EE627 - Speech Signal Processing

Assignment # 3

1. Show that for a frame interval of L samples (i.e., time decimation of L samples), $2L$ samples of $|X(nL, k)|$ over $[0, \pi]$ in frequency are required to uniquely recover a sequence $x[n]$. Assume the window length $N_w \geq L/2$. For simplicity, assume N_w and L are even.

2. Consider modifying the STFT to obtain

$$Y(n, \omega) = X(n, k\omega)H(\omega)$$

where the modifying function is given by

$$H(\omega) = e^{jn_0\omega}$$

i.e., a linear-phase modification. Suppose a sequence $y[n]$ is computed by the filter bank summation (FBS) synthesis method:

$$y[n] = \left[\frac{1}{NW(0)}\right] \sum_{k=0}^{N-1} Y(n, k) e^{j\frac{2\pi}{N}kn}$$

Derive an expression for $y[n]$ in terms of original sequence $x[n]$ and the window $\omega[n]$. Consider two different cases: (1) The length of $\omega[n]$ is less than N, and (2) The length of $\omega[n]$ greater than or equal to N. Give $y[n]$ for each case.

3. (MATLAB) In this problem, use the voiced speech waveform *speech2_10k* in the workspace *ex7M1.mat* to design the time-scale modification systems based on OLA and LSE synthesis methods. The speech was sampled at 10000 samples/s.

   (a) Write a MATLAB function to compute the STFT of the sequence *speech2_10k* using a 30 ms triangular analysis window, created using MATLAB function *triang.m*, at a 15 ms frame interval. Then reconstruct the original waveform from the STFT using OLA approach for synthesis. Ignore tapering end effects of the first and last frames.

   (b) Design in MATLAB an OLA-based synthesis method to time-scale expand the speech signal by a factor of two by repeating every frame.

   (c) Repeat parts (a) and (b) using the LSE-based synthesis approach. Compare the time-scaled signals from each approach and comment on the differences.

   (d) Repeat parts (b) and (c) using "pitch-synchronized" OLA and LSE-based synthesis approaches. This will require that you estimate a consistent time-instant (e.g., the waveform peak or the glottal pulse time) within a glottal cycle in order to synchronize consecutive frames. How does this approach improve your synthesis from parts (b) and (c)?

   (e) Repeat parts (a)-(d) with a speech utterance from your own voice recording. For part (d), consider manually marking voiced and unvoiced regions and invoke pitch synchrony during the voiced regions.

4. (MATLAB) In this MATLAB exercise, use the workspace *ex7M2.mat*, as well as the function *uniform_bank.m* to explore the conditions for the recovery of a sequence using the filter bank summation (FBS) method.

   (a) There are four different analysis windows (or " analysis filters") in the workspace *ex7M2.mat filter1*, *filter2*, *filter3*, and *filter4*. Using MATLAB command *subplot*, plot all the filters. Note that *filter4* is simply a shifter version of *filter3*.

   (b) The function *uniform_bank.m* creates a filter bank $h_k[n]$. The output of the function is a $2-$D array of modulated filter impulse response, but without the demodulation term $e^{-j\frac{2\pi}{N}kn}$ (as shown in Figure 7.5 in book Discrete Time Speech Signal Processing by Thomas F. Quatieri). Also, the impulse responses are real because the complex conjugate response pairs have been combined. Note that the first and last filter do not correspond to complex conjugate pairs. Explain why.

   (c) Do a *help* command on *uniform_bank.m* and run the function with a $250-$Hz spacing between filters, the analysis filter *filter1*, and a plot factor of 100 (scaling the output for a good display). The function plots the frequency response magnitude of the filters using a 1024-point DFT. Assume the time sampling rate is 10000 samples/s so that the 512th frequency bin corresponds to 5000 Hz. Having the impulse responses $h_k[n]$ of your filter bank, write a MATLAB function to filter the unit sample *impulse* in *ex7M2.mat* with your filter bank, and then combine all impulse responses to create the composite (summed) filter bank impulse response. Explain the observed composite impulse response using the FBS constraint. Finally, plot the impulse response to the second and the fifteenth bandpass filter and explain the difference in time structure of the two signals [recall that the filter bank is missing the STFT demodulation term $e^{-j\frac{2\pi}{N}kn}$ (Figure 7.5 in book Discrete Time Speech Signal Processing by Thomas F. Quatieri)]

   (d) Repeat part (c) with analysis filter *filter2*. Using the FBS constraint, explain the reason for the deviation in the composite response from an impulse.

   (e) Repeat part (c) with *filter3* (using a plot factor of 2). Superimpose the composite impulse response on *filter3* (plotting the first 400 samples) and again explain your observation using the FBS constraint.

   (f) Repeat part (c) with *filter4*, which is *filter3* shifted by fifty samples to the right. Superimpose the composite impulse response on *filter4* (plotting the first 450 samples) and again explain your observation using the FBS constraint. Also comment on the difficulty in creating an impulsive composite response by simply shifting the analysis filter.