Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform

JONT B. ALLEN

Abstract—**A theory of short term spectral analysis, synthesis, and modification is presented with an attempt at pointing out certain practical and theoretical questions. The methods discussed here are useful in designing filter banks when the filter bank outputs are to be used for synthesis after multiplicative modifications are made to the spectrum.**

IN THIS paper, some practical and theoretical questions are considered concerning the analysis of and synthesis from a signal's short term spectrum. The short term spectrum, or time spectrum, are those signals which result from analyzing a single input signal with a set of filters which are selective over a range of frequencies [1]-[3]. In the case of analysis by a spectrum analyzer, the filters are either spaced contiguously or one filter is heterodyned over the frequency range of interest. For many applications, this is quite adequate. However, when one is interested in both analysis and synthesis, a more rigorous approach is in order.

In this paper, we shall restrict ourselves to the case of uniformly spaced, symmetric bandpass filters. We are also not concerned with bandwidth reduction. In general, the channel capacity in the short term spectral domain will be greater than that of the original signal. We will be interested, however, in being able to modify the short term spectrum in either its phase or amplitude content without introducing undesired distortion in the synthesized signal.

Previous filter bank analysis-synthesis techniques have been given by Flanagan and Golden [1], Schafer and Rabiner [2], and Portnoff [3]. Our approach differs in several important ways. Previous approaches have used contiguous filter banks in the analysis process. We shall show that this results in an undersampled spectrum and, as a result, synthesis becomes very sensitive to phase or delay modifications. We will then show that by using a properly sampled overlapping filter set, we may avoid this sensitivity. By recognizing the need for a greater number of filters, both the analysis and synthesis procedures are simplified. However, the number of samples of data in the short term spectral domain that results per sample of input data is greater than one; thus, the relevance of the present method to bandwidth reduction remains unclear.

A second important difference between our approach and that of others is during synthesis. All previous authors [1]-[3] have summed the filter outputs for the synthesis; we synthesize in a way that is similar to the overlap add method [4]. As a result, our method does not require an interpolating filter prior to the add, and thus, there is a savings in the amount of additional filtering. In our method, only a small number of adds (four for a Hamming window) per sample are required.

The major advantage of the present scheme is that it allows arbitrary modifications of the short term spectrum. These modifications may be directly interpreted in the time domain as a filter whose impulse response is given by the Fourier transform of the modification. The price paid for allowing modifications of the spectrum will be seen to be an increase in the number of frequency channels required. A modification made prior to the usual method of synthesis, namely, of adding the filter outputs of a contiguous filter set, does *not* satisfy the convolution rule.

Manuscript received September 29, 1976; revised January 25, 1977.
The author is with Bell Laboratories, Murray Hill, NJ 07974.

Analysis of Short Term Spectra

We have defined the short term spectra as an output derived from a bank of filters. At each filter frequency, we require two filters which have the same magnitude response but differ in phase by 90°.

It is well known [2], [3] that a filter bank of this form may be realized by weighting the input signal $x(t)$ by a sliding low-pass filter impulse response $w(t)$ and Fourier transforming the result. Thus, we have

$$X(f,t)=\int_{-\infty}^{\infty} w(t-\tau)\,x(\tau)\,e^{j2\pi f\tau}\,d\tau \tag{1}$$

where $X(f,t)$ is the short term frequency spectrum, t is the time variable, $w(t-\tau)$ is the shifted window, $x(\tau)$ is the input signal, and $\exp(j2\pi f\tau)$ is the complex exponential.

We define $W(f)$ as the Fourier transform of $w(t)$:

$$W(f)=\int_{-\infty}^{\infty} w(\tau)\,e^{j2\pi f\tau}\,d\tau. \tag{2}$$

$W(f)$ is assumed to be small for frequencies above some critical frequency. The short term spectrum $X(f,t)$ is equivalent to frequency shifting the frequency band of $x(t)$ centered at f down to zero frequency with the complex exponential $\exp(j2\pi ft)$ and low-pass filtering the result with the low-pass filter $w(t)$. The resulting $X(f,t)$ is a complex function of time (see Fig. 1 and [1]-[3]). In the applications considered here, $x(t)$ is to be a sampled data signal $x(k)$ and $X(f,t)$ will be found by replacing the Fourier transform by a discrete Fourier transform (DFT).

An important question is relevant at this point, namely, how many frequency and time samples are required to fully represent the data $X(f,t)$ in a sampled data system. This question is answered by applying the Nyquist theorem twice. $w(t)$ has two characteristic "lengths," one in the time domain and one

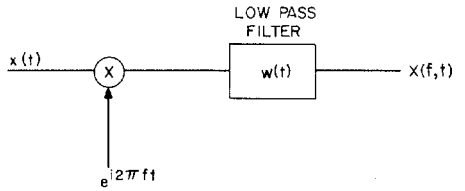


Fig. 1. Analog equivalent of the windowed Fourier transform method of simulating a filter bank.

in the frequency domain. These lengths are bounded by the so-called "uncertainty principle" [5]. We define the length in time to be that time period T over which $w(t)$ is significant. The frequency length is that frequency range F over which $W(f)$ is significant. For the case of the Hamming window,

$$w(t) = \begin{cases} 0.54 + 0.46 \cos(2\pi t/T_0), & -T_0/2 \leqslant t \leqslant T_0/2 \\ 0, & |t| > T_0/2 \end{cases} \tag{3}$$

reasonable definitions of T and F are

$$T = T_0$$

$$F = 4/T_0. \tag{4}$$

In Fig. 2 we see the meaning of these definitions graphically. $w(t)$ is exactly time limited; thus the definition of T is based on the window's nonzero length. $W(f)$, on the other hand, is only approximately bandlimited. Using a 42 dB criteria, $W(f)$ is effectively frequency limited with a maximum frequency of $2/T_0$. Note that our definition of the characteristic frequency length includes negative as well as positive frequencies since the spectrum is symmetric about $f = 0$.

F and T may be used to define sampling rates in the time and frequency domains, respectively. For a fixed f, $X(f, t)$ is a low-pass signal. Thus, from the Nyquist theorem, X may be sampled in time at a rate greater than or equal to twice its highest frequency. An equivalent statement of the Nyquist theorem is that *the density of time samples must be greater than the characteristic frequency length.* We define this sample period as D.

$$D = 1/F. \tag{5}$$

The time samples of interest are then given by

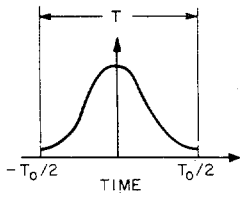
$$t_n = nD. \tag{6a}$$

Equation (5) says that the interval between time samples equals the reciprocal of the characteristic frequency length. F will be called the frame rate. For a Hamming window, D is equal to one quarter of the window's length.

By a completely analogous argument, the continuous frequency spectra may be replaced by a discrete set of frequencies. This is again the Nyquist theorem; however, now it is applied to the time domain. In this case, the sample density is in frequency and the characteristic length is in time. Thus, we may sample in the frequency domain with a spacing of $1/T$. The frequencies of interest are

$$f_m = m/T. \tag{6b}$$

Equation (6b) says that the interval between frequency samples is the reciprocal of the characteristic time length T.



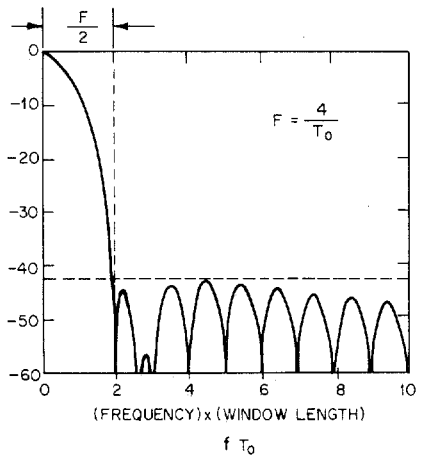


Fig. 2. Basic definitions of the characteristic time and frequency lengths T, F for the case of a Hamming window.

Properly sampled in both the time and frequency domains, the short term spectra may be found by a discrete Fourier transform (DFT) when the input is a sequence $x(k)$. In the discrete case, T is normalized by the sequence sampling period which we call T_s. T_s is chosen so that T is an integral power of 2.

In the following, we denote the time index n and the frequency index m as subscripts. *There is no significant information lost if we replace the short term spectrum X(f, t) with its sampled version X_{nm} by sampling it at the two Nyquist periods 1/F and 1/T.*

$$X_{nm} = X(m/T, n/F), \quad n \text{ and } m \text{ integers.} \tag{7}$$

Using this notation, we may specify the formula for the Nyquist sampled short term spectrum as

$$X_{nm} = \sum_{k=0}^{T-1} w(nD - k)\, x(k)\, e^{j2\pi km/T}. \tag{8}$$

Equation (8) is the sampled version of (1). Note that three different sample periods have been defined: T_s, the sample period for the input sequence $x(k)$; D, the frame period for the bandlimited signals at each frequency f; and $1/T$, the frequency sampling period or filter spacing.

If we define F and F^{-1} as the DFT and its inverse,

$$X_{nm} = F\{w(nD - k)\, x(k)\}. \tag{9}$$

In (9), k and m are the transform variables and n is a parameter. Note that by applying an inverse Fourier transform to (9), we may find a relationship which will be useful later:

$$x(k)\, w(nD - k) = F^{-1}\{X_{nm}\} \tag{10}$$

where

$$F^{-1}\{X_{nm}\} \triangleq \frac{1}{T}\sum_{m=0}^{T-1} X_{nm}\, e^{-j2\pi km/T}. \tag{11}$$

When X_{nm} is interpreted as the output of a bank of filters, each n corresponds to a time sample at the lower sampling rate resulting from the bandlimiting nature ("frequency length") of each filter. Each m corresponds to a different filter frequency of the filter set. The frequency domain may be sampled because of the finite time length of the window.

Next we look at the increase in the data sample density defined as the number of samples of short term spectra generated for each input sample of $x(k)$. For a real signal $x(k)$, assume that we take a segment T long and compute the number of samples of short term spectra required during that period. A DFT is required every D samples and each DFT has T unique frequency values ($T/2$ complex values). Thus, we get

$$T(\text{samples per frame}) \times \frac{T}{D}\ (\text{frames per } T \text{ interval}) \tag{12}$$

samples for every T time samples. The number of samples generated for each input sample is then given by dividing (12) by T:

$$\frac{T}{D} = TF. \tag{13}$$

For the case of a Hamming window, this factor becomes 4.

Synthesis Without Modification

The synthesis procedure is based on the following identity proved in the Appendix for any bandlimited window $w(k)$:

$$\sum_{n=-\infty}^{\infty} w(nD - k) = 1 \tag{14}$$

where we have assumed with no loss of generality that

$$W(0) = D. \tag{15}$$

Relation (14) is exactly true if the window is truly bandlimited to $F/2$. Note that (14) is independent of k.

If we multiply (14) by $x(k)$ and use the inverse DFT relation (10), we obtain the synthesis rule

$$x(k) = \sum_{n=-\infty}^{\infty} F^{-1}\{X_{nm}\} \tag{16}$$

where

$$F^{-1}\{X_{nm}\} = \frac{1}{T}\sum_{m=0}^{T-1} X_{nm}\, e^{-j2\pi km/T}. \tag{17}$$

Equation (16) states that $x(k)$ is a sum over inverse DFT's. It is similar to the overlap add rule, as discussed by Stockham [4], which may be used to do continuous convolution using FFT's; it differs in that the sections are taken as overlapping and are not rectangular windows.

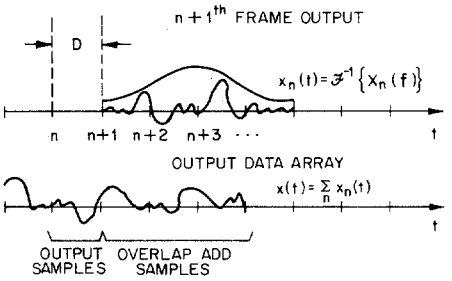


Fig. 3. The synthesis is performed in three steps. First, D samples are output from the left end of the output buffer. Then the output buffer is shifted D samples to the left to the position shown. Finally, the samples $x_n(t)$ are added to those in the output buffer. The steps are then repeated for the next frame.

In practice, during analysis the signal $x(k)$ is shifted by D and the window remains fixed. For synthesis, we define the signals $x_n(k)$

$$x_n(k) = F^{-1}\{X_{nm}\}.$$

During synthesis, the output buffer $x(k)$ is shifted by D and the new samples $x_n(k)$ are added to $x(k)$. (This might be viewed as a block recursive calculation.) This process is then repeated for the next time frame, each frame being defined as a shift of the window by D samples (Fig. 3).

Synthesis with Spectrum Modifications

The analysis-synthesis system described above may be generalized to allow for modifications of the short term spectrum. To do this, we must recognize that spectral modifications are equivalent to filtering. Since the time characteristic length is always increased by filtering, the number of frequency points must be increased as indicated by (6b). This is done by appending zeros to the window $w(k)$ prior to the analysis DFT. The number of points in the DFT is also increased to include the appended zeros. This is similar to appending zeros to the window in the overlap add procedure after sectioning [4]. The number of appended zeros, and thus the DFT length, must be great enough to accommodate the modification which is to be made. When the modification has been properly accounted for, time aliasing will not occur during synthesis. In practice, some time aliasing does occur, but by proper choice of the initial DFT length, the error may be made negligible.

The effect of a fixed spectral modification P_m may be understood by factoring it out as follows:

$$y(k) = \sum_n F^{-1}\{P_m X_{nm}\} \tag{18}$$

$$= \sum_n F^{-1}\{P_m\} * F^{-1}\{X_{nm}\} \tag{19}$$

$$= p(k) * x(k) \tag{20}$$

where

$$p(k) = F^{-1}\{P_m\} \tag{21}$$

and "$*$" denotes convolution. Thus, a fixed modification to the short term spectrum is equivalent to convolution with $p(k)$.

Fixed modifications may not be made in the schemes pre-

sented by other authors [2], [3]. For example, the synthesis techniques of both authors are particularly sensitive to a 180° phase modification in one channel. Such a modification will produce zeros in the transition regions between bands, and will therefore not result in the desired all-pass modification.

The next obvious step is to allow the modification to become a function of time, giving rise to the final synthesis rule:

$$y(k) = \sum_{n=-\infty}^{\infty} F^{-1}\{P_{nm}\, X_{nm}\}. \tag{22}$$

The effect of a time-varying spectral modification is beyond the scope of the present paper.

Finally, we would like to point out some further differences between the present and previous [2], [3] results. It is true, in general, that the short term spectra may be subsampled in either frequency or time and $x(k)$ can still be determined from X_{nm}. When this happens, however, the synthesis is no longer robust to modifications. That this is true may be seen from two examples.

Suppose we generate the time subsampled, short term spectra using a Hamming window which has been shifted by its full period

$$X_{nm} = F\{w(nT_0 - k)\, x(k)\}$$

where $w(k)$ is the Hamming window with length T_0. $x(k)$ may then be recovered by first using (16) and then correcting for the window function. However, when a modification P has occurred, such as a pure delay, the window correction will no longer be correct.

The second example is the case of Portnoff [3]. Our analysis procedure is equivalent to his if we remove all filters that overlap, leaving a contiguous subsampled set. Portnoff has discussed in detail a method of synthesis from this undersampled data set. He requires a restriction on the window function for his synthesis to work. The extra restriction is that the frequency response of the filters add to one. This requirement is particularly sensitive to phase modifications. Portnoff himself points out the problem of this condition not being properly met: "· · · the resulting distortion will be perceived as reverberation in the output signal" [3].

Summary

A theory of short term spectral analysis-synthesis with modifications has been discussed with particular attention being given to the number of time and frequency points required to properly represent the short term spectra. The analysis is performed in frames by a sliding low-pass filter window and a DFT, a frame being defined by the Nyquist period of the bandlimited window. The synthesis is reminiscent of the overlap add process as discussed by Stockham [4], and consists of an inverse DFT and a vector add each frame. Spectral modifications may be included if zeros are appended to the window function prior to the analysis, the number of zeros being equal to the time characteristic length of the modification.

Advantages of the new technique are that modifications may be included and no interpolation is necessary during synthesis. A possible disadvantage is the increased amount of bandwidth required to transmit the short term spectrum as compared to that required to transmit the original signal.

Appendix

We wish to show that given any function $w(k)$ which is bandlimited to a frequency of $1/(2D)$ and normalized as given by (15), then (14) is true, namely, that the sum of any set of samples of $w(k)$ taken with a period D is one.

This is easily proved using the Poisson summation formula [5, eq. (3-56), p. 47]. If $W(f)$ is the Fourier transform of $w(k)$, then

$$\sum_{n=-\infty}^{\infty} w(nD - k) = \frac{1}{D} \sum_{m=-\infty}^{\infty} e^{-j2\pi mk/D}\, W(m/D). \tag{A1}$$

Since W is bandlimited, only the $m = 0$ term is nonzero. Thus, (14) follows from (15).

Acknowledgment

The author would like to acknowledge the helpful comments of D. A. Berkley.

References

[1] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell Syst. Tech. J.*, vol. 45, pp. 1493–1509, Nov. 1966.

[2] 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.

[3] M. R. Portnoff, "Implementation of the digital phase vocoder using the fast Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 243–246, June 1976.

[4] T. G. Stockham, "High-speed convolution and correlation," reprinted in *Digital Signal Processing*, L. R. Rabiner and C. M. Rader, Ed. New York: IEEE Press, 1972, p. 330.

[5] A. Papoulis, *The Fourier Integral and Its Applications*. New York: McGraw-Hill, 1962, p. 63.