this is straightforward for the discrete-time STFT, a number of important STFT concepts are introduced for addressing the more challenging task of synthesis from the discrete STFT. The basic theory part of the chapter is essentially concluded in Section 7.4 in treating the magnitude of the STFT as a transform in its own right. Next, in Section 7.5, we consider the important practical problem of estimating a signal from a modified STFT or STFT magnitude that does not satisfy the definitional constraints of the STFT, leading to many practical applications of the STFT and STFT magnitude in speech processing. The particular applications of time-scale modification and noise reduction are introduced in Section 7.6.

7.2 Short-Time Analysis

In this section, following the development in [13],[20], we explore two different views of the STFT: (1) the Fourier transform view and (2) the filter bank view. We begin with the first perspective, which draws on the STFT representation of a sequence being analogous to that of the Fourier transform.

7.2.1 Fourier Transform View

The expression for the discrete-time STFT at time n was given in Chapter 3 as[1]

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x[m]w[n - m]e^{-j\omega m}, \tag{7.1}$$

where $w[n]$ is assumed to be non-zero only in the interval $[0, N_w - 1]$ and is referred to as the *analysis window* or sometimes as the *analysis filter* for reasons that will become clear later in this chapter. The sequence $f_n[m] = x[m]w[n - m]$ is called a short-time section of $x[m]$ at time n. This sequence is obtained by time-reversing the analysis window, $w[m]$, shifting the result by n points, and multiplying it with $x[m]$. With the short-time section for time n, we can take its Fourier transform to obtain the frequency function $X(n, \omega)$. This series of operations is illustrated in Figure 7.1. To obtain $X(n + 1, \omega)$, we slide the time-reversed analysis window one point from its previous position, multiply it with $x[m]$, and take the Fourier transform of the resulting short-time section. Continuing this way, we generate a set of discrete-time Fourier transforms that together constitute the discrete-time STFT. Typically, the analysis window is selected to have a much shorter duration than the signal $x[n]$ for which the STFT is computed; as we have seen for a speech waveform, the window duration is typically set at about 20-30 ms or a few pitch periods.

By analogy with the discrete Fourier transform (DFT), the discrete STFT is obtained from the discrete-time STFT through the following relation:

$$X(n, k) = X(n, \omega)|_{\omega = \frac{2\pi}{N}k}, \tag{7.2}$$

where we have sampled the discrete-time STFT with a *frequency sampling interval*[2] of $\frac{2\pi}{N}$ in order to obtain the discrete STFT. We refer to N as the *frequency sampling factor*. Substituting

[1] We have changed our notation slightly from that in Chapter 3, where we denoted the STFT by $X(\omega, \tau)$.

[2] More strictly, the sampling occurs over one period so that $X(n, k) = X(n, \omega)|_{\omega = \frac{2\pi}{N}k}$ for $k = 0, 1, \ldots N - 1$, and zero elsewhere, but here we think of the DFT as periodic with period N [14].