)$$

$$\approx \sum_{r=-\infty}^{\infty} f(m)\, h(m-r)\, t(n-r, m)$$

$$= \sum_{r=-\infty}^{\infty} h(m-n+r)\, t(r, m) f(m)$$

$$\bar{t}(n, m) \approx [h(-r) *_r t(r, m)] \big|_{r=n-m} \cdot f(m). \tag{57}$$

In this case, $t(n, m)$ is windowed by $f(m)$ in m and smoothed by $h(-n)$ in n, but an additional "smearing" is introduced because the convolution of (57) is evaluated for $(n - m)$ rather than n. In the limit, $f(n)$ can be chosen as unity and $\bar{t}(n, m)$ becomes (exactly)

$$\bar{t}(n, m) = h(-r) *_r t(r, m)\big|_{r=n-m}. \tag{58}$$

The third special case is the implementation of a slowly time-varying filter. If $t(n, m)$ can be approximated as stationary over the duration of the synthesis window so that $f(r)\, t(n - r, m) \approx f(r)\, t(n, m)$, then $\bar{t}(n, m)$ becomes

$$\bar{t}(n, m) = \sum_{r=-\infty}^{\infty} f(r)\, h(m-r)\, t(n-r, m)$$

$$\approx \sum_{r=-\infty}^{\infty} f(r)\, h(m-r)\, t(n, m)$$

$$\bar{t}(n, m) \approx \{f(m) *_m h(m)\}\, t(n, m). \tag{59}$$

Thus, $\bar{t}(n, m)$ is just the product of $t(n, m)$ with the effective window $f(m) * h(m)$. Alternatively, (59) is also valid if $t(n, m)$ is slowly-varying in the sense that the bandwidth of $T_1(\psi, m)$ in ψ is narrow compared with the bandwidth of $F(\psi)$.

In practice, the limit on simultaneous time and frequency variation is generally not a serious restriction. In fact, it can often be exploited. A common application of short-time Fourier analysis is for adaptive filtering. Here, a signal is filtered by a time-varying system, the characteristics of which depend on the local characteristics of the input signal. By using the formulation leading to (56), filter design by windowing [1]-[3] can be accomplished automatically. Furthermore, by using the formulation leading to (55), linear smoothing of the time variation of the time-varying frequency response can also be introduced.

Part II

V. Discrete Short-Time Fourier Analysis and Synthesis

The short-time Fourier transform $X_2(n, \omega)$ represents the sequence $x(n)$ by a function of the continuous variable ω for each value of the index n and thus contains redundant information about the signal and the analysis window. This section considers the problem of representing a sequence by samples of its short-time Fourier transform with the result that for proper choices of sampling rates in both time and frequency and for certain choices of analysis and synthesis filters, the short-time Fourier transform $X_2(n, \omega)$ of $x(n)$ can be sampled using, on the average, one sample per value of $x(n)$.

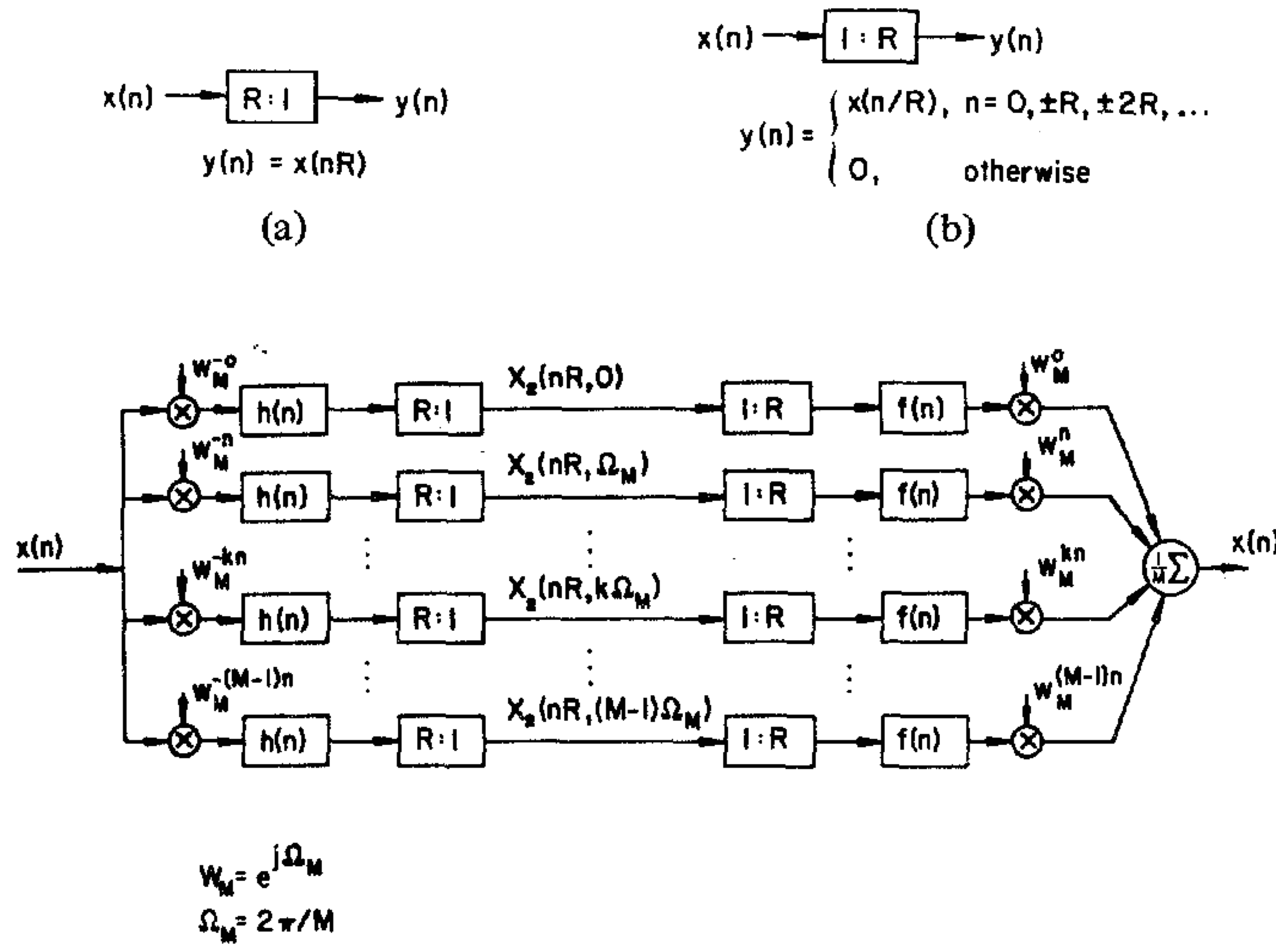


Fig. 6. (a) R:1 compressor. (b) 1:R expander. (c) Filter-bank analog for discrete short-time Fourier analysis/synthesis.

Define the *discrete* short-time Fourier transform of $x(n)$ as

$$X_2(sR, k\Omega_M) = \sum_{m=-\infty}^{\infty} h(sR - m)\, x(m) \exp[-j\Omega_M km]$$

$$\text{for} \quad k = 0, 1, \cdots, M-1 \tag{60}$$

corresponding to samples of the short-time Fourier transform specified every R samples in time and $\Omega_M = 2\pi/M$ radians in frequency. For certain choices of the sampling parameters R and M and the filters $h(n)$ and $f(n)$, it will be shown that $x(n)$ can be recovered by means of the synthesis formula

$$x(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=-\infty}^{\infty} f(n - sR)\, X_2(sR, k\Omega_M) \exp[j\Omega_M kn]. \tag{61}$$

The pair of equations (60) and (61) describe the M-channel filter-bank analysis/synthesis system depicted in Fig. 6. $X_2(sR, k\Omega_M)$ is the output of the analysis filter $h(n)$ in the kth channel, uniformly sampled every R samples. The synthesized signal is generated by interpolating (time expanding and filtering) each of the M channels with $f(n)$, modulating by $\exp[j\Omega_M kn]$, and summing over all the channels.

To show that $x(n)$ can be recovered by means of (61), refer to Fig. 6 and observe that the identical combination of the analysis filter $h(n)$ followed by an $R:1$ compressor, and a $1:R$ interpolator [$1:R$ expander followed by the synthesis filter, $f(n)$] appears in each of the M channels of the filter bank. This combination can be replaced in each channel, as illustrated in Fig. 7, by the equivalent linear time-varying system $w(n, m)$ given by

$$w(n, m) = \sum_{s=-\infty}^{\infty} f(n - sR)\, h(sR - n + m) \tag{62}$$

where $w(n, m)$ denotes the response of the system at sample n to a unit sample applied at time $(n - m)$. For each value of n, (62) corresponds to uniformly sampling $h(n)$ every R samples,

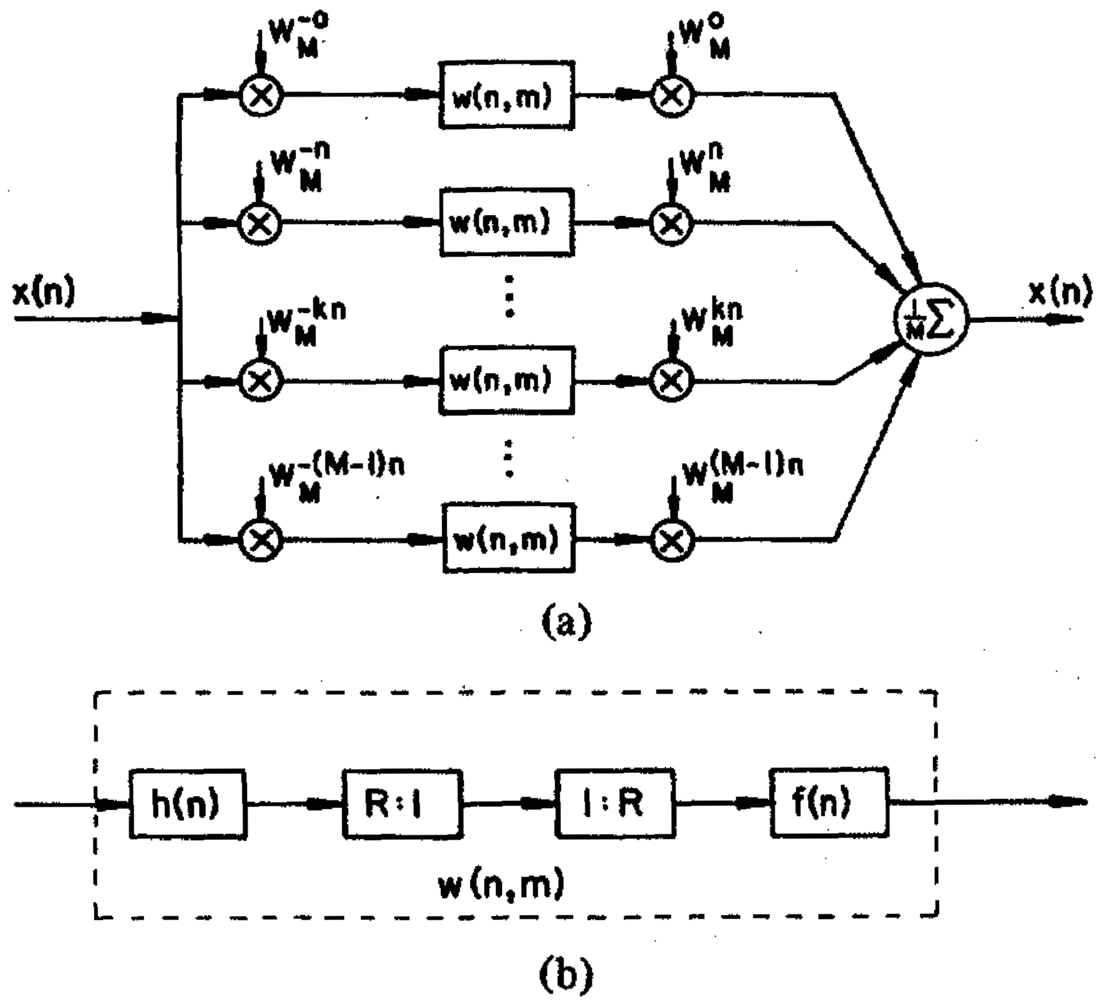


Fig. 7. (a) Filter-bank equivalent to Fig. 6(c). (b) Equivalent linear time-varying filter for each channel.

starting at sample $(n - m)$, then interpolating by $1:R$ with $f(n)$. The time variation of $w(n, m)$ results because decimation/interpolation is not a time-invariant operation.

The condition such that $x(n)$ can be recovered exactly from $X_2(sR, k\Omega_M)$ by means of (61) is

$$w(n, pM) = \delta(p) \qquad \text{for all } n. \tag{63}$$

To derive this condition, define $\tilde{x}(n)$ as the right-hand side of (61), i.e.,

$$\tilde{x}(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=-\infty}^{\infty} f(n - sR)\, X_2(sR, k\Omega_M) \exp\,[j\Omega_M kn].$$

Replacing $X_2(sR, k\Omega_M)$ by its definition (60) gives

$$\tilde{x}(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=-\infty}^{\infty} f(n - sR) \cdot \left\{ \sum_{m=-\infty}^{\infty} h(sR - m)\, x(m) \exp\,[-j\Omega_M km] \right\} \cdot \exp\,[j\Omega_M kn]$$

$$= \sum_{s=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} f(n - sR)\, h(sR - m)\, x(m) \cdot \left\{ \frac{1}{M} \sum_{k=0}^{M-1} \exp\,[j\Omega_M k(n - m)] \right\}$$

or

$$\tilde{x}(n) = \sum_{p=-\infty}^{\infty} f(n - sR)\, h(sR - n + pM)\, x(n - pM).$$

Thus, $\tilde{x}(n) = x(n)$ if and only if

$$\sum_{s=-\infty}^{\infty} f(n - sR)\, h(sR - n + pM) = \delta(p) \qquad \text{for all } n. \tag{64}$$

Using the definition (62) for $w(n, m)$ yields the desired condition (63).

The representation of a sequence by its sampled short-time Fourier transform with no redundancy results when the decimation ratio, R, is equal to the number of frequency samples, M. In this case, (62) and (63) becomes

$$\sum_{s=-\infty}^{\infty} f(n - sM)\, h((p + s) M - n) = \delta(p) \qquad \text{for all } n. \tag{65}$$

Because the left-hand side of (65) is periodic in n with period M, the problem of determining the synthesis filter $f(n)$, given a particular analysis window $h(n)$, reduces to M independent inverse filtering problems, one for each value of n in the range $0 \leqslant n < M$.

An alternate formulation of the condition for the exact resynthesis of $x(n)$, according to (61), can be obtained by taking the partial Fourier transform of (64) with respect to n, giving

$$F_2(pM, \omega) \left\{ \frac{2\pi}{R} \sum_{q=0}^{R-1} u_o(\omega - 2\pi q/R) \right\} = 2\pi\delta(p)\, u_o(\omega) \tag{66}$$

where

$$F_2(m, \omega) = \sum_{n=-\infty}^{\infty} h(m - n)\, f(n) \exp\,[-j\omega n]$$

is the short-time Fourier transform of the synthesis filter $f(n)$, and $u_o(\omega)$ denotes the unit-impulse function. Equation (66) requires that $F_2(m, \omega)$ have the property that

$$F_2(m, \omega) = \begin{cases} R & \text{for } m = 0 \text{ and } \omega = 0 \\ 0 & \text{for } m = \pm M, \pm 2M, \cdots \\ 0 & \text{for } \omega = 2\pi/R, 4\pi/R, \cdots, 2\pi(R - 1)/R. \end{cases} \tag{67}$$

This formulation of the condition that must be satisfied by the analysis and synthesis filters will be useful for dealing with the problem of linear filtering based on discrete short-time Fourier analysis considered in Section VII.

VI. Short-Time Fourier Analysis and Synthesis Based on the FFT

One difficulty in implementing systems based on short-time Fourier analysis has been the rapid increase in the amount of computation required for the analysis and synthesis as the number of frequency samples becomes large. This section discusses computationally efficient implementations for discrete short-time Fourier analysis and synthesis using the FFT algorithm [11], [12].

In order to simplify notation in this section, the discrete short-time Fourier transform (for $R = 1$) is denoted as

$$X_k[n] = X_2(n, 2\pi k/M) \tag{68}$$

and the complex exponentials corresponding to the Mth roots of unity are denoted as

$$W_M^k = \exp\,[j2\pi k/M]. \tag{69}$$

Thus, the definition of the discrete short-time Fourier transform of the sequence $x(n)$, for $R = 1$, becomes

$$X_k[n] = \sum_{m=-\infty}^{\infty} h(n-m)\,x(m)\,W_M^{-mk}. \tag{70}$$

A. Implementation of the Short-Time Fourier Analyzer

If the number of frequency samples M is chosen to be a highly composite number (usually an integral power of 2), then the FFT algorithm can be employed to compute efficiently the short-time Fourier transform $X_k[n]$ defined by (70). Because (70) does not have the form of a discrete Fourier transform (DFT), it cannot be directly computed with the FFT algorithm. The limits on the summation are given as infinite, but in practice are finite and determined by the length of $h(n)$. By recognizing $X_k[n]$ for fixed n, as samples equally spaced in frequency, of the (continuously valued) partial Fourier transform of $x(m)\,h(n-m)$, $X_k[n]$ can be obtained by time-domain aliasing $x(m)\,h(n-m)$ and then computing the DFT of the aliased sequence [11]. Thus,

$$X_k[n] = W_M^{-nk} \sum_{m=0}^{M-1} \underset{\sim}{x}_m[n]\, W_M^{-mk} \tag{71}$$

where $\underset{\sim}{x}_m[n]$ is the aliased short-time sequence

$$\underset{\sim}{x}_m[n] = \sum_{p=-\infty}^{\infty} x(n+pM+m)\,h(-pM-m). \tag{72}$$

In addition to the computational savings gained by computing the short-time Fourier transform using the FFT, further savings may be gained by avoiding the complex multiplications by W_M^{-nk} in (71). By circularly shifting $\underset{\sim}{x}_m[n]$ in m, prior to computing its DFT, the multiplications by W_M^{-nk} are avoided and (71) can be rewritten as [12]

$$X_k[n] = \sum_{m=0}^{M-1} \underset{\sim}{x}_{((m-n))_M}[n]\, W_M^{-mk},$$

or

$$X_k[n] = \sum_{m=0}^{M-1} x_m[n]\, W_M^{-mk} \tag{73}$$

where

$$x_m[n] = \underset{\sim}{x}_{((m-n))_M}[n] \tag{74}$$

and $((n))_M$ denotes "n reduced modulo M." Although (72)–(74) provide a convenient formulation for implementation [11], [12], (73) is, in fact, just the discrete Fourier transform formulation of (27) where $x_m[n]$ is the aliased short-time function

$$x_m[n] = \sum_{p=-\infty}^{\infty} x(n, m+pM). \tag{75}$$

B. Implementation of the Short-Time Fourier Synthesizer

Let $Y_k[sR]$ denote the samples of data available to the synthesizer and let $y(n)$ denote the time sequence to be synthesized. The discrete short-time Fourier synthesis formula

$$y(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=-\infty}^{\infty} f(n-sR)\, Y_k[sR]\, W_M^{nk} \tag{76}$$

can be interpreted according to the discussion in Section V as interpolating each of the M sequences $Y_k[sR]$ for $k = 0, 1, \cdots, M-1$, by $1:R$, modulating each by W_M^{nk}, and summing the resulting signals. Clearly, this procedure could be implemented directly [11]; however, it is computationally intractable for large values of the transform size M.

A synthesis procedure can be formulated, which, for a highly composite number M, permits $y(n)$ to be computed from the samples $Y_k[sR]$ using the FFT algorithm [12]. In addition to the computational savings afforded by employing the FFT, the number of computations required to perform the $1:R$ interpolation is reduced by the factor M.

Assume that $f(n)$ is the unit-sample response of a $1:R$ FIR interpolating filter with length $2QR + 1$. The synthesis formula (76) then becomes

$$y(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=L^-}^{L^+} f(n-sR)\, Y_k[sR]\, W_M^{nk} \tag{77}$$

where the limits on the inner sum, determined by the length of $f(n)$, are

$$L^+(n) = \lfloor n/R \rfloor + Q$$

$$L^-(n) = \lfloor n/R \rfloor - Q + 1$$

and where $\lfloor N \rfloor$ means "the largest integer contained in N." Since the limits on both sums are finite, the order of summation can be interchanged to give

$$y(n) = \sum_{s=L^-}^{L^+} f(n-sR) \left\{ \frac{1}{M} \sum_{k=0}^{M-1} Y_k[sR]\, W_M^{nk} \right\} \tag{78}$$

or

$$y(n) = \sum_{s=L^-}^{L^+} f(n-sR)\, y_n[sR] \tag{79}$$

where

$$y_n[sR] = \frac{1}{M} \sum_{k=0}^{M-1} Y_k[sR]\, W_M^{nk}. \tag{80}$$

Thus, for fixed values of s, $y_n[sR]$ is the inverse partial DFT of $Y_k[sR]$ with respect to k, and can, therefore, be computed using the FFT algorithm. It is important to observe that $y_n[sR]$ is periodic in n with period M. Since the FFT only computes values of $y_n[sR]$ for one period ($n = 0, 1, \cdots, M-1$), the subscript n in (79) and (80) is interpreted as reduced modulo M. Equation (79) is the discrete short-time Fourier transform formulation of (47a) for a time-invariant synthesis filter.[4]

[4]For the implementation of (79), Crochiere has observed [27] that, for $R < M$, a savings in storage over that required by "helical interpolation" [11] can be achieved by first computing the short-time sequence

$$\eta_n[sR] = f(n-sR)\, y_n[sR]$$

and then projecting it to obtain

$$y(n) = \sum_{s=L^-}^{L^+} \eta_n[sR].$$

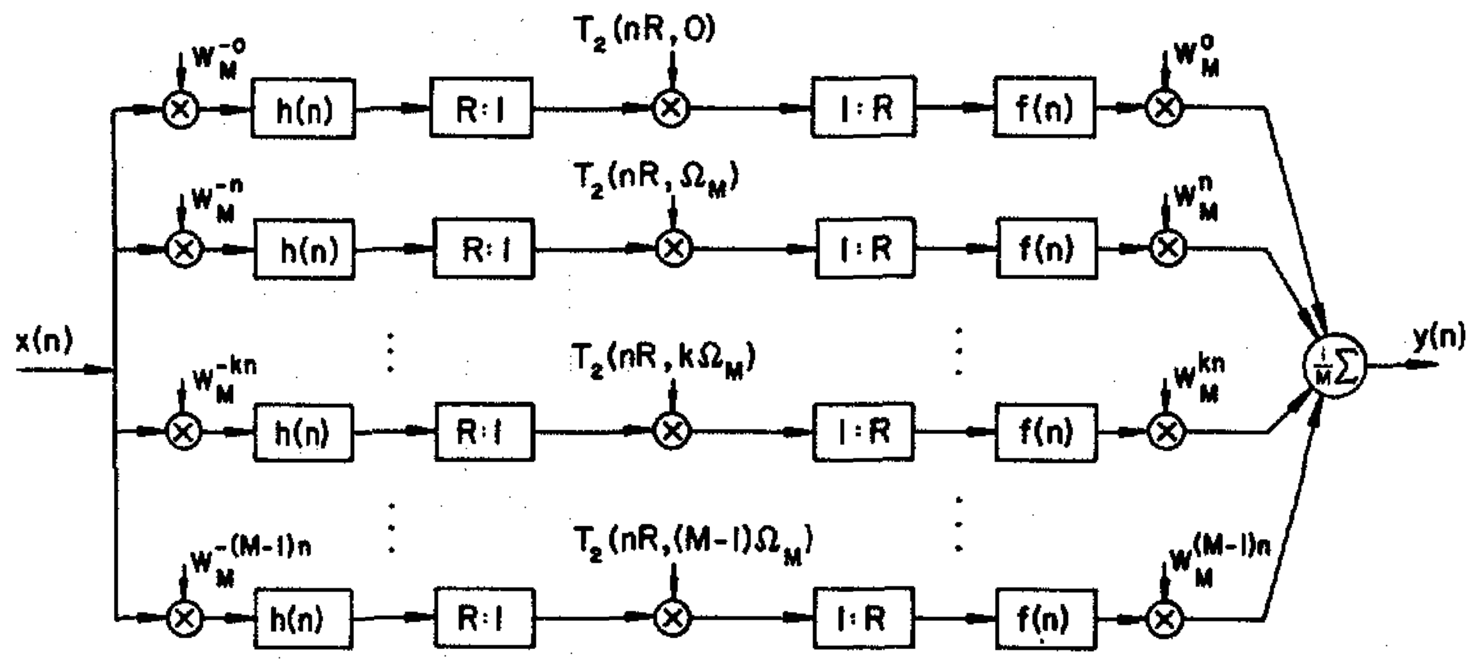


Fig. 8. Filter-bank analog for a linear time-varying filtering based on short-time Fourier analysis.

VII. Linear Filtering Based on Discrete Short-Time Fourier Analysis

Section V showed that a sequence could be represented by the discrete short-time Fourier transform with no redundancy. This section now considers the problem of representing linear filtering as the product of the discrete short-time Fourier transform of the input signal multiplied by samples of the time-varying frequency response of the filter. If the sampled-transform implementation is to be equivalent to the formulation of Section IV, then the sampling rates in both time and frequency must be sufficiently high to prevent aliasing in either time or frequency.

A. Linear Time-Varying Filters

Let $X(sR, k\Omega_M)$ represent samples of the short-time Fourier transform of $x(n)$ as defined by (60), and let $T_2(sR, k\Omega_M)$ represent samples of the time-varying frequency response of the linear time-varying system with unit-sample response $t(n, m)$. Furthermore, assume that

$$T_1(\psi, m) = \sum_{n=-\infty}^{\infty} t(n, m) \exp[-j\psi n] \tag{81}$$

is band limited in ψ and time limited in m so that $t(n, m)$ can be represented, without aliasing, by samples of $T_2(n, \omega)$.

Define $Y_2(sR, k\Omega_M)$ as the product

$$Y_2(sR, k\Omega_M) = T_2(sR, k\Omega_M) X_2(sR, k\Omega_M) \tag{82}$$

and define $y(n)$, using the discrete short-time Fourier synthesis formula, as

$$y(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=-\infty}^{\infty} f(n - sR) Y_2(sR, k\Omega_M) \exp[j\Omega_M kn]. \tag{83}$$

Thus,

$$y(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=-\infty}^{\infty} f(n - sR) T_2(sR, k\Omega_M) X_2(sR, k\Omega_M) \cdot \exp[j\Omega_M kn] \tag{84}$$

and $y(n)$ corresponds to the output of the system depicted in Fig. 8.

The overall unit-sample response of this system is obtained by replacing $X_2(sR, k\Omega_M)$ in (84) by its definition (60). Thus, for a given input $x(n)$, the corresponding output $y(n)$ is given by

$$\begin{aligned} y(n) &= \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=-\infty}^{\infty} f(n - sR) T_2(sR, k\Omega_M) \\ &\quad \cdot \sum_{m=-\infty}^{\infty} h(sR - m) x(m) \exp[-j\Omega_M km] \\ &\quad \cdot \exp[j\Omega_M kn] \\ &= \sum_{s=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} f(n - sR) h(sR - m) \\ &\quad \cdot \frac{1}{M} \sum_{k=0}^{M-1} T_2(sR, k\Omega_M) \exp[j\Omega_M k(n - m)]\, x(m). \end{aligned}$$

Now, letting $m' = n - m$, evaluating the sum over k, and interchanging the order of summation gives

$$y(n) = \sum_{m'=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} f(n - sR) h(sR - n + m') \cdot t(sR, m' - pM) x(n - m'),$$

or

$$y(n) = \sum_{m'=-\infty}^{\infty} \tilde{t}(n, m') x(n - m') \tag{85}$$

where $\tilde{t}(n, m)$ is given by

$$\tilde{t}(n, m) = \sum_{p=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} f(n - sR) h(sR - n + m) \cdot t(sR, m - pM). \tag{86}$$

Thus, $\tilde{t}(n, m)$ corresponds to the overall unit-sample response of the system depicted in Fig. 8.

The conditions on the filters and sampling rates, such that $\tilde{t}(n, m)$ defined by (86) for the sampled-transform implementation is identical to $\bar{t}(n, m)$ defined by (53) for the nonsampled implementation, become apparent by expressing the partial Fourier transform $\tilde{T}_1(\psi, m)$ in terms of $T_1(\psi, m)$. Transforming (86) with respect to n gives

$$T_1(\psi, m) = \sum_{n=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} f(n - sR) h(sR - n + m) \cdot t(sR, m - pM) \exp[-j\psi n]$$

and, letting $r = n - sR$,

$$\begin{aligned}\tilde{T}_1(\psi, m) &= \sum_{r=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} f(r)\, h(m-r) \\ &\quad \cdot t(sR, m - pM) \exp\left[-j\psi(r + sR)\right] \\ &= \sum_{r=-\infty}^{\infty} f(r)\, h(m-r) \exp\left[-j\psi r\right] \\ &\quad \cdot \sum_{p=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} t(sR, m-pM) \exp\left[-j\psi sR\right];\end{aligned}$$

hence

$$\begin{aligned}\tilde{T}_1(\psi, m) &= F_2(m, \psi) \frac{1}{R} \sum_{q=0}^{R-1} \sum_{p=-\infty}^{\infty} \\ &\quad \cdot T_1(\psi - 2\pi q/R, m - pM)\end{aligned} \tag{87}$$

where $T_1(\psi, m)$ is the partial Fourier transform of $t(n, m)$ and $F_2(m, \psi)$ is the short-time Fourier transform of $f(n)$ given by

$$\begin{aligned}F_2(m, \psi) &= \sum_{r=-\infty}^{\infty} h(m-r) f(r) \exp\left[-j\psi r\right] \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\psi + \varphi) H(\varphi) \exp\left[j\varphi m\right] d\varphi.\end{aligned} \tag{88}$$

Equation (87) states that $T_1(\psi, m)$ is aliased in both time (m) and frequency (ψ), then windowed by $F_2(m, \psi)$ to produce $\tilde{T}_1(\psi, m)$.

For the *sampled-transform implementation* to be equivalent to the *nonsampled implementation* of Section IV, $\tilde{T}_1(\psi, m)$, given by (87), must be identical to $\overline{T}_1(\psi, m)$, given by (54). To prevent distortion due to aliasing, the number of frequency samples M must be at least as great as the duration of $T_1(\psi, m)$ in m, and the temporal sampling frequency $2\pi/R$ must be greater than the bandwidth of $T_1(\psi, m)$ in ψ. Further, to eliminate the images of $T_1(\psi, m)$ in (87), M must be, in general, at least as great as the duration of $F_2(m, \psi)$ in m and $2\pi/R$ must be greater than the bandwidth of $F_2(m, \psi)$ in ψ. The latter condition is not necessary for the special case of $T_1(\psi, m)$ and $F_2(\psi, m)$ having structures such that regions of zero of $F_2(m, \psi)$ exactly cancel the images of $T_1(\psi, m)$ in (87). One example of this situation is the implementation of fast convolution for linear time-invariant filters treated in the following section.

B. Fast Convolution

To conclude this section, the methods of fast convolution [22] are considered as special cases of linear filtering based on discrete short-time Fourier analysis. Suppose a signal is processed by multiplying its discrete short-time Fourier transform by samples of the frequency response of a linear time-invariant system and a new signal is synthesized from the product. If the short-time Fourier analysis and synthesis are implemented using the FFT, as discussed in Section VI, and if the analysis and synthesis filters have appropriate rectangular unit-sample responses, then the overall processing is equivalent to the overlap-save or overlap-add method of fast convolution, depending upon the designs for the unit-sample responses of the analysis and synthesis filters.

In particular, let L denote the length of the unit-sample response $t(m)$ of the filter to be implemented so that

$$t(n, m) = \begin{cases} t(m) & \text{for } 0 \leqslant m < L \\ 0 & \text{otherwise.} \end{cases} \tag{89}$$

The overlap-save method results for

$$\begin{aligned} h(m) &= \begin{cases} 1 & \text{for } 0 \leqslant m < M \\ 0 & \text{otherwise} \end{cases} \\ f(m) &= \begin{cases} 1 & \text{for } -R < m \leqslant 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{90}$$

with the parameters L, M, and R satisfying the constraint

$$M \geqslant L + R - 1. \tag{91}$$

In the terminology associated with the fast-convolution method, L is the length of the filter unit-sample response, M is the length of each input-data section which is equal to the DFT size, and R is the number of "good" points per output section which is equal to the spacing between overlapping input-data sections.

If $y(n)$ denotes the result of filtering $x(n)$ with $t(n)$, then the convolution

$$y(n) = \sum_{m=0}^{L-1} t(m)\, x(n-m) \tag{92}$$

can be implemented as follows. The discrete short-time Fourier transform

$$X_2(sR, k\Omega_M) = \sum_{m=-\infty}^{\infty} h(sR - m)\, x(m) \exp\left[-j\Omega_M km\right] \tag{93a}$$

is calculated using the FFT and multiplied by samples of the frequency response of $t(m)$ to obtain

$$Y_2(sR, k\Omega_M) = T(k\Omega_M)\, X_2(sR, k\Omega_M), \quad \text{for } k = 0, 1, \cdots, M-1. \tag{93b}$$

The desired output signal $y(n)$ is synthesized according to the discrete short-time Fourier synthesis formula

$$y(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{s=-\infty}^{\infty} f(n - sR)\, Y_2(sR, k\Omega_M) \exp\left[j\Omega_M kn\right] \tag{93c}$$

implemented using (79), (80), and the FFT.

To show that $y(n)$, given by (93), is indeed the desired linear convolution (92), recall from the previous section that $y(n)$ corresponds to the linear convolution of $x(n)$ with the modified unit-sample response $\tilde{t}(n, m)$ given by (86). For a linear time-invariant system, $t(n, m) = t(m)$; thus, it must be shown that $\tilde{t}(n, m) = t(m)$ for all n, or equivalently, that

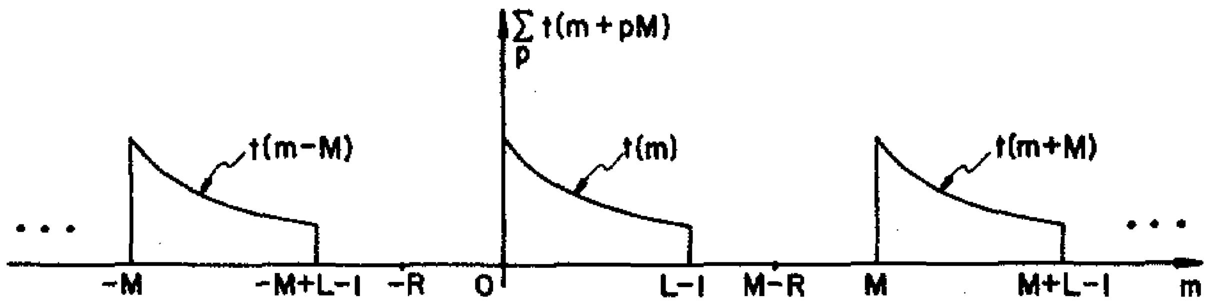


Fig. 9. Finite duration unit-sample response $t(m)$ periodically repeated in time with period M.

$\tilde{T}_1(\psi, m) = T_1(\psi, m) = 2\pi u_o(\psi)\ t(m)$ for all ψ. Substituting $T_1(\psi, m) = 2\pi u_o(\psi)\ t(m)$ into (87) gives

$$\begin{aligned}\tilde{T}_1(\psi, m) &= F_2(m, \psi) \frac{1}{R} \sum_{q=0}^{R-1} \sum_{p=-\infty}^{\infty} 2\pi u_o(\psi - 2\pi q/R) \\ &\quad \cdot t(m - pM) \\ &= \frac{2\pi}{R} \sum_{q=0}^{R-1} \sum_{p=-\infty}^{\infty} F_2(m, 2\pi q/R)\ t(m - pM) \\ &\quad \cdot u_o(\psi - 2\pi q/R) \end{aligned} \tag{94}$$

where $F_2(\omega, \psi)$, defined by (88), can be shown to be given by

$$F_2(m, \psi) = \begin{cases} e^{j\psi(R-1-m)/2} \cdot \dfrac{\sin[\psi(m+R)/2]}{\sin(\psi/2)} & \text{for} \quad -R < m \leqslant 0 \\ e^{j\psi(R-1)/2} \cdot \dfrac{\sin[\psi R/2]}{\sin(\psi/2)} & \text{for} \quad 0 \leqslant m \leqslant M-R \\ e^{j\psi(M-1-m)/2} \cdot \dfrac{\sin[\psi(M-m)/2]}{\sin(\psi/2)} & \text{for} \quad M-R \leqslant m < M \\ 0 & \text{otherwise.} \end{cases} \tag{95}$$

By considering $\tilde{T}_1(\psi, m)$ for each of the four conditions in (95), $\tilde{T}_1(\psi, m)$ can be seen to reduce to $\tilde{T}_1(\psi, m) = 2\pi u_o(\psi)\ t(m)$. For $m \leqslant -R$ or $m \geqslant M$, $\tilde{T}_1(\psi, m) = T_1(\psi, m) = 0$ because $F_2(m, \psi) = 0$. For m in the range $-R < m \leqslant 0$, it can be seen, by referring to Fig. 9, that $\tilde{T}_1(\psi, m) = T_1(\psi, m) = 0$ because $t(m) = 0$ for $m < 0$ or $m \geqslant L$, and because $-M + L - 1 \leqslant -R$ from condition (91). For m in the range $0 \leqslant m \leqslant M - R$, $\tilde{T}_1(\psi, m) = T_1(\psi, m) = 2\pi u_o(\psi)\ t(m)$ because $F_2(m, 2\pi q/R) = R$ for $q = 0$ and $F_2(m, 2\pi q/R) = 0$ for $1 \leqslant q < R$. For m in the range $M - R < m < M$, $\tilde{T}_1(\psi, m) = T_1(\psi, m) = 0$ because, again referring to Fig. 9, $t(m) = 0$ for $m < 0$ or $m \geqslant L$ and because $L - 1 \leqslant M - R$ from condition (91). Consequently, $\tilde{T}_1(\psi, m) = T_1(\psi, m)$ for all ψ and m. Thus, the overall unit-sample response of the system $\tilde{t}(n, m)$ is exactly $t(m)$ for all n, and the desired linear convolution (92) is achieved.

The overlap-add method results when the designs for $h(m)$ and $f(m)$ are interchanged so that

$$h(m) = \begin{cases} 1 & \text{for} \ -R < m \leqslant 0 \\ 0 & \text{otherwise} \end{cases} \tag{96}$$

$$f(m) = \begin{cases} 1 & \text{for} \ 0 \leqslant m < M \\ 0 & \text{otherwise} \end{cases}$$

and the analysis/synthesis is again implemented using the FFT subject to the constraint (91). Here R is the length of each of the contiguous input-data sections which is equal to the spacing between overlapping output sections and M is the length of each of the output-data sections which is equal to the DFT size.

To prove that the filters (96) for the overlap-add method result in the desired linear convolution, observe that interchanging the designs of $f(m)$ and $h(m)$ corresponds to replacing ψ with $-\psi$ in $F_2(m, \psi)$ and multiplying by $\exp[-j\psi m]$. Thus, the short-time Fourier transform of $f(m)$ for the overlap-add method is just the Fourier transform of $f(m)$ for the overlap-save method (95) with ψ replaced by $-\psi$ and multiplied by $\exp[-j\psi m]$. By the previous argument $\tilde{t}(n, m) = t(m)$ and the desired linear convolution is again achieved.

The difference between the formulation based on short-time Fourier analysis and the conventional formulation [22] is that in the conventional formulation the analogous filter-bank channel signals are bandpass signals as illustrated in Fig. 10(a), whereas, in the formulation based on short-time Fourier analysis, the channel signals are low-pass signals as illustrated in Fig. 10(b). However, because the modulation operations on the channel signals correspond to circular shifts of the data, as discussed in Section VI-A; the computational difference between the two formulations simply amounts to a difference in indexing schemes. Consequently, fast convolution can be considered a special case of linear filtering based on discrete short-time Fourier analysis.

VIII. Summary

This paper has developed certain aspects of the theory of short-time Fourier analysis and synthesis. It was shown that not only short-time Fourier analysis, but short-time Fourier synthesis, as well, can be viewed equivalently in terms of a filter bank or in terms of ordinary Fourier transforms of a set of weighted sequences. The representation of linear filtering in terms of the time-varying frequency response was considered with the result that a class of such filters could be represented as the product of the time-varying frequency response of the filter multiplied by the short-time Fourier transform of the input signal.

The problems of discretizing the short-time Fourier transform representation were discussed. For certain analysis-synthesis filter pairs, it was shown that a signal could be represented by the discrete short-time Fourier transform with no redundancy. Furthermore, a class of linear-filtering operations could be represented as the product of samples of the time-varying frequency response of the filter multiplied by the discrete short-time Fourier transform of the input signal. This representation is of practical significance because the short-time Fourier analysis and synthesis can be implemented efficiently using the FFT algorithm. Finally, for linear time-invariant filters, this representation implies an implementation that is computationally equivalent to the methods of fast convolution.

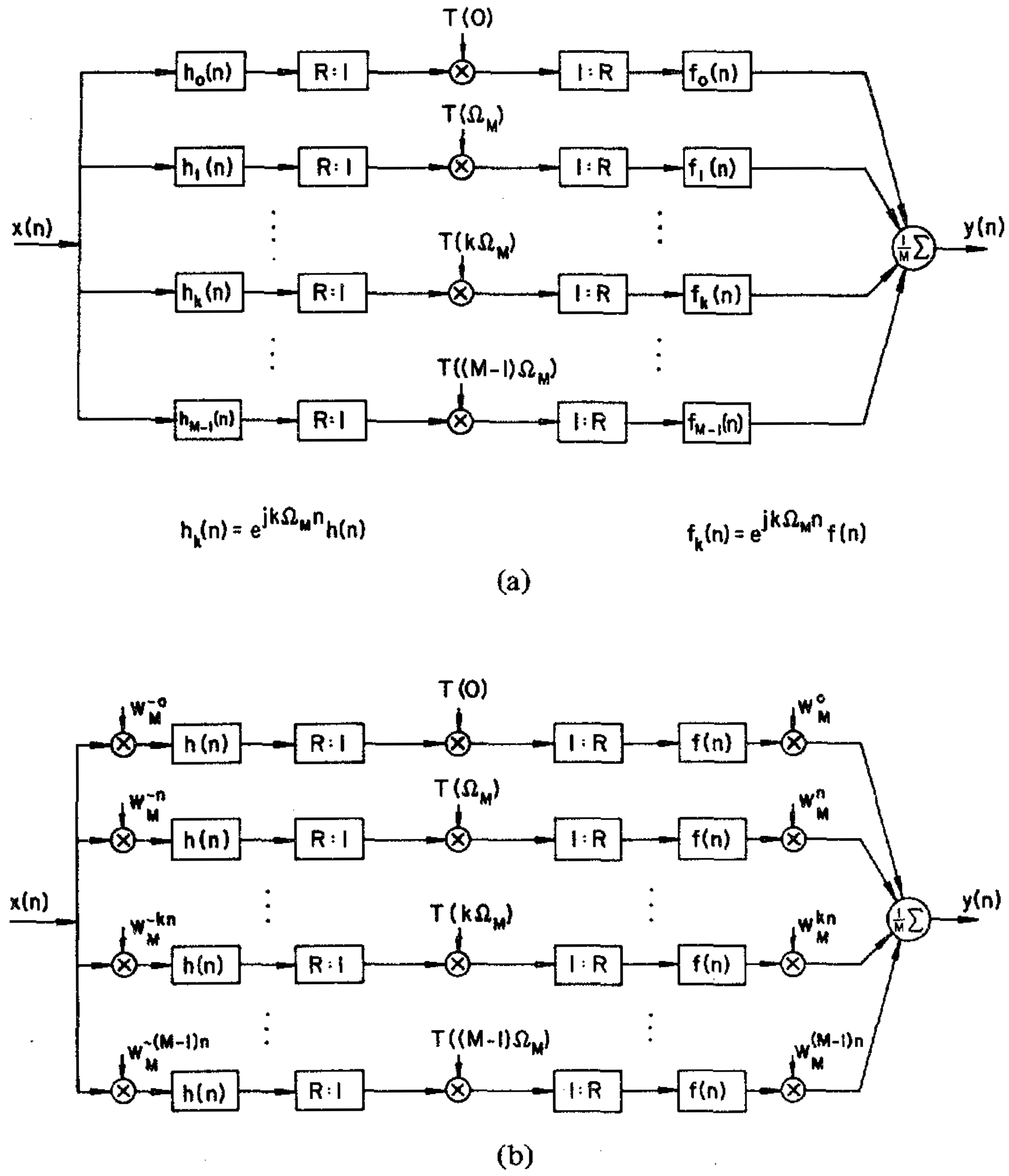


Fig. 10. (a) Filter-bank analog for the conventional method of fast convolution. (b) Filter-bank analog for the short-time Fourier transform method of fast convolution.

ACKNOWLEDGMENT

The author would like to thank A. V. Oppenheim, J. H. McClellan, and K. N. Stevens of the Massachusetts Institute of Technology, Cambridge, and Dr. D. Dudgeon of M.I.T. Lincoln Laboratory who read and commented on earlier versions of this manuscript.

REFERENCES

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

[2] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[3] A. Peled and B. Liu, *Digital Signal Processing, Theory, Design, and Implementation.* New York: Wiley, 1976.

[4] L. A. Zadeh, "A general theory of signal transmission systems," *J. Franklin Inst.*, vol. 253, pp. 293-312, Apr. 1952.

[5] T. Kailath, "Sampling models for linear time-variant filters," Research Laboratory of Electronics Massachusetts Inst. Tech., Cambridge, Tech. Rep. 352, May 1959.

[6] T. Kailath, "Channel characterization: Time-variant dispersive channels," *Lectures on Communication System Theory*, E. J. Baghdaddy, Ed. New York: McGraw-Hill, 1961, pp. 95-123.

[7] A. Gersho and N. DeClaris, "Duality concepts in time-varying linear systems," in *1964 IEEE Int. Conv. Rec.*, part 1, pp. 344-356.

[8] D. Gabor, "Theory of communication," *J. IEE*, part III, pp. 429-411, Nov. 1946.

[9] A. A. Kharkevich, *Spectra and Analysis* (translated from the Russian). New York: Consultants Bureau Enterprises, 1960.

[10] C. J. Weinstein, "Short-time Fourier analysis and its inverse," S.M. thesis, Dep. Elec. Eng., Massachusetts Inst. Tech., Cambridge, 1966.

[11] R. W. Schafer and L. R. Rabiner, "Design and simulation of a speech analysis-synthesis system based on short-time fourier analysis," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 165-174, June 1973.

[12] M. R. Portnoff, "Implementation of the digital phase vocoder using the fast fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 243-248, June 1976.

[13] M. W. Callahan, "Acoustic signal processing based on the short-time spectrum," Ph.D. dissertation, Dep. Comput. Sci., Univ. Utah, Salt Lake City, UT, Tech. Rep. UTEC-CSc-76-209, Mar. 1976.

[14] T. W. Parsons, "Separation of speech from interfering speech by means of harmonic selection," *J. Acoust. Soc. Amer.*, vol. 60, pp. 911-918, Oct. 1976.

[15] J. B. Allen, "Short-term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 235-238, June 1977.

[16] J. B. Allen and L. R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, pp. 1558-1564, Nov. 1977.

[17] M. R. Portnoff, "Time-scale modification of speech based on short-time Fourier analysis," Sc.D. dissertation, Dep. Elec. Eng. Comput. Sci., Massachusetts Inst. Tech., Cambridge, Apr. 1978.

[18] R. M. Lerner, "Representation of signals," *Lectures on Communication System Theory*, E. J. Baghdaddy, Ed. New York: McGraw-Hill, 1961, pp. 204-242.

[19] A. W. Rihaczek, "Signal energy distribution in time and frequency," *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 321-327, May 1968.

[20] M. H. Ackroyd, "Short-time spectra and time-frequency energy distributions," *J. Acoust. Soc. Amer.*, vol. 50, part 1, pp. 1229-1231, May 1971.

[21] M. R. Portnoff, "Magnitude-phase relationships for short-time

Fourier transforms based on Gaussian analysis windows," in *Rec. 1979 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Washington, DC, Apr. 1979, pp. 186–189.

[22] T. G. Stockham, Jr., "High-speed convolution and correlation," in 1966 *Conf. Proc. AFIPS Spring Joint Comput. Conf.* Reprinted in *Digital Signal Processing*, L. R. Rabiner and C. M. Rader, Ed. New York: IEEE Press, 1972.

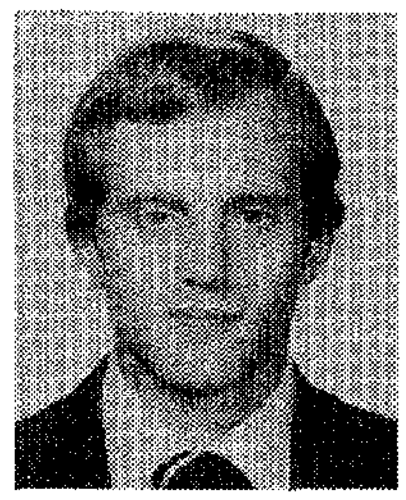
[23] C. R. Patisaul and J. C. Hammett, Jr., "Time-frequency resolution experiment in speech analysis and synthesis," *J. Acoust. Soc. Amer.*, vol. 58, pp. 1296–1307, Dec. 1975.

[24] R. J. Wang, "Optimum window length for the measurement of time-varying power spectra," *J. Acoust. Soc. Amer.*, vol. 52, part 1, pp. 33–38, Jan. 1972.

[25] G. Gambardella, "A contribution to the theory of short-time spectral analysis with nonuniform bandwidth filters," *IEEE Trans. Circuit Theory*, vol. CT-18, July 1971.

[26] J. E. Youngberg and S. F. Boll, "Constant-Q signal analysis and synthesis," in *Rec. 1978 IEEE Int. Conf. Acoust. Speech, Signal Processing*, Tulsa, OK, Apr. 1978, pp. 375–378.

[27] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoust., Speech, Signal Processing*, this issue, pp. 99–102.



Michael R. Portnoff (S'69–M'77) was born in Newark, NJ, on July 1, 1949. He was educated at the Massachusetts Institute of Technology, Cambridge, MA, receiving the S.B., S.M., and E.E. degrees in electrical science and engineering in 1973, and the Sc.D. degree in electrical engineering and computer science in 1978.

From 1969–1971, he was a cooperative student at Bell Telephone Laboratories, Inc. From 1971–1978, he was a Research Assistant in the M.I.T. Research Laboratory of Electronics, Digital Signal Processing Group, and a Teaching Assistant in the M.I.T. Department of Electrical Engineering and Computer Science. From 1978–1979, he was a Research Associate in the Digital Signal Processing Group at M.I.T. In 1979 he joined the University of California Lawrence Livermore Laboratory, where he is currently a Staff Member in the Engineering Research Division. His research interests are in the theory of digital signal processing and its application to speech, image, and seismic signal processing.

Dr. Portnoff received the 1977 IEEE Browder J. Thompson Memorial Prize Award and is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.

On the Implementation of a Short-Time Spectral Analysis Method for System Identification

LAWRENCE R. RABINER, FELLOW, IEEE, AND JONT B. ALLEN, MEMBER, IEEE

Abstract—**Recent work has demonstrated the utility of a short-time spectral analysis approach to the problems of spectral estimation and system identification. In this paper several important aspects of the implementation are discussed. Included is a discussion of the computational effects (e.g., storage, running time) of the various analysis parameters. A computer program is included which illustrates one implementation of the method.**

I. Introduction

The problems of spectral estimation and system identification have been of great importance for a variety of applications. Although classical techniques have had various degrees of success, particular problems often require specialized techniques for the most efficient cost-effective solutions. Recently, a new method for spectral estimation and system identification was proposed based on the theory of short-time spectral analysis [1], [2]. This method was shown to be theoretically equivalent to the classical least squares method when the number of data points (N) was infinite [1]. For finite N the method has the property that the "misalignment" error (between the actual and computed system impulse responses) tends to zero as $1/N$, i.e., the solution rapidly approaches the least squares solution.

The purpose of this paper is to describe one implementation of the method described in [2]. Following a brief review of the basic method (Section II), we describe a DFT implementation in which the relevant quantities used in the analysis equation are computed entirely in the frequency domain (Section III). In Section IV we discuss the issues of computation speed, storage, and accuracy and show that tradeoffs between these factors can be made. Finally, in Section V we present a flowchart of one implementation of the method which is fairly general purpose.

Manuscript received April 20, 1979; revised July 30, 1979.

The authors are with the Acoustics Research Department, Bell Laboratories, Murray Hill, NJ 07974.

II. Review of the Short-Time Spectral Analysis Approach to System Identification

Assume the input to the system to be identified is $x(n)$ and the output of the system [corrupted by additive noise $q(n)$] is $y(n)$, i.e.,

$$y(n) = x(n) * h(n) + q(n) \tag{1}$$

0096-3518/80/0200-0069$00.75 © 1980 IEEE