Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications. Alfred Mertins

Print ISBN 0-471-98626-7 Electronic ISBN 0-470-84183-4

Chapter 7

Short-Time Fourier Analysis

A fundamental problem in signal analysis is to find the spectral components contained in a measured signal $x(t)$ and/or to provide information about the time intervals when certain frequencies occur. An example of what we are looking for is a sheet of music, which clearly assigns time to frequency, see Figure 7.1. The classical Fourier analysis only partly solves the problem, because it does not allow an assignment of spectral components to time. Therefore one seeks other transforms which give insight into signal properties in a different way. The short-time Fourier transform is such a transform. It involves both time and frequency and allows a time-frequency analysis, or in other words, a signal representation in the time-frequency plane.

7.1 Continuous-Time Signals

7.1.1 Definition

The *short-time Fourier transform* (STFT) is the classical method of time-frequency analysis. The concept is very simple. We multiply $x(t)$, which is to be analyzed, with an analysis window $\gamma^*(t-\tau)$ and then compute the Fourier

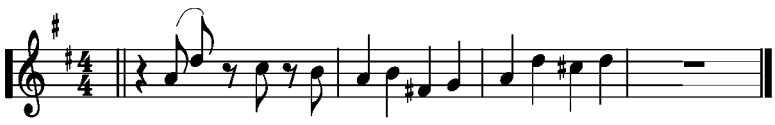

Figure 7.1. Time-frequency representation.

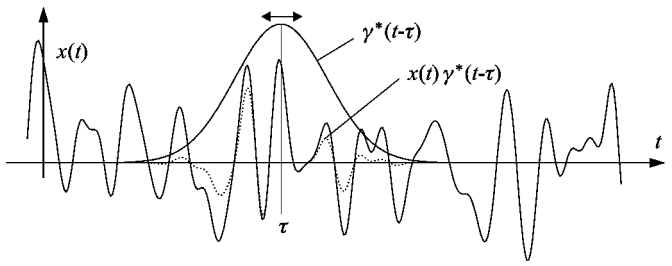

Figure 7.2. Short-time Fourier transform.

transform of the windowed signal:

$$\mathcal{F}_x^{\gamma}(\tau, \omega) = \int_{-\infty}^{\infty} x(t)\ \gamma^*(t-\tau)\ e^{-j\omega t}\ dt. \tag{7.1}$$

The analysis window $\gamma^*(t-\tau)$ suppresses $x(t)$ outside a certain region, and the Fourier transform yields a local spectrum. Figure 7.2 illustrates the application of the window. Typically, one will choose a real-valued window, which may be regarded as the impulse response of a lowpass. Nevertheless, the following derivations will be given for the general complex-valued case.

If we choose the Gaussian function to be the window, we speak of the *Gabor transform*, because Gabor introduced the short-time Fourier transform with this particular window [61].

Shift Properties. As we see from the analysis equation (7.1), a time shift $x(t) \rightarrow x(t-t_0)$ leads to a shift of the short-time Fourier transform by t_0. Moreover, a modulation $x(t) \rightarrow x(t)\ e^{j\omega_0 t}$ leads to a shift of the short-time Fourier transform by ω_0. As we will see later, other transforms, such as the discrete wavelet transform, do not necessarily have this property.

7.1.2 Time-Frequency Resolution

Applying the shift and modulation principle of the Fourier transform we find the correspondence

$$\begin{array}{rcl} \gamma_{\tau;\omega}(t) & := & \gamma(t-\tau)\ e^{j\omega t} \\ & \updownarrow & \\ \Gamma_{\tau;\omega}(\nu) & := & \displaystyle\int_{-\infty}^{\infty} \gamma(t-\tau)\ e^{-j(\nu-\omega)t}\ dt = \Gamma(\nu-\omega)\ e^{-j(\nu-\omega)\tau} \end{array} \tag{7.2}$$

From Parseval's relation in the form

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{\gamma}_{\tau;\omega} \rangle &= \int_{-\infty}^{\infty} x(t)\ \gamma^*(t-\tau)\ e^{-j\omega t}\ dt \\ &= \tfrac{1}{2\pi} \langle \boldsymbol{X}\ , \boldsymbol{\Gamma}_{\tau;\omega} \rangle \\ &= \tfrac{1}{2\pi} \int_{-\infty}^{\infty} X(\nu)\ \Gamma^*(\nu-\omega)\ e^{j(\nu-\omega)\tau} d\nu \end{aligned} \tag{7.3}$$

we conclude

$$\mathcal{F}_x^{\gamma}(\tau,\omega) = e^{-j\omega\tau}\ \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\nu)\ \Gamma^*(\nu-\omega)\ e^{j\nu\tau}\ d\nu. \tag{7.4}$$

That is, windowing in the time domain with $\gamma^*(t-\tau)$ simultaneously leads to windowing in the spectral domain with the window $\Gamma^*(\nu-\omega)$.

Let us assume that $\gamma^*(t-\tau)$ and $\Gamma^*(\nu-\omega)$ are concentrated in the time and frequency intervals

$$[\tau + t_0 - \Delta_t\ ,\ \tau + t_0 + \Delta_t] \tag{7.5}$$

and

$$[\omega + \omega_0 - \Delta_\omega\ ,\ \omega + \omega_0 + \Delta_\omega], \tag{7.6}$$

respectively. Then $\mathcal{F}_x^{\gamma}(\tau,\omega)$ gives information on a signal $x(t)$ and its spectrum $X(\omega)$ in the time-frequency window

$$[\tau + t_0 - \Delta_t\ ,\ \tau + t_0 + \Delta_t]\ \times\ [\omega + \omega_0 - \Delta_\omega\ ,\ \omega + \omega_0 + \Delta_\omega]. \tag{7.7}$$

The position of the time-frequency window is determined by the parameters τ and ω. The form of the time-frequency window is independent of τ and ω, so that we obtain a uniform resolution in the time-frequency plane, as indicated in Figure 7.3.

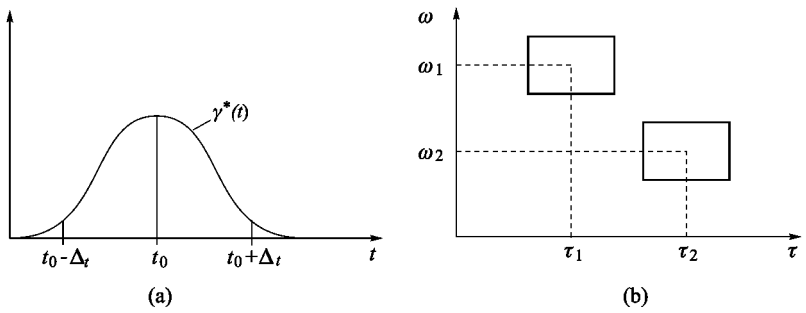

Figure 7.3. Time-frequency window of the short-time Fourier transform.

Let us now have a closer look at the size and position of the time-frequency window. Basic requirements for $\gamma^*(t)$ to be called a time window are $\gamma^*(t) \in L_2(\mathbb{R})$ and $t\,\gamma^*(t) \in L_2(\mathbb{R})$. Correspondingly, we demand that $\Gamma^*(\omega) \in L_2(\mathbb{R})$ and $\omega\,\Gamma^*(\omega) \in L_2(\mathbb{R})$ for $\Gamma^*(\omega)$ being a frequency window. The center t_0 and the radius Δ_t of the time window $\gamma^*(t)$ are defined analogous to the mean value and the standard deviation of a random variable:

$$t_0 = \int_{-\infty}^{\infty} t \cdot \frac{|\gamma(t)|^2}{\|\gamma\|^2}\, dt, \tag{7.8}$$

$$\Delta_t = \left(\int_{-\infty}^{\infty} (t - t_0)^2 \cdot \frac{|\gamma(t)|^2}{\|\gamma\|^2}\, dt \right)^{\frac{1}{2}}. \tag{7.9}$$

Accordingly, the center ω_0 and the radius Δ_ω of the frequency window $\Gamma^*(\omega)$ are defined as

$$\omega_0 = \int_{-\infty}^{\infty} \omega \cdot \frac{|\Gamma(\omega)|^2}{\|\Gamma\|^2}\, d\omega, \tag{7.10}$$

$$\Delta_\omega = \left(\int_{-\infty}^{\infty} (\omega - \omega_0)^2 \cdot \frac{|\Gamma(\omega)|^2}{\|\Gamma\|^2}\, d\omega \right)^{\frac{1}{2}}. \tag{7.11}$$

The radius Δ_ω may be viewed as half of the bandwidth of the filter $\gamma^*(-t)$.

In time-frequency analysis one intends to achieve both high time and frequency resolution if possible. In other words, one aims at a time-frequency window that is as small as possible. However, the uncertainty principle applies, giving a lower bound for the area of the window. Choosing a short time window leads to good time resolution and, inevitably, to poor frequency resolution. On the other hand, a long time window yields poor time resolution, but good frequency resolution.

7.1.3 The Uncertainty Principle

Let us consider the term $(\Delta_t \Delta_\omega)^2$, which is the square of the area in the time-frequency plane being covered by the window. Without loss of generality we may assume $\int t\ |\gamma(t)|^2\ dt = 0$ and $\int \omega\ |\Gamma(\omega)|^2\ d\omega = 0$, because these properties are easily achieved for any arbitrary window by applying a time shift and a modulation. With (7.9) and (7.11) we have

$$(\Delta_t \Delta_\omega)^2 = \frac{\left(\int_{-\infty}^{\infty} t^2\ |\gamma(t)|^2\ dt\right)\left(\int_{-\infty}^{\infty} \omega^2\ |\Gamma(\omega)|^2\ d\omega\right)}{\|\gamma\|^2\ \|\Gamma\|^2}. \tag{7.12}$$

For the left term in the numerator of (7.12), we may write

$$\int_{-\infty}^{\infty} t^2\ |\gamma(t)|^2\ dt = \|\xi\|^2 \tag{7.13}$$

with $\xi(t) = t\,\gamma(t)$. Using the differentiation principle of the Fourier transform, the right term in the numerator of (7.12) may be written as

$$\begin{aligned} \int_{-\infty}^{\infty} \omega^2\ |\Gamma(\omega)|^2\ d\omega &= \int_{-\infty}^{\infty} |\mathcal{F}\left\{\gamma'(t)\right\}|^2\ d\omega \\ &= 2\pi\ \|\gamma'\|^2 \end{aligned} \tag{7.14}$$

where $\gamma'(t) = \frac{d}{dt}\gamma(t)$. With (7.13), (7.14) and $\|\Gamma\|^2 = 2\pi\ \|\gamma\|^2$ we get for (7.12)

$$(\Delta_t \Delta_\omega)^2 = \frac{1}{\|\gamma\|^4}\ \|\xi\|^2\ \|\gamma'\|^2\,. \tag{7.15}$$

Applying the Schwarz inequality yields

$$\begin{aligned} (\Delta_t \Delta_\omega)^2 &\geq \tfrac{1}{\|\gamma\|^4}\ |\ \langle \xi, \gamma' \rangle\ |^2 \\ &\geq \tfrac{1}{\|\gamma\|^4}\ |\Re\left\{\langle \xi, \gamma' \rangle\right\}|^2 \\ &= \tfrac{1}{\|\gamma\|^4}\ \left|\Re\left\{\int_{-\infty}^{\infty} t\ \gamma(t)\ \gamma'^*(t)\ dt\right\}\right|^2. \end{aligned} \tag{7.16}$$

By making use of the relationship

$$\Re\left\{t\,\gamma(t)\,\gamma'^*(t)\right\} = \frac{1}{2}\ t\ \frac{d}{dt}\,|\gamma(t)|^2\,, \tag{7.17}$$

which can easily be verified, we may write the integral in (7.16) as

$$\Re\left\{\int_{-\infty}^{\infty} t\ \gamma(t)\ \gamma'^*(t)\ dt\right\} = \frac{1}{2}\int_{-\infty}^{\infty} t\ \frac{d}{dt}\,|\gamma(t)|^2\ dt. \tag{7.18}$$

Partial integration yields

$$\frac{1}{2}\int_{-\infty}^{\infty} t\ \frac{d}{dt}\,|\gamma(t)|^2\ dt = \frac{1}{2}\ t\ |\gamma(t)|^2\,\Big|_{-\infty}^{\infty} - \frac{1}{2}\int_{-\infty}^{\infty} |\gamma(t)|^2\ dt. \tag{7.19}$$

The property

$$\lim_{|t|\to\infty} t\ |\gamma(t)|^2 = 0, \tag{7.20}$$

which immediately follows from $t\ \gamma(t) \in L_2$, implies that

$$\Re\left\{\int_{-\infty}^{\infty} t\ \gamma(t)\ \gamma'^*(t)\ dt\right\} = -\frac{1}{2}\ \|\gamma\|^2, \tag{7.21}$$

so that we may conclude that

$$(\Delta_t \Delta_\omega)^2\ \geq\ \frac{1}{4}, \tag{7.22}$$

that is

$$\Delta_t \Delta_\omega\ \geq\ \frac{1}{2}. \tag{7.23}$$

The relation (7.23) is known as the *uncertainty principle*. It shows that the size of a time-frequency windows cannot be made arbitrarily small and that a perfect time-frequency resolution cannot be achieved.

In (7.16) we see that equality in (7.23) is only given if $t\,\gamma(t)$ is a multiple of $\gamma'(t)$. In other words, $\gamma(t)$ must satisfy the differential equation

$$t\,\gamma(t) = c\ \gamma'(t), \tag{7.24}$$

whose general solution is given by

$$\gamma(t) = \alpha\ e^{-\frac{t^2}{2\beta^2}}. \tag{7.25}$$

Hence, equality in (7.23) is achieved only if $\gamma(t)$ is the Gaussian function. If we relax the conditions on the center of the time-frequency window of $\gamma(t)$, the general solution with a time-frequency window of minimum size is a modulated and time-shifted Gaussian.

7.1.4 The Spectrogram

Since the short-time Fourier transform is complex-valued in general, we often use the so-called *spectrogram* for display purposes or for further processing stages. This is the squared magnitude of the short-time Fourier transform:

$$S_x(\tau,\omega) = |\ \mathcal{F}_x^{\gamma}(\tau,\omega)|^2 = \left|\ \int_{-\infty}^{\infty} x(t)\ \gamma^*(t-\tau)\ e^{-j\omega t}\ dt\ \right|^2. \tag{7.26}$$

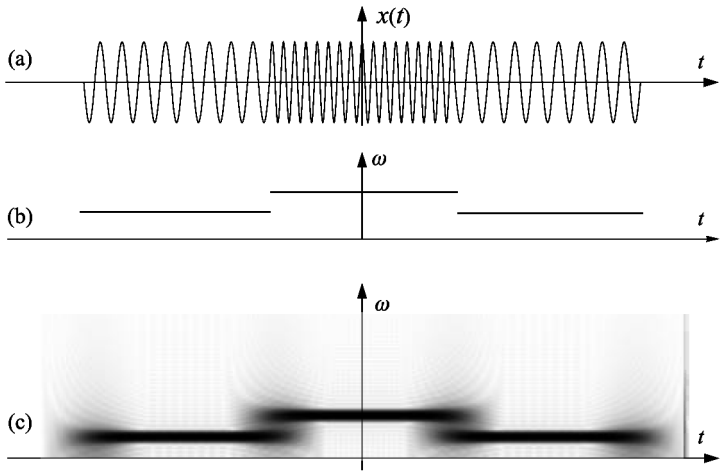

Figure 7.4. Example of a short-time Fourier analysis; (a) test signal; (b) ideal time-frequency representation; (c) spectrogram.

Figure 7.4 gives an example of a spectrogram; the values $S_x(\tau, \omega)$ are represented by different shades of gray. The uncertainty of the STFT in both time and frequency can be seen by comparing the result in Figure 7.4(c) with the ideal time-frequency representation in Figure 7.4(b).

A second example that shows the application in speech analysis is pictured in Figure 7.5. The regular vertical striations of varying density are due to the pitch in speech production. Each striation corresponds to a single pitch period. A high pitch is indicated by narrow spacing of the striations. Resonances in the vocal tract in voiced speech show up as darker regions in the striations. The resonance frequencies are known as the formant frequencies. We see three of them in the voiced section in Figure 7.5. Fricative or unvoiced sounds are shown as broadband noise.

7.1.5 Reconstruction

A reconstruction of $x(t)$ from $\mathcal{F}_x^\gamma(\tau, \omega)$ is possible in the form

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{F}_x^\gamma(\tau, \omega)\ g(t-\tau)\ e^{j\omega t}\ d\tau\ d\omega. \tag{7.27}$$

Here, the synthesis window $g(t)$ must satisfy the condition

$$\int_{-\infty}^{\infty} \gamma^*(t)\ g(t)\ dt\ =\ 1. \tag{7.28}$$

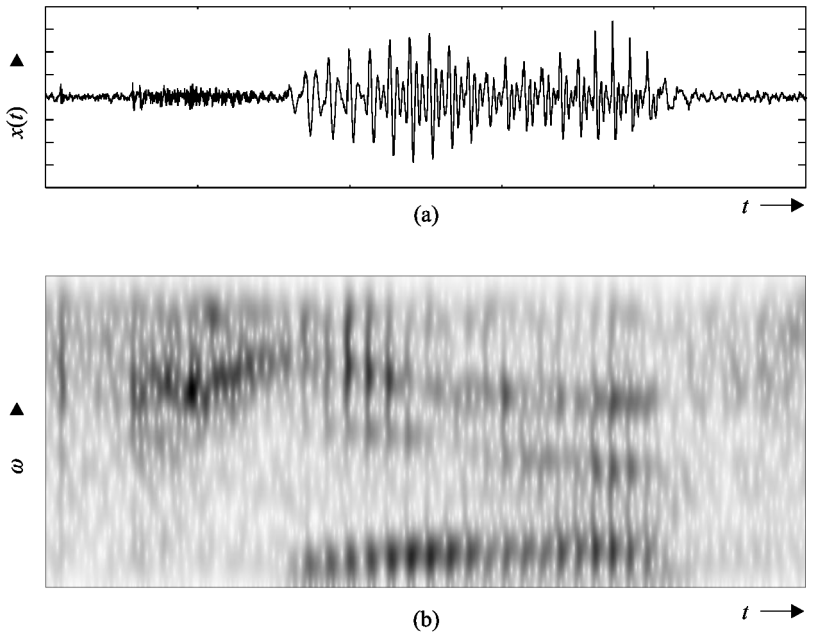

Figure 7.5. Spectrogram of a speech signal; (a) signal; (b) spectrogram.

We can verify this by substituting (7.1) into (7.27) and by rewriting the expression obtained:

$$\begin{aligned} x(t) &= \frac{1}{2\pi}\int\int\int x(t')\ \gamma^*(t'-\tau)\ e^{-j\omega t'}\ dt'\ g(t-\tau)\ e^{j\omega t}\ d\tau\ d\omega \\ &= \int x(t')\int \gamma^*(t'-\tau)\ g(t-\tau)\ \frac{1}{2\pi}\int e^{j\omega(t-t')}\ d\omega\ d\tau\ dt' \qquad (7.29) \\ &= \int x(t')\int \gamma^*(t'-\tau)\ g(t-\tau)\ \delta(t-t')\ d\tau\ dt'. \end{aligned}$$

For (7.29) to be satisfied,

$$\delta(t-t') = \int_{-\infty}^{\infty} \gamma^*(t'-\tau)\ g(t-\tau)\ \delta(t-t')\ d\tau \qquad (7.30)$$

must hold, which is true if (7.28) is satisfied.

The restriction (7.28) is not very tight, so that an infinite number of windows $g(t)$ can be found which satisfy (7.28). The disadvantage of (7.27) is of course that the complete short-time spectrum must be known and must be involved in the reconstruction.

7.1.6 Reconstruction via Series Expansion

Since the transform (7.1) represents a one-dimensional signal in the two-dimensional plane, the signal representation is redundant. For reconstruction purposes this redundancy can be exploited by using only certain regions or points of the time-frequency plane. Reconstruction from discrete samples in the time-frequency plane is of special practical interest. For this we usually choose a grid consisting of equidistant samples as shown in Figure 7.6.

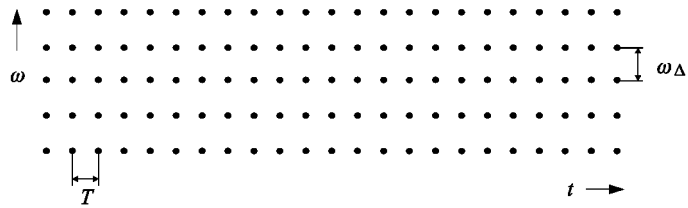

Figure 7.6. Sampling the short-time Fourier transform.

Reconstruction is given by

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \mathcal{F}_x^{\gamma}(mT, k\omega_\Delta)\; g(t - mT)\; e^{jk\omega_\Delta t}. \tag{7.31}$$

The sample values $\mathcal{F}_x^{\gamma}(mT, k\omega_\Delta)$, $m, k \in \mathbb{Z}$ of the short-time Fourier transform are nothing but the coefficients of a series expansion of $x(t)$. In (7.31) we observe that the set of functions used for signal reconstruction is built from time-shifted and modulated versions of the same prototype $g(t)$. Thus, each of the synthesis functions covers a distinct area of the time-frequency plane of fixed size and shape. This type of series expansion was introduced by Gabor [61] and is also called a *Gabor expansion.*

Perfect reconstruction according to (7.31) is possible if the condition

$$\frac{2\pi}{\omega_\Delta} \sum_{m=-\infty}^{\infty} g(t - mT)\; \gamma^*(t - mT - \ell\frac{2\pi}{\omega_\Delta}) = \delta_{\ell 0} \quad \forall\, t \tag{7.32}$$

is satisfied [72], where $\delta_{\ell 0}$ is the Kronecker delta. For a given window $\gamma(t)$, (7.32) represents a linear set of equations for determining $g(t)$. However, here, as with Shannon's sampling theorem, a minimal sampling rate must be guaranteed, since (7.32) can be satisfied only for [35, 72]

$$T\; \omega_\Delta \leq 2\pi. \tag{7.33}$$

Unfortunately, for critical sampling, that is for $T\,\omega_\Delta = 2\pi$, and equal analysis and synthesis windows, it is impossible to have both a good time and a good frequency resolution. If $\gamma(t) = g(t)$ is a window that allows perfect reconstruction with critical sampling, then either Δ_t or Δ_ω is infinite. This relationship is known as the *Balian–Low theorem* [36]. It shows that it is impossible to construct an orthonormal short-time Fourier basis where the window is differentiable and has compact support.

7.2 Discrete-Time Signals

The short-time Fourier transform of a discrete-time signal $x(n)$ is obtained by replacing the integration in (7.1) by a summation. It is then given by [4, 119, 32]

$$\mathcal{F}_x^\gamma(m, e^{j\omega}) = \sum_n x(n)\ \gamma^*(n - mN)\ e^{-j\omega n}. \tag{7.34}$$

Here we assume that the sampling rate of the signal is higher (by the factor $N \in \mathbb{N}$) than the rate used for calculating the spectrum. The analysis and synthesis windows are denoted as γ^* and g, as in Section 7.1; in the following they are meant to be discrete-time. Frequency ω is normalized to the sampling frequency.

In (7.34) we must observe that the short-time spectrum is a function of the discrete parameter m and the continuous parameter ω. However, in practice one would consider only the discrete frequencies

$$\omega_k = 2\pi k/M, \qquad k = 0, \ldots, M-1. \tag{7.35}$$

Then the discrete values of the short-time spectrum can be given by

$$X(m,k) = \sum_n x(n)\ \gamma^*(n - mN)\ W_M^{kn}, \tag{7.36}$$

where

$$X(m,k) = \mathcal{F}_x^\gamma(m, e^{j\omega_k}) \tag{7.37}$$

and

$$W_M = e^{-j2\pi/M}. \tag{7.38}$$

Synthesis. As in (7.31), signal reconstruction from discrete values of the spectrum can be carried out in the form

$$\hat{x}(n) = \sum_{m=-\infty}^{\infty} \sum_{k=0}^{M-1} X(m,k)\ g(n - mN)\ W_M^{-kn}. \tag{7.39}$$

The reconstruction is especially easy for the case $N = 1$ (no subsampling), because then all PR conditions are satisfied for $g(n) = \delta_{n0} \longleftrightarrow G(e^{j\omega}) = 1$ and any arbitrary length-M analysis window $\gamma(n)$ with $\gamma(0) = 1/M$ [4, 119]. The analysis and synthesis equations (7.36) and (7.39) then become

$$X(m,k) = \sum_n x(n)\ \gamma^*(n-m)\ W_M^{kn} \tag{7.40}$$

and

$$\hat{x}(n) = \sum_{k=0}^{M-1} X(n,k)\ W_M^{-kn}. \tag{7.41}$$

This reconstruction method is known as *spectral summation*. The validity of $\hat{x}(n) = x(n)$ provided $\gamma(0) = 1/M$ can easily be verified by combining these expressions.

Regarding the design of windows allowing perfect reconstruction in the subsampled case, the reader is referred to Chapter 6. As we will see below, the STFT may be understood as a DFT filter bank.

Realizations using Filter Banks. The short-time Fourier transform, which has been defined as the Fourier transform of a windowed signal, can be realized with filter banks as well. The analysis equation (7.36) can be interpreted as filtering the modulated signals $x(n)W_M^{kn}$ with a filter

$$h(n) = \gamma^*(-n). \tag{7.42}$$

The synthesis equation (7.39) can be seen as filtering the short-time spectrum with subsequent modulation. Figure 7.7 shows the realization of the short-time Fourier transform by means of a filter bank. The windows $g(n)$ and $\gamma(n)$ typically have a lowpass characteristic.

Alternatively, signal analysis and synthesis can be carried out by means of equivalent bandpass filters. By rewriting (7.36) as

$$X(m,k) = W_M^{kmN} \sum_n x(n)\ \gamma^*(n-mN)\ W_M^{k(n-mN)} \tag{7.43}$$

we see that the analysis can also be realized by filtering the sequence $x(n)$ with the bandpass filters

$$h_k(n) = \gamma^*(-n)\ W_M^{-kn}, \qquad k = 0, \ldots, M-1 \tag{7.44}$$

and by subsequent modulation.

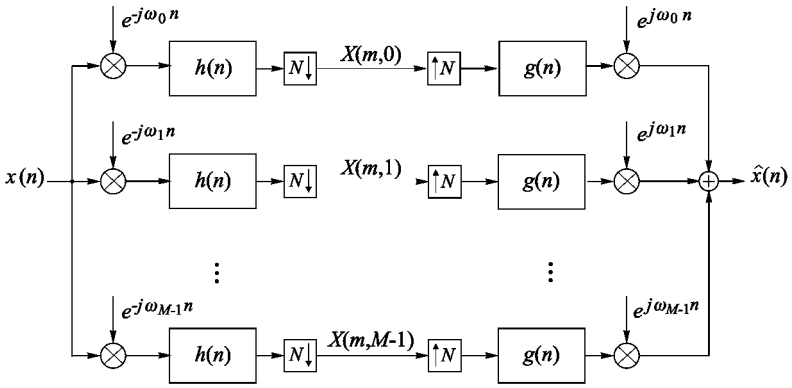

Figure 7.7. Lowpass realization of the short-time Fourier transform.

Rewriting (7.39) as

$$\hat{x}(n) = \sum_{m=-\infty}^{\infty} \sum_{k=0}^{M-1} X(m,k)\, W_M^{-kmN}\, g(n-mN)\, W_M^{-k(n-mN)} \tag{7.45}$$

shows that synthesis can be achieved with modulated filters as well. To accomplish this, first the short-time spectrum is modulated, then filtering with the bandpass filters

$$g_k(n) = g(n)\, W_M^{-kn}, \qquad k = 0, \ldots, M-1, \tag{7.46}$$

takes place; see Figure 7.8.

We realize that the short-time Fourier transform belongs to the class of modulated filter banks. On the other hand, it has been introduced as a transform, which illustrates the close relationship between filter banks and short-time transforms.

The most efficient realization of the STFT is achieved when implementing it as a DFT polyphase filter bank as outlined in Chapter 6.

7.3 Spectral Subtraction based on the STFT

In many real-word situations one encounters signals distorted by additive noise. Several methods are available for reducing the effect of noise in a more or less optimal way. For example, in Chapter 5.5 optimal linear filters that yield a maximum signal-to-noise ratio were presented. However, linear methods are

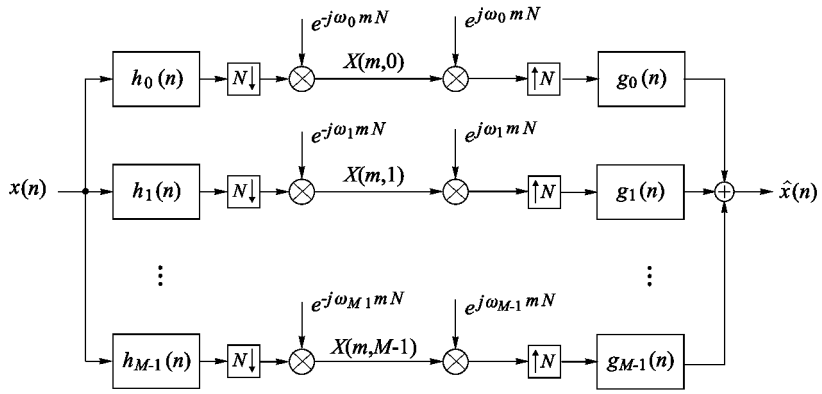

Figure 7.8. Bandpass realization of the short-time Fourier transform.

not necessarily the optimal ones, especially if a subjective signal quality with respect to human perception is of importance. Spectral subtraction is a nonlinear method for noise reduction, which is very well suited for the restoration of speech signals.

We start with the model

$$y(t) = x(t) + n(t), \tag{7.47}$$

where we assume that the additive noise process $n(t)$ is statistically independent of the signal $x(t)$. Assuming that the Fourier transform of $y(t)$ exists, we have

$$Y(\omega) = X(\omega) + N(\omega) \tag{7.48}$$

in the frequency domain. Due to statistical independence between signal and noise, the energy density may be written as

$$|Y(\omega)|^2 = |X(\omega)|^2 + |N(\omega)|^2. \tag{7.49}$$

If we now assume that $E\left\{|N(\omega)|^2\right\}$ is known, the least squares estimate for $|X(\omega)|^2$ can be obtained as

$$|\hat{X}(\omega)|^2 = |Y(\omega)|^2 - E\left\{|N(\omega)|^2\right\}. \tag{7.50}$$

In spectral subtraction, one only tries to restore the magnitude of the spectrum, while the phase is not attached. Thus, the denoised signal is given in the frequency domain as

$$\hat{X}(\omega) = |\hat{X}(\omega)| \angle Y(\omega). \tag{7.51}$$

Keeping the noisy phase is motivated by the fact that the phase is of minor importance for speech quality.

So far, the time dependence of the statistical properties of the signal and the noise process has not been considered. Speech signals are highly nonstationary, but within intervals of about 20 msec, the signal properties do not change significantly, and the assumption of stationarity is valid on a short-time basis. Therefore, one replaces the above spectra by the short-time spectra computed by the STFT. Assuming a discrete implementation, this yields

$$Y(m,k) = X(m,k) + N(m,k), \tag{7.52}$$

where m is the time and k is the frequency index. $Y(m,k)$ is the STFT of $y(m)$.

Instead of subtracting an average noise spectrum $E\{|N(\omega)|^2\}$, one tries to keep track of the actual (time-varying) noise process. This can for instance be done by estimating the noise spectrum in the pauses of a speech signal. Equations (7.50) and (7.51) are then replaced by

$$|\hat{X}(m,k)|^2 = |Y(m,k)|^2 - |\widehat{N(m,k)}|^2 \tag{7.53}$$

and

$$\hat{X}(m,k) = |\hat{X}(m,k)| \angle Y(m,k), \tag{7.54}$$

where $|\widehat{N(m,k)}|^2$ is the estimated noise spectrum.

Since it cannot be assured that the short-time spectra satisfy $|Y(m,k)|^2 - |\widehat{N(m,k)}|^2 > 0$, $\forall m,k$, one has to introduce further modifications such as clipping. Several methods for solving this problem and for keeping track of the time-varying noise have been proposed. For more detail, the reader is referred to [12, 50, 51, 60, 49]. Finally, note that a closely related technique, known as wavelet-based denoising, will be studied in Section 8.10.