ECE 537 Homework 2

Due: September 11, Urbana time

Submit your answers on Gradescope by no later than the end of the day on September 11, Urbana time. You may write your answers on this page, or use your own paper. You may submit either typeset or handwritten answers, as long as your answers are legible. Full credit is given if you show your work and your answer is mostly correct. Full credit is not given for an answer provided without any derivation, though the distributed solutions will be in that form.

1. Voiced speech is not perfectly periodic; every voiced period contains some energy caused by aspiration noise. For this reason, compared to the models we've discussed in class, a slightly better model of voiced speech is

$$x[n] = v[n] + \sum_{k=-K}^{K} a_k e^{j(2\pi k\omega_0(n-n_0)+\theta_k)},$$

where ω_0 is the pitch frequency in radians/sample. The aspiration noise, $v[n]$, can be reasonably approximated as white noise with zero mean and unit variance, uncorrelated with any of the harmonics. Assume that the phase θ_k is deterministic, and only the measurement offset n_0 is random. Notice that, here, we are using a_k to denote the total amplitude of the k^{th} harmonic, rather than denoting the amplitude of the vocal tract transfer function at that frequency as we did in lecture.

In terms of a_k, θ_k, and ω_0, what are the autocorrelation and power spectrum, $R_{XX}[m]$ and $P_{XX}(\omega)$, of this signal? What are the cross-correlation and cross-power-spectrum of this signal with the test tone $y[n] = \cos(3\omega_0 n)$? You may assume that $x[n]$ is real-valued and that therefore $a_k = a_{-k}$ and $\theta_k = -\theta_{-k}$.

Solution: Since the noise and the harmonics are all uncorrelated with each other, their autocorrelations and power spectra add, thus

$$R_{XX}[m] = \delta[m] + \sum_{k=-K}^{K} |a_k|^2 e^{j2\pi k\omega_0 m}$$

$$P_{XX}(\omega) = 1 + \sum_{k=-K}^{K} 2\pi |a_k|^2 \delta\left(\omega - k\omega_0\right)$$

$x[n]$ has random time alignment, while $y[n]$ has deterministic time alignment. Averaged over all possible time alignments, the expectation is

$$R_{XY}[m] = 0$$

$$P_{XY}(\omega) = 0$$

Note that this result was a typo in the problem statement; I originally intended to ask about $y[n] = \frac{1}{2}\left(e^{j3\omega_0(nn_0)} + e^{-j3\omega_0(n-n_0)}\right)$, i.e., $y[n]$ having the same time alignment as $x[n]$. In that case we'd get

$$R_{XY}[m] = a_3 \cos\left(3\omega_0 m + \theta_3\right)$$

$$P_{XY}(\omega) = \pi a_3 \left(e^{j\theta_3}\delta(\omega - 3\omega_0) + e^{-j\theta_3}\delta(\omega + 3\omega_0)\right)$$

2. Classic spectrogram studies distinguish two types of spectrograms. A narrowband spectrogram is one computed using a window sufficiently long so that the pitch harmonics are visible as separate peaks in each periodogram, and therefore show up as horizontal bands in the spectrogram image. A wideband spectrogram is one computed using a window sufficiently short so that the spectrum smoothly interpolates between pitch harmonic peaks.

Assume that the spectrum of any window function is triangular in frequency, i.e., piece-wise linear between its peak at $\omega = 0$ and its first null, and that sidelobes don't matter (this approximation is pretty accurate for the Hamming window, and less accurate for other windows). In terms of the pitch period T_0 (in samples), which Hamming windows will produce wideband spectrograms, and which will produce narrowband spectrograms? Which rectangular windows will produce wideband spectrograms, and which will produce narrowband spectrograms? A wideband spectrogram eliminates the horizontal bands marking pitch harmonics, but if computed with a rectangular window (or a short enough Hamming window), it will display information about the pitch period in some other way: How might information about the pitch period be displayed in a wideband spectrogram?

Solution: The spacing between pitch harmonics is $\omega_0 = \frac{2\pi}{T_0}$ radians/sample. The Hamming window mainlobe has a halfwidth of $\frac{4\pi}{N}$, so wideband spectrograms will be produced if $N < 2T_0$, narrowband spectrograms otherwise (a triangular mainlobe would generate a perfect linear interpolation between harmonics if $N = 2T_0$ exactly; a true Hamming window will be not quite so perfect). The rectangular window mainlobe has a halfwidth of $\frac{2\pi}{N}$, so wideband spectrograms will be produced if $N < T_0$, narrowband spectrograms otherwise. If $N < T_0$, then each window is shorter than one pitch period, so the total energy in each windowed frame will vary depending on how it aligns with the glottal closure instants, thus the pitch period is visible in the spectrogram as a periodic fluctuation of $|X_n(\omega)|$ in the time domain (with period T_0).

3. The STFT can be computed as a lowpass filter, or as a bank of bandpass filters. The equations for those two implementations are repeated here, with the frequency $\hat{\omega}$ replaced by the FFT bin frequency $\frac{2\pi k}{N}$:

$$\begin{aligned} X_n\left(\frac{2\pi k}{N}\right) &= w[m] * \left(x[m]e^{-j\frac{2\pi k m}{N}}\right) \\ &= e^{-j\frac{2\pi k n}{N}}\left(h_k[m] * x[m]\right) \end{aligned}$$

For convenience, let's define the signals $x_k[n] = h_k[n] * x[n]$ and $s_k[n] = X_n\left(\frac{2\pi k}{N}\right) = e^{-j\frac{2\pi k n}{N}} x_k[n]$.

(a) Describe or sketch the spectra of each of these two signals, $|\text{DTFT}\{x_k[n]\}|$ and $|\text{DTFT}\{s_k[n]\}|$, as functions of ω. Over what frequencies are they significantly nonzero? Assume a Hamming window.

Solution: $|\text{DTFT}\{s_k[n]\}|$ is significantly nonzero for $-\frac{4\pi}{N} \le \omega \le \frac{4\pi}{N}$. $|\text{DTFT}\{x_k[n]\}|$ is significantly nonzero for $\frac{2\pi k - 4\pi}{N} \le \omega \le \frac{2\pi k + 4\pi}{N}$.

(b) Since $s_k[n]$ is a lowpass signal and $x_k[n]$ is not, it's useful to visualize the latter in terms of the former. Unfortunately, visualization is complicated by their imaginary parts, so let's just focus on the real parts. If $x[n]$ is real then $\Re x_k[n] = \frac{1}{2}\left(x_k[n] + x_{-k}[n]\right)$. Express $\Re x_k[n]$ as a function of $|s_k[n]|$, $\angle s_k[n]$, and $\omega_k = \frac{2\pi k n}{N}$.

Solution:

$$\begin{aligned} x_k[n] &= e^{j\frac{2\pi k n}{N}} s_k[n] = e^{j\frac{2\pi k n}{N} + \angle s_k[n]}|s_k[n]| \\ \Re x_k[n] &= \frac{1}{2}|s_k[n]| \cos\left(\frac{2\pi k n}{N} + \angle s_k[n]\right) \end{aligned}$$

In words: the real part of the bandpass signal equals the real part of the lowpass signal, modulated by $\cos\left(\frac{2\pi k n}{N}\right)$.

(c) Assume that $s_k[n] = X_n(\omega_k)$ is computed using a rectangular window. Write the ISTFT formula that recomputes $x[m]$ from $s_k[n]$, recalling that, for an STFT without downsampling, $\sum_n w[n-m] = N$. From this, find a formula that recomputes $x[m]$ from $x_k[n]$. Comment on how this formula might simplify even further if $s_k[n]$ has been downsampled, in time, to a frame rate of only one frame per N samples.

> **Solution:** The STFT formula is
>
> $$\begin{aligned} w[m-n]x[m] &= \sum_{k=0}^{N-1} s_k[n]e^{j\omega_k m} \\ x[m] &= \frac{\sum_n w[n-m]x[m]}{\sum_n w[n-m]} \\ &= \frac{1}{N}\sum_n \sum_{k=0}^{N-1} s_k[n]e^{j\omega_k m} \\ &= \frac{1}{N}\sum_n \sum_{k=0}^{N-1} x_k[n] \end{aligned}$$
>
> If this has been downsampled to one frame per N samples, then the denominator sum changes from N to 1, and the numerator sum over n has only one nonzero term.