EE6640 Speech Signal Processing
Homework #1
Out: 2016 / 10 / 03
Due: 2016 / 10 / 30

Problem 1

A speech signal is sampled at a rate of 10,000 samples/sec (i.e., Fs=10,000). A Hamming window of length ***L*** samples is used to compute the STFT of the speech signal. The STFT is sampled in time with period ***R***, and in frequency at ***N=1024***frequencies.

(a) It can be shown that the main lobe of the Hamming window has a symmetric full width of approximately $\boldsymbol{8\pi/L}$.How should ***L*** be chosen if we want the full width of the main lobe to correspond to approximately 200 Hz analog frequency?

(b) How should ***R*** be chosen if we wish to compute the STFT every 10 msec.

(c) What is the spacing (in Hz) between sample points in the frequency domain?

Problem 2

The short-time autocorrelation function of the signal x(m) is defined as follow:

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m+k)w(n-k-m)$$

(a) Show that $R_n(k) = R_n(-k)$, i.e. that it is an even function of k

(b) Show that $R_n(k)$ can be expressed as

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)x(m-k)h_k(n-m)$$

where,

$$h_k(n-m) = w(n)w(n+k)$$

(c) Suppose that

$$w(n) = \begin{cases} na^n, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

Find the impulse response $h_k(n)$, for computing the $\mathrm{k^{th}}$ lag.

(d) If we define the short-time power density, $S_n(e^{jw})$, of this signal $x(m)$ in terms of its STFT (short time Fourier transform), $X_n(e^{jw})$, as the following

$$S_n(e^{jw}) = |X_n(e^{jw})|^2$$

Now, if

$$X_n(e^{jw}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-jwm}$$

Show that $S_n(e^{jw})$ is the Fourier transform of $R_n(k)$.

Problem 3

Consider the sequence

$$x[n] = \delta[n] + \alpha\delta[n - N_p]$$

(a) Find the complex cepstrum of $x[n]$ and sketch the result

(b) Sketch the real cepstrum, $c[n]$, for $x[n]$

(c) Suppose that an approximation to the cepstrum, $\tilde{c}[n]$, is computed as follows:

$$X_p[k] = \sum_{n=0}^{N-1} x[n]e^{\frac{-j2\pi}{N}kn}, \quad 0 \le k \le N-1$$

$$\tilde{c}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \log|X_p[k]|e^{\frac{j2\pi}{N}kn}, \quad 0 \le n \le N-1$$

Sketch $\tilde{c}[n]$ for $0 \le n \le N-1$ for the case $N_p = N/6$. What if N is not divisible by N_p?

(d) If the largest impulse in the cepstrum approximation, $\tilde{c}[n]$ is used to detect N_p, how large must N be in order to avoid confusion?

Problem 4

(a) Give a block diagram for a general MFCC feature extraction procedure and explain each step.

(b) (Matlab) load the file **sound.mat**. **x**is the voice signal, and **fs**is samplingfrequency.

(c) (Matlab) Follow the steps according to your block diagram in (a), and calculate the first 15 mel-frequency cepstral coefficients.

(d) Please explain the difference between LPCC and MFCC.

(e) Discuss the relative merits and demerits of rectangular and Hamming window and, as applied to speech processing. Why is the Hamming window often preferred to the rectangular window?

Problem 5

(a) (Matlab) Record a short sentence of your voice by using **audiorecord**.

(b) Estimate the window size to plot spectrogram.

(c) (Matlab) Use the window size you estimate in (b) to plot **spectrogram**.

(d) (Matlab) Adjust the window size. Which size do you think is suitable for spectrogram.

Problem 6

In implementing STFT representations, we employ sampling in both the time and frequency dimensions. In this problem, we investigate the effects of both types of sampling. Consider a sequence x[n] with DTFT

$$X(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x[n]e^{-j\omega m}$$

(a) If the periodic function $X(e^{j\omega})$ is sampled at frequencies $\omega_k = 2\pi k/N, \quad k = 0, 1, \ldots, N-1$, we obtain

$$\tilde{X}[k] = \sum_{m=-\infty}^{\infty} x[m]e^{-j\frac{2\pi}{N}km}$$

These samples can be thought of as the DFT of the sequence $\tilde{x}[n]$ given by

$$\tilde{x}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{X}[k]e^{j\frac{2\pi}{N}kn}.$$

Show that ,

$$\tilde{x}[n] = \sum_{r=-\infty}^{\infty} x[n+rN]$$

(b) What are the conditions on x[n] so that no aliasing distortion occurs in the time domain when $X(e^{j\omega})$ is sampled?

(c) Now consider "sampling" the sequence x[n]; i.e., let us form the new sequence

$$y[n] = x[nM]$$

consisting of every Mth sample of x[n]. Show that the Fourier transform of y[n] is

$$Y(e^{j\omega}) = \frac{1}{M}\sum_{k=0}^{M-1} X(e^{j(\omega-2\pi k)/M})$$

In proving this result, you may wish to begin by considering the sequence

$$v[n] = x[n]p[n],$$

where

$$p[n] = \sum_{r=-\infty}^{\infty} \delta[n+rM].$$

Then note that $y[n] = v[nM] = x[nM]$

(d) What are the conditions on $X(e^{j\omega})$ so that no aliasing distortion in the frequency domain occurs when $x[n]$ is sampled?

Problem 7

In deriving the lattice formulation, the i^{th} order prediction error filter was defined as

$$A^{(i)}(z) = 1 - \sum_{k=1}^{i} \alpha_k^{(i)} z^{-k}.$$

The predictor coefficients satisfy the following relations:

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}, \quad 1 \le j \le i-1$$
$$\alpha_i^{(i)} = k_i.$$

Using the relations, derive the following recursive from of the predictor error filter:

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}).$$

Problem 8

Consider an all-pole model of the vocal tract transfer function of the form

$$V(z) = \frac{1}{\prod_{k=1}^{q}(1-c_k z^{-1})(1-c_k^* z^{-1})},$$

where

$$c_k = r_k e^{j\theta_k}.$$

Show that the corresponding cepstrum is

$$\hat{v}(n) = 2\sum_{k=1}^{q} \frac{(r_k)^n}{n} \cos(\theta_k n).$$