## Filter Banks: The Second Interpretation of the STDTFT

Recall our equation for the STDTFT:

$$\tilde{S}(\omega, n_0) = \sum_{n=-\infty}^{\infty} \tilde{s}[n, n_0] e^{-j\omega n}$$

$$= \sum_{n=-\infty}^{\infty} s[n] w[n_0 - n] e^{-j\omega n}$$

1. This has the form of a convolution of $s[n]e^{-j\omega n}$ with $w[n]$, where the output variable is $n_0$
2. Using the commutative property of convolution, we can rewrite this as

$$\tilde{S}(\omega, n_0) = \sum_{n=-\infty}^{\infty} s[n_0 - n] w[n] e^{-j\omega(n_0 - n)}$$

3. If we think of $n_0$ as our time variable, we can express this as

$$\tilde{S}(\omega, n_0) = e^{-j\omega n_0} \sum_{n=-\infty}^{\infty} s[n_0 - n] w[n] e^{j\omega n}$$

$$= e^{-j\omega n_0} \left( s[n_0] * \left( w[n_0] e^{j\omega n_0} \right) \right)$$

$$= e^{-j\omega n_0} \left( s[n_0] * h_\omega[n_0] \right)$$

where

$$h_\omega[n_0] \triangleq w[n_0] e^{j\omega n_0}$$

4. Since we are now regarding $n_0$ as our time-variable, we will drop the zero subscript, and replace $n_0$ by $n$

5. Then we have

$$\tilde{S}(\omega,n) = e^{-j\omega n}\left(s[n] * h_\omega[n]\right)$$

and

$$h_\omega[n] \triangleq w[n]e^{j\omega n}$$

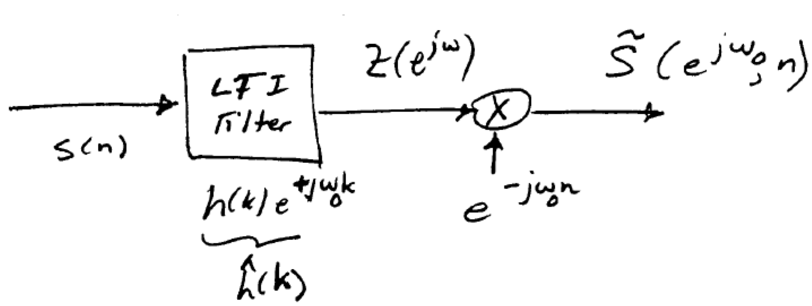
6. Further, we now wish to regard the frequency $\omega$ as being fixed at $\omega = \omega_0$; so we write

$$\tilde{S}(\omega_0,n) = e^{-j\omega_0 n}\left(s[n] * h_{\omega_0}[n]\right)$$

and

$$h_{\omega_0}[n] \triangleq w[n]e^{j\omega_0 n}$$

7. Thus, looks like we are filtering $s[n]$ with $h_{\omega_0}[n]$ followed with multiplication by the complex exponential $e^{-j\omega_0 n}$.




8. To see what is really happening here, we recall the modulation property for the DTFT:

$$x[n]e^{j\omega_0 n} \overset{\text{DTFT}}{\leftrightarrow} X\left(\omega - \omega_0\right)$$

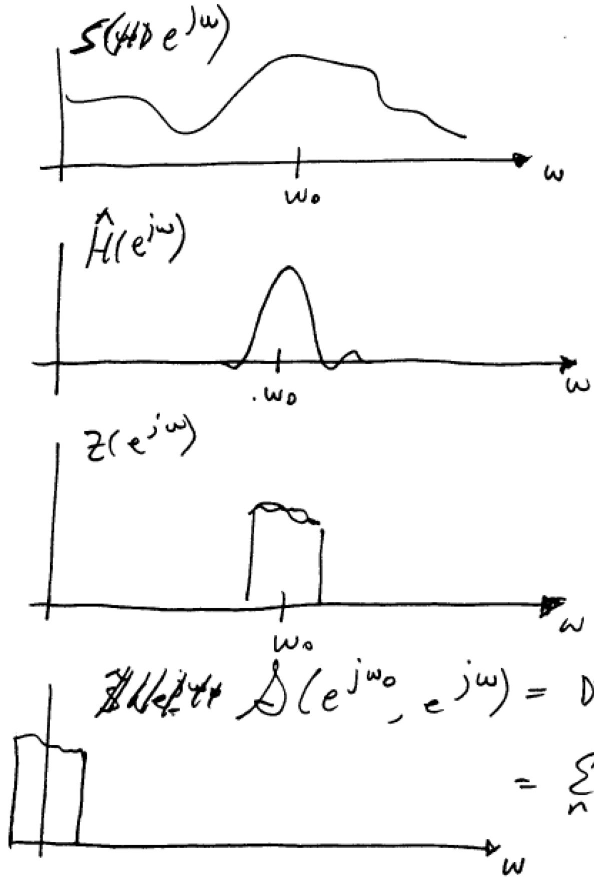
9. So we can also say that we are
   a. shifting the DTFT $W(\omega)$ of the window $w[n]$ to frequency $\omega_0$

b. multiplying the spectrum $S(\omega)$ of the speech waveform $s[n]$ by this filter frequency response, which selects a region of $S(\omega)$ centered at $\omega_0$ with width equal to the bandwidth of $w[n]$, or equivalently the support of $W(\omega)$

c. then, shifting the resulting frequency domain product by $-\omega_0$ down to the baseband

10. To express this directly in the frequency domain, we take the DTFT of $\tilde{S}(\omega_0, n)$ with respect to the time variable $n$, treating $\omega_0$ as a constant:
$$\tilde{\mathbf{S}}(\omega, \omega_0) = \sum_{n=-\infty}^{\infty} \tilde{S}(\omega_0, n) e^{-j\omega n}$$
11. This is a bit strange, because $\tilde{S}(\omega_0, n)$ is already a DTFT of a signal. However, now, we are viewing the original frequency as fixed parameter $\omega_0$, and we are taking the DTFT with respect to the location of the center of the shifted window, which we now call $n$
12. Applying standard properties of the DTFT to
$$\tilde{S}(\omega_0, n) = e^{-j\omega_0 n}\left(s[n] * h_{\omega_0}[n]\right)$$
we obtain
$$\tilde{\mathbf{S}}(\omega, \omega_0) = H_{\omega_0}(\omega + \omega_0) S(\omega + \omega_0)$$
where
$$H_{\omega_0}(\omega) = W(\omega - \omega_0)$$

Graphically, we have




$$\mathcal{S}(e^{j\omega_0}, e^{j\omega}) = \mathrm{DTFT}\{\tilde{S}(e^{j\omega_0}, n)\}$$
$$= \sum_n \tilde{S}(e^{j\omega_0}, n) e^{-j\omega n}$$

**This interpretation is particularly relevant to the case of a narrowband spectrogram, because**

1. The window $w[n]$ is very long, implying that the filter frequency response $H_{\omega_0}(\omega) = W(\omega - \omega_0)$ is very localized
2. Hence the nomenclature "narrowband"
3. In contrast to the wideband spectrogram, which gives fine resolution along the time axis, the narrowband spectrogram gives fine resolution along the frequency axis, and can resolve the discrete spectral lines separated by frequency $2\pi / P$ radians/sample in a voiced phoneme
4. We call this a filter bank because for each different value of $\omega_0$, we get a different filter.