EE 438 Digital Signal Processing with Applications: Short Time Fourier Analysis

Prof Jan P Allebach
School of Electrical and Computer Engineering
Purdue University
West Lafayette IN 47907-1285
allebach@ecn.purdue.edu

March 31, 2000

We will be using both τ and t to denote the variable "time" in this derivation. Let $s(\tau)$ denote our continuous-time signal and $h(t-\tau)$ the window used to limit $s(\tau)$ to the desired region of interest centered at time t. We define the short-time continuous-time Fourier transform (STCTFT) according to

$$\mathcal{S}(f,t) = \int_{-\infty}^{\infty} s(\tau)h(t-\tau)e^{-j2\pi f\tau}d\tau. \tag{1}$$

We use the calligraphic letter $\mathcal{S}$ to distinguish the STCTFT of $s(\tau)$ from the usual CTFT $S(f)$ of $s(\tau)$. Consider the following model for a speech signal

$$s(\tau) = \text{rep}_P\left[\text{rect}(\tau/D)\right]\cos(2\pi f_1\tau). \tag{2}$$

What we have here is a train of cosine pulses. The function $\text{rect}(\tau/D)\cos(2\pi f_1\tau)$ represents the vocal tract impulse response. The vocal tract has a single resonant frequency f_1. The parameter P represents the pitch period. Generally, $1/P << f_1$.

To find the STCTFT of $s(\tau)$, we observe that with t fixed, (1) is simply the CTFT with respect to τ of the product of $s(\tau)$ and $h(t-\tau)$. Thus, in the frequency domain, we will simply have the convolution of the Fourier transforms of these two terms. Applying the usual transform relations to (2), we obtain

$$\begin{aligned} S(f) &= \frac{1}{P}\text{comb}_{\frac{1}{P}}\left[D\text{sinc}(D(f)\right] * \frac{1}{2}\left(\delta(f-f_1)+\delta(f+f_1)\right) \\ &= \frac{D}{2P}\sum_{k=-\infty}^{\infty}\text{sinc}\left(\frac{Dk}{P}\right)\delta\left(f-\frac{k}{P}\right) * \frac{1}{2}\left(\delta(f-f_1)+\delta(f+f_1)\right) \\ &= \frac{D}{2P}\sum_{k=-\infty}^{\infty}\text{sinc}\left(\frac{Dk}{P}\right)\left(\delta\left(f-f_1-\frac{k}{P}\right)+\delta\left(f+f_1-\frac{k}{P}\right)\right) \end{aligned} \tag{3}$$

We then take the CTFT of $h(t-\tau)$ with respect to τ where t is fixed. Noting that $h(t-\tau) = h(-(\tau - t))$, we apply the reflection property $h(-\tau) \overset{CTFT}{\leftrightarrow} H(-f)$, followed by shifting to get

$$\mathcal{F}\{h(t-\tau)\} = H(-f)e^{-j2\pi ft} \tag{4}$$

where it should again be emphasized that these transforms are taken with respect to τ with t fixed.

We are finally ready to perform the convolution between (3) and (4), which yields

$$\begin{aligned}
\mathcal{S}(f,t) &= \int_{-\infty}^{\infty} S(\rho)H(-(f-\rho))e^{-j2\pi(f-\rho)t}d\rho \\
&= \frac{D}{2P}\sum_{k=-\infty}^{\infty} \text{sinc}\left(\frac{Dk}{P}\right)\int_{-\infty}^{\infty}\left(\delta\left(\rho - f_1 - \frac{k}{P}\right) + \delta\left(\rho + f_1 - \frac{k}{P}\right)\right)H(-(f-\rho))e^{-j2\pi(f-\rho)t}d\rho \\
&= \frac{D}{2P}\sum_{k=-\infty}^{\infty} \text{sinc}\left(\frac{Dk}{P}\right)\left[H\left(-\left(f - f_1 + \frac{k}{P}\right)\right)e^{-j2\pi(f-f_1+\frac{k}{P})t}\right. \\
&\quad \left. + H\left(-\left(f + f_1 + \frac{k}{P}\right)\right)e^{-j2\pi(f+f_1+\frac{k}{P})t}\right]
\end{aligned} \tag{5}$$

If the window is symmetric, i.e. $h(-t) = h(t)$, this result simplifies somewhat

$$\begin{aligned}
\mathcal{S}(f,t) &= \frac{D}{2P}\sum_{k=-\infty}^{\infty} \text{sinc}\left(\frac{Dk}{P}\right)\left[H\left(f - f_1 + \frac{k}{P}\right)e^{-j2\pi(f-f_1+\frac{k}{P})t}\right. \\
&\quad \left. + H\left(f + f_1 + \frac{k}{P}\right)e^{-j2\pi(f+f_1+\frac{k}{P})t}\right]
\end{aligned} \tag{6}$$

We distinguish two cases. In the first case, the duration of the window $h(\tau)$ is much greater than the pitch period P. In this case, $H(f)$ will be narrow compared to the pitch frequency $1/P$; and the terms under the sum over k in (6) will not overlap. Then, if we are only interested in the magnitude of the STCTFT, we can write

$$\mathcal{S}(f,t) = \frac{D}{2P}\sum_{k=-\infty}^{\infty}\left|sinc\left(\frac{Dk}{P}\right)\right|\left(\left|H\left(f - f_1 + \frac{k}{P}\right)\right| + \left|H\left(f + f_1 + \frac{k}{P}\right)\right|\right) \tag{7}$$

This corresponds to a narrowband spectrogram. The time dependence has disappeared; and the spectral lines separated by $1/P$ due to the pitch period are clearly visible.

In the second case, the duration of the window $h(\tau)$ is about the same as the pitch period P. In this case, we see blips of energy as the window slides along the time axis (t) over the individual vocal tract pulses. Along the frequency axis, The CTFT $H(f)$ of the window has broadened to such an extent that the terms under the sum over k overlap somewhat; and the individual spectral lines are no longer visible. Instead, we see an envelope corresponding to the CTFT $\text{sinc}(Df)$ of the vocal tract response envelope $\text{rect}(t/D)$.