We begin in Section 13.2 with the Fourier transform and filtering perspectives of the short-time Fourier transform (STFT) as bases for approaches to the problems of reducing additive noise and convolutional distortion. We describe, as a foundation, *spectral subtraction* that operates on STFT short-time sections for additive noise reduction, and *cepstral mean subtraction* that operates on STFT bandpass filter outputs for removing stationary convolutional distortion. In Section 13.3, the *Wiener filter* and its adaptive renditions for additive noise removal are then developed. A variety of Wiener-filter-based approaches are given, including a method that adapts to spectral change in a signal to help preserve signal nonstationarity while exploiting auditory temporal masking, and also a stochastic-theoretical approach to obtain a minimum mean-squared error estimate of the desired spectral magnitude. Sections 13.4 and 13.5 then develop all-pole model-based and further auditory-based approaches to additive noise reduction, respectively. The auditory-based methods use frequency-domain perceptual masking principles to conceal annoying residual noise under the spectral components of interest. Section 13.6 next generalizes cepstral mean subtraction (CMS), introduced in Section 13.2, to reduce *time-varying* convolutional distortion. This generalization, commonly referred to as *RASTA*, (along with CMS) can be viewed as homomorphic filtering along the time dimension of the STFT, rather than with respect to its frequency dimension. These approaches represent a fascinating application of homomorphic filtering theory in a domain different from that studied in Chapter 6. Moreover, we will see that CMS and RASTA can be viewed as members of a larger class of enhancement algorithms that filter *nonlinearly* transformed temporal envelopes of STFT filter outputs.

13.2 Preliminaries

In this section, we first formulate the additive noise and convolutional distortion problems in the context of the STFT, from both the Fourier transform and filtering viewpoints introduced in Chapter 7. With this framework, we then develop the method of spectral subtraction for additive noise suppression and cepstral mean subtraction for reducing a stationary convolutional distortion.

13.2.1 Problem Formulation

Additive Noise — Let $y[n]$ be a discrete-time noisy sequence

$$y[n] = x[n] + b[n] \tag{13.1}$$

where $x[n]$ is the desired signal, which we also refer to as the "object," and $b[n]$ is the unwanted background noise. For the moment, we assume $x[n]$ and $b[n]$ to be wide-sense stationary, uncorrelated random processes with power spectral density functions (Appendix 5.A) denoted by $S_x(\omega)$ and $S_b(\omega)$, respectively. One approach to recovering the desired signal $x[n]$ relies on the additivity of the power spectra, i.e.,

$$S_y(\omega) = S_x(\omega) + S_b(\omega). \tag{13.2}$$

With STFT analysis, however, we work with the short-time segments given by

$$y_{pL}[n] = w[pL - n](x[n] + b[n])$$

where L is the frame length and p is an integer, which in the frequency domain is expressed as

$$Y(pL, \omega) = X(pL, \omega) + B(pL, \omega)$$