ECE 537 Fundamentals of Speech Processing
Problem Set 2
UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering

Assigned: Monday, 8/29/2022; Due: Monday, 9/5/2022

1. Suppose that, in the style of the vocoder, we have a voiced excitation signal

\[ e[n] = \sum_{m=-\infty}^{\infty} \delta[n - mN_0] \]

We want to filter it through ten bandpass filters \( H_l(\omega) \) (1 \( \leq l \leq 10 \)), then scale each band by amplitude \( A_l \), in order to match the energy of each band in the original speech signal \( s[n] \). To be more precise, we want it to be the case that, for each \( l \),

\[
\sum_{k=0}^{N_0-1} |H_l \left( \frac{2\pi k}{N_0} \right)|^2 |S_k|^2 = \sum_{k=0}^{N_0-1} |A_l|^2 |H_l \left( \frac{2\pi k}{N_0} \right)|^2 |E_k|^2,
\]

where \( S_k \) are the discrete-time Fourier series coefficients of \( s[n] \), and \( E_k \) are the Fourier series coefficients of \( e[n] \).

(a) (1 point) Suppose that the filters are ideal bandpass filters with bandwidth \( \frac{2\pi b}{N_0} \) and center frequency \( \frac{2\pi a}{N_0} \) for some integers \( a \) and \( b \), in other words:

\[
H \left( \frac{2\pi k}{N_0} \right) = \begin{cases} 
1 & a - \frac{b}{2} \leq k < a + \frac{b}{2} \\
1 & N_0 - a - \frac{b}{2} \leq k \leq N_0 - a + \frac{b}{2} \\
0 & \text{otherwise}
\end{cases}
\]

Find positive real numbers \( A_l \) in terms of \( S_k \) so that Eq. 1 is satisfied.
Solution:

\[ |A_i|^2 = \frac{\sum_{k=0}^{N_0-1} |H_i\left(\frac{2\pi k}{N_0}\right)|^2 |S_k|^2}{\sum_{k=0}^{N_0-1} |H_i\left(\frac{2\pi k}{N_0}\right)|^2 |E_k|^2} \]

\[ = \frac{\sum_{k=a-\frac{1}{2}}^{a+\frac{1}{2}-1} |S_k|^2 + \sum_{k=N_0-(a+\frac{1}{2}-1)}^{N_0-(a-\frac{1}{2})} |S_k|^2}{\sum_{k=a-\frac{1}{2}}^{a+\frac{1}{2}-1} \frac{1}{N_0} |S_k|^2 + \sum_{k=N_0-(a+\frac{1}{2}-1)}^{N_0-(a-\frac{1}{2})} \frac{1}{N_0} |S_k|^2} \]

\[ = \frac{N_0^2}{2b} \left( \sum_{k=a-\frac{b}{2}}^{a+\frac{b}{2}-1} |S_k|^2 + \sum_{k=N_0-(a+\frac{b}{2}-1)}^{N_0-(a-\frac{b}{2})} |S_k|^2 \right) \]

where the second line takes advantage of the definition of \( H(\omega) \), and of the formula for the Fourier series of an impulse train. Thus we have

\[ A_i = \sqrt{\frac{N_0^2}{b} \sum_{k=a-\frac{b}{2}}^{a+\frac{b}{2}-1} |S_k|^2} \]

(b) (1 point) In 1940, Fourier analyzers were not very cheap. Instead, most spectral analysis was done by using bandpass filters to compute

\[ s_i[n] = h_i[n] * s[n], \]

and then finding the power of the signal in the time domain,

\[ P_i = \frac{1}{N} \sum_{n=0}^{N-1} s_i[n]^2 \]

Assuming ideal bandpass filters as in part (a), find \( A_i \) in terms of \( P_i, a, \) and/or \( b \). State any assumptions that you need to make about \( N \).

Solution: Parseval's theorem for the discrete-time Fourier series is

\[ \sum_{k=0}^{N_0-1} |X_k|^2 = \frac{1}{N_0} \sum_{n=0}^{N_0-1} x^2[n], \]
which, in our case, translates to

\[ P_l = \frac{1}{N_0} \sum_{n=0}^{N_0-1} s_l^2[n] \]

\[ = \sum_{k=0}^{N_0-1} \left| H_l \left( \frac{2\pi k}{N_0} \right) \right|^2 |S_k|^2 \]

\[ = \sum_{k=a-\frac{\beta}{2} - 1}^{a+\frac{\beta}{2} - 1} |S_k|^2 + \sum_{k=N_0-(a-\frac{\beta}{2})}^{N_0-(a+\frac{\beta}{2} - 1)} |S_k|^2 \]

\[ = 2 \sum_{k=a-\frac{\beta}{2}}^{a+\frac{\beta}{2} - 1} |S_k|^2 \]

Using the answer from part (a), we find that \( A_l \) should therefore be

\[ A_l = \sqrt{\frac{N_0^2 P_l}{2b}} \]

2. Suppose that, in the style of the vocoder, we have an unvoiced excitation signal, \( e[n] \). Assume that \( e[n] \) is zero-mean, unit variance, and uncorrelated, i.e.,

\[ E[e[n]] = 0 \]
\[ E[e^2[n]] = 1 \]
\[ E[e[n]e[n-m]] = 0, \quad m \neq 0 \]

The last two lines in the equation above can be summarized by saying that the autocorrelation is a delta function, \( \rho_{ee}[m] = \delta[m] \), where the autocorrelation is defined to be

\[ \rho_{ee}[m] = E[e[n]e^*[n-m]] \]

where \( e^*[n] \) is the complex conjugate of \( e[n] \), which is part of the definition of autocorrelation, but is not relevant for this problem, because this problem uses only real-valued signals. We want to filter \( e[n] \) through ten bandpass filters \( H_l(\omega) \) \((1 \leq l \leq 10)\), then scale each band by a positive real number \( A_l \), in order to match the energy of each band in the original speech signal \( s[n] \). To be more precise, we want it to be the case that, for each \( l \),

\[ \int_{-\pi}^{\pi} |H_l(\omega)|^2 R_{sss}(\omega)d\omega = A_l^2 \int_{-\pi}^{\pi} |H_l(\omega)|^2 R_{pee}(\omega)d\omega \tag{2} \]

where \( R_{ee}(\omega) \) and \( R_{ss}(\omega) \) are the power spectra of \( e[n] \) and \( s[n] \), respectively.

(a) (1 point) Suppose that the filters are ideal bandpass filters with bandwidth \( \beta \) and center frequency \( \alpha \) radians/sample, in other words:

\[ H(\omega) = \begin{cases} 
1 & \alpha - \frac{\beta}{2} \leq \omega < \alpha + \frac{\beta}{2} \\
1 & -\alpha - \frac{\beta}{2} \leq \omega \leq -\alpha + \frac{\beta}{2} \\
0 & \text{otherwise}
\end{cases} \]

Find positive real numbers \( A_l \) in terms of \( R_{ss}(\omega) \) so that Eq.(2) is satisfied.
Solution:

\[ |A_l|^2 = \frac{\int_{-\pi}^{\pi} \mathcal{H}(\omega)^2 R_{ss}(\omega) d\omega}{\int_{-\pi}^{\pi} \mathcal{H}(\omega)^2 R_{ee}(\omega) d\omega} = \frac{1}{\beta} \int_{\omega=\alpha-\frac{\beta}{2}}^{\alpha+\frac{\beta}{2}} R_{ss}(\omega) d\omega \]

where the second line takes advantage of the definition of \( \mathcal{H}(\omega) \), and of the formula for the Fourier transform of an impulse. Thus we have

\[ A_l = \sqrt{\frac{\pi}{\beta} P_l} \]

(b) (1 point) In 1940, Fourier analyzers were not very cheap. Instead, most spectral analysis was done by using bandpass filters to compute

\[ s_l[n] = h_l[n] * s[n], \]

and then finding the power of the signal in the time domain,

\[ P_l = \frac{1}{N} \sum_{n=0}^{N-1} s_l^2[n] \]

For stochastic analysis, we need to assume that \( N \) is long enough so that

\[ P_l = \mathbb{E}[s_l^2[n]] = R_{s_l s_l}[0], \] (3)

where \( R_{s_l s_l}[m] \) is the autocorrelation of \( s_l[n] \). Assuming ideal bandpass filters as in part (a), find \( A_l \) in terms of \( P_l \), \( \alpha \), and/or \( \beta \).

Solution: The power spectrum is the Fourier transform of autocorrelation, so

\[ R_{s_l s_l}[m] = \frac{1}{2\pi} \int_{-\pi}^{\pi} R_{s_l s_l}(\omega)e^{j\omega m} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\mathcal{H}(\omega)|^2 R_{ss}(\omega) e^{j\omega m} d\omega = \frac{1}{\pi} \int_{\omega=\alpha-\frac{\beta}{2}}^{\alpha+\frac{\beta}{2}} R_{ss}(\omega) e^{j\omega m} d\omega, \]

and therefore

\[ P_l = R_{s_l s_l}[0] = \frac{1}{\pi} \int_{\omega=\alpha-\frac{\beta}{2}}^{\alpha+\frac{\beta}{2}} R_{ss}(\omega) d\omega. \]

and

\[ A_l = \sqrt{\frac{\pi}{\beta} P_l} \]

3. One of the advantages of Licklider’s duplex theory of pitch perception is its potential for detecting periodic signals in noisy listening conditions. Consider the signal

\[ x[n] = v[n] + s[n], \]
where \( v[n] \) is zero-mean, unit-variance white noise, and \( s[n] \) is a periodic speech signal with the form

\[
s[n] = \sum_{k=0}^{N_0-1} S_k e^{j \frac{2\pi kn}{N_0}}
\]

In a realistic measurement scenario, the timing of the input signal is unknown, so we can treat \( n \) as a uniformly distributed random variable. For example,

\[
E[s[n]] = \sum_{k=0}^{N_0-1} S_k E\left[e^{j \frac{2\pi kn}{N_0}}\right] = 0,
\]

where the second line follows by taking expectation over the random variable \( n \).

(a) (1 point) Find the autocorrelation of \( s[n] \). In this case, even though \( s[n] \) is real-valued, the components of the Fourier series are not, so use the formula

\[
R_{ss}[m] = E[s[n]s^*[n-m]],
\]

and then use the assumption that \( n \) is random.

**Solution:**

\[
R_{ss}[m] = E[x[n][n-m]] = E[v[n]v[n-m]] + 2E[v[n]s[n-m]] + E[s[n]s[n-m]]
\]

\[
= \delta[m] + E\left[\sum_{k=0}^{N_0-1} X_k e^{j \frac{2\pi kn}{N_0}}\right]\left[\sum_{l=0}^{N_0-1} X^*_l e^{-j \frac{2\pi l(n-m)}{N_0}}\right]
\]

\[
= \delta[m] + E\left[\sum_{k=0}^{N_0-1} \sum_{l=0}^{N_0-1} X_k X^*_l e^{j \frac{2\pi (k-l)(n-m)}{N_0}}\right]
\]

\[
= \delta[m] + \sum_{l=0}^{N_0-1} |X_l|^2 e^{j \frac{2\pi ml}{N_0}}
\]

(b) (1 point) In the preceding section, we saw that the noise power was \( R_{vv}[0] = 1 \), while the power of the periodic part of the signal is \( R_{ss}[0] = R_{ss}[N_0] \). Licklider’s model reduces the noise power, without reducing the signal power. Suppose, for example, that we bandpass filter \( x[n] \) through ideal bandpass filters with some bandwidth less than the pitch period:

\[
\beta = \frac{2\pi}{N_0}, \quad 0 < \gamma < 1.
\]

Now suppose that we compute the autocorrelations in every channel, and then we add together only the channels that have significant energy at a delay of \( m = N_0 \). What is the resulting noise power, \( R_{vv}[0] \)?

**Solution:** Using Parseval’s theorem, the power of each bandpass-filtered noise signal \( v_l[n] \) is

\[
\frac{\beta}{2\pi} = \frac{\gamma}{N_0}.
\]

When we add together \( N_0 \) of these, we get a total power of

\[
R_{vvv}[0] = \gamma
\]