of the $\log[Y(pL, \omega)]$ is computed and subtracted directly.[5] Because, in practice, the mean is computed over a finite number of frames, we can think of CMS as a highpass, non-causal FIR filtering operation [2],[23].

13.3 Wiener Filtering

An alternative to spectral subtraction for recovering an object sequence $x[n]$ corrupted by additive noise $b[n]$, i.e., from a sequence $y[n] = x[n] + b[n]$, is to find a linear filter $h[n]$ such that the sequence $\hat{x}[n] = y[n] * h[n]$ minimizes the expected value of $(\hat{x}[n] - x[n])^2$. Under the condition that the signals $x[n]$ and $b[n]$ are uncorrelated and stationary, the frequency-domain solution to this stochastic optimization problem is given by the suppression filter (Exercise 13.2)

$$H_s(\omega) = \frac{S_x(\omega)}{S_x(\omega) + S_b(\omega)} \tag{13.10}$$

which is referred to as the *Wiener filter* [32]. When the signals $x[n]$ and $b[n]$ meet the conditions under which the Wiener filter is derived, i.e., uncorrelated and stationary object and background, the Wiener filter provides noise suppression without considerable distortion in the object estimate and background residual. The required power spectra, $S_x(\omega)$ and $S_b(\omega)$, can be estimated by averaging over multiple frames when sample functions of $x[n]$ and $b[n]$ are provided. Typically, however, the desired signal and background are nonstationary in the sense that their power spectra change over time, i.e., they can be expressed as time-varying functions $S_x(n, \omega)$ and $S_b(n, \omega)$. Thus, ideally, each frame of the STFT is processed by a different Wiener filter. For the simplifying case of a stationary background, we can express the *time-varying* Wiener filter as

$$H_s(pL, \omega) = \frac{\hat{S}_x(pL, \omega)}{\hat{S}_x(pL, \omega) + \hat{S}_b(\omega)}$$

where $\hat{S}_x(pL, \omega)$ is an estimate of the time-varying power spectrum of $x[n]$, $S_x(n, \omega)$, on each frame, and $\hat{S}_b(\omega)$ is an estimate of the power spectrum of a stationary background, $S_b(\omega)$. The time-varying Wiener filter can also be expressed as (Exercise 13.2)

$$H_s(pL, \omega) = \left[1 + \frac{1}{R(pL, \omega)}\right]^{-1} \tag{13.11}$$

with a signal-to-noise ratio $R(pL, \omega) = \hat{S}_x(pL, \omega)/\hat{S}_b(\omega)$. A comparison of the suppression curves for spectral subtraction and Wiener filtering is shown in Figure 13.1, where we see the attenuation of low SNR regions relative to the high SNR regions to be somewhat stronger for the Wiener filter, consistent with the filter in Equation (13.8) being a compressed (square-rooted) form of that in Equation (13.11). A second important difference from spectral subtraction is that the Wiener filter does not invoke an absolute thresholding. Finally,

[5] In practice, only the STFT magnitude is used in recognition applications and the 0th cepstral value is computed by computing the mean of $\log|Y(pL, \omega)|$ along p, rather than computing an explicit inverse Fourier transform. Often, however, in recognition applications, the cepstrum of $\log|Y(pL, \omega)|$ is computed with respect to ω to obtain a cepstral feature vector for each frame. Equivalently, one can then subtract the mean cepstrum (across frames) to remove the distortion component, and hence we have an alternative and, perhaps, more legitimate motivation for the nomenclature *cepstral mean subtraction*.