UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering
Instructor: Mark Hasegawa-Johnson

ECE 537 Speech Processing

Problem Set 5
Fall 2009

Issued: Wed Sep. 30, 2009 **Due:** Wed Oct. 7, 2009

Reading for problem set 5: Flanagan, Allen & Hasegawa-Johnson 5.1–5.3

For this problem set, you may want to download the voicebox toolbox for matlab from

`http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html`

You can use the voicebox toolbox to check your answers in problem 5.2, but you should also be able to implement any of the requested computations on your own, without help of the toolbox.

Problem 5.1

A reasonable model for the vocal tract impulse response $h[n]$ during vowel production is

$$h[n] = \left(e^{-\sigma_1 n}\sin(\omega_1 n)u[n]\right) * \left(e^{-\sigma_2 n}\sin(\omega_2 n)u[n]\right) * \left(e^{-\sigma_3 n}\sin(\omega_3 n)u[n]\right) \quad (1)$$

The formant frequencies are $\omega_1, \omega_2, \omega_3$ radians/sample, the bandwidths are $2\sigma_1, 2\sigma_2, 2\sigma_3$ radians/sample, $*$ denotes convolution, and $u[n]$ is the unit step function.

(a) Find the vocal tract frequency response, $H(e^{j\omega})$, by taking the Fourier transform of Eq. 1.

(b) The speech signal $s[n]$ is a periodic repetition, with period T_0, of the impulse response, i.e.,

$$s[n] = \sum_{k=-\infty}^{\infty} h[n - kT_0] = p[n] * h[n]$$

where $p[n]$ is a periodic pulse train. Find $P(e^{j\omega})$, then specify $S(e^{j\omega})$ in terms of $H(e^{j\omega})$ and $P(e^{j\omega})$.

(c) The first $s_0[n] = w[n]s[n]$, where $w[n]$ is a rectangular window of length $N = 2.5T_0$ samples. What is $S_0(e^{j\omega})$? Find the exact solution (in terms of $H(e^{j\omega})$, T_0, and $W_R(e^{j\omega})$ the transform of a rectangular window), then sketch $|S(e^{j\omega})|$ in the vicinity of the first formant, showing important features.

(d) Repeat part (c), but now using a Hamming window (of length $N = 2.5T_0$) instead of a rectangular window. Write your answer in terms of $W_H(e^{j\omega})$, the transform of a Hamming window, then sketch $|S_0(e^{j\omega})|$.

(e) Suppose that your result in part (c) can be usefullly approximated as follows:

$$S_0(e^{j\omega}) \approx H(e^{j\omega}) \left(1 + G \sum_{k=-\infty}^{\infty} \delta(\omega - \frac{2\pi k}{T_0}\right)$$

where G is some constant whose value we will consider irrelevant. The complex cepstra are computed as follows:

$$\tilde{s}[n] = \mathcal{F}^{-1}\left(\ln S_0(e^{j\omega})\right)$$

then windowed to produce:

$$\tilde{y}[n] = w[n]\tilde{s}[n]$$

where $w[n] = u[n + (T_0 - 1)/2] - u[n - (T_0 + 1)/2]$ is a length-T_0 rectangular window centered at $n = 0$. This window is only possible if T_0 is an odd integer, so let's assume that T_0 is an odd integer.

Find $\tilde{Y}(e^{j\omega})$, the Fourier transform of $\tilde{y}[n]$. Assume that $\ln H(e^{j\omega})$ is a low-quefrency function, so that the liftering operation passes it unchanged. Assume that $\sin(\omega N/2)/\sin(\omega/2) \approx N\text{sinc}(\omega N/2)$ for small enough values of ω.

(f) Consider now the same speech STFT as in part (e), but instead of using a complex cepstrum, you are going to use a magnitude cepstrum, the inverse DCT of $\ln|S_0(e^{j\omega})|$:

$$c[n] = \frac{1}{\pi}\int_0^{\pi} \ln|S_0(e^{j\omega})| \cos(\omega n) d\omega$$

The magnitude cepstrum is then liftered to produce:

$$f[n] = \begin{cases} 0.5c[n] & n = 0 \\ c[n] & 0 \le n \le \frac{T_0 - 1}{2} \\ 0 & \text{else} \end{cases}$$

and then DCT'd to produce:

$$F(e^{j\omega}) = 2\sum_{n=-\infty}^{\infty} f[n]\cos(\omega n)$$

Prove that $F(e^{j\omega})$ is the real part of $\tilde{Y}(e^{j\omega})$.

Problem 5.2

A newborn infant has a high larynx (near the bend in the vocal tract), therefore he or she can't easily manipulate the frequencies of multiple formants. His or her formant frequencies are pretty high, anyway, because of the short vocal tract. Suppose that a particular infant produces a vowel sound whose LPC analysis filter is

$$A(z) = (1 - 0.8z^{-1})(1 - 0.95e^{j\pi/3}z^{-1})(1 - 0.95e^{-j\pi/3}z^{-1})$$

Assume $F_s = 8\text{kHz}$.

(a) Draw a pole-zero plot of the filter $H(z) = 1/A(z)$.

(b) What is the first formant frequency, in Hertz?

(c) What is the first formant bandwidth, in Hertz?

(d) Sketch a reflection-line model of the LPC synthesis filter. Specify values of the reflection coefficients.

(e) Find the log area ratios.

(f) Sketch a vocal tract area function that would produce this $H(z) = 1/A(z)$. Specify the terminating glottal impedance, the terminating radiation impedance, and all of the tube areas and tube lengths.

(g) Find $Q(z)$ and $P(z)$, the polynomials whose roots are the line spectral frequencies.

(h) Use the identities $\cos(2\omega) = 2\cos^2\omega - 1$ and $\sin(2\omega) = 2\cos(\omega)\sin(\omega)$ to express $Q(z)$ and $P(z)$ as polynomials $Q(x, y)$ and $P(y)$, where $x = \sin(\omega)$ and $y = \cos(\omega)$. Solve to find the LSFs.

Problem 5.3

Record your own voice, saying the vowel /ɑ/. Downsample to 8kHz sampling rate, so that LPC knows to which frequencies it should pay attention. Compute a tenth-order LPC model of this vowel (you may get better results if you chop off the silences at either end, first). Make a pole-zero plot, and identify the frequencies and bandwidths (in Hertz) of the first two formants.

Compute the log area ratios (you may use voicebox for this). From the log area ratios, compute the vocal tract area function, using the assumption that $Z_R = 0$, and assuming that the area of the lips is about 5cm^2—note that these assumptions are different from the assumptions made by the voicebox toolbox `lpcrf2aa.m`, so you should not use it for this problem. Plot the resulting area function, $A(x)$, where x is measured in centimeters from the glottis (remember to compute the length of each tube section. Remember: distance from glottis, not distance from lips).

Problem 5.4

Record your own voice, saying any utterance of less than one second duration (up to four or five syllables).

Plot the wideband spectrogram. Remember that a spectrogram is a picture in which the brightness or color of each spot is proportional to $20\log_{10}|S_t(e^{j2\pi f/F_s})|$, for $0 \le f \le F_s/2$; the `specgram` function in matlab will do this, as will the `spgrambw` function in the voicebox toolbox. Use Hamming windows of at least 2.5 pitch periods in length (recommendation: 25ms). The step between Hamming windows should be adjusted so that you get a nice number of columns in your spectrogram, e.g., a 2ms step size will give you an image 500 pixels wide. Make sure that your abscissa is labeled in seconds, and your ordinate in Hertz.

Now compute an LPC polynomial once every 10ms or so, using 25ms Hamming windows. Compute the formant frequencies (imaginary parts of the roots of the LPC polynomial, scaled to Hertz) as functions of time, and overlay them on the spectrogram using magenta 'x's, e.g.,

```
hold on;
plot(T,F,'mx');
```

where `T` is a vector containing time stamps in seconds, and `F` is a matrix containing, in each column, formant frequencies in Hertz at the corresponding time.

Finally, compute the line spectral frequencies, using the same LPC polynomials. Scale the line spectral frequencies to Hertz. Plot the q_n roots as cyan circles, and the p_n roots as yellow triangles, e.g.,

```
hold on;
plot(T,Q,'co',T,P,'y^');
```

If these color schemes don't work on your display or on your printer, find some other color scheme that works. Try to get the spectrogram, formants, and line spectral frequencies all on the same plot.