ECE 797: Speech and Audio Processing

Assignment #2: All-pole analysis and synthesis of speech

The Due date: Report and MATLAB code to be submitted before class on Thursday March 18.

The report: Should include plots and descriptions of results and answers to specific questions. A hand-written copy of the mathematical derivations for Part A is acceptable.

The code: The MATLAB M-files generating the results in your report should be submitted.

The rules: Each student should work on their own code and report. If you get stuck, you can ask for advice from another student, but please do not show each other your mathematical derivations for Part A. If you get really stuck, you can ask me.

The Assignment:

Part A: Frequency domain interpretation of linear prediction analysis.

On page 205–207 of Quatieri, a frequency domain of linear prediction analysis is developed. Some steps in the derivation are skipped. Your job is to complete those steps.

1. Show that $\int_{-\pi}^{\pi} \log |A(\omega)|^2 \frac{d\omega}{2\pi} = 0$, and consequently $\int_{-\pi}^{\pi} \log \left| \frac{A}{A(\omega)} \right|^2 \frac{d\omega}{2\pi} = \log A^2$.

2. Show that minimization of the error function I with respect to the parameters α_k of $A(\omega)$ is equivalent to minimization of the average squared prediction-error filter output $\int_{-\pi}^{\pi} |E(\omega)|^2 d\omega = E_n$, i.e., it gives the same solution as the autocorrelation method.

3. Show that minimization of the error function I with respect to the squared gain A^2 gives the solution $A = \sqrt{E_n}$, the same solution as the autocorrelation method.

Part B: Implementation of linear prediction analysis in MATLAB.

In this part, use the supplied .wav file `assgn2.wav`, which contains 250 ms of a voiced speech segment sampled at 10 kHz.

1. Apply a 25 ms Hamming window to the speech segment. Compute the autocorrelation of the resulting windowed signal. Plot both the signal and the autocorrelation function.

2. Assume that four resonances represent the signal and model the vocal tract with 8 poles. Set up the autocorrelation matrix R_n, using your result from above.

3. Solve directly for the linear predictor coefficients using matrix inversion.

4. Plot the log-magnitude of the resulting frequency response:

$$H(\omega) = \frac{A}{1 - \sum_{k=1}^{p} \alpha_k e^{-j\omega k}},$$

    where the gain A is given by Eq. (5.30) in Quatieri. Compare your result with the log-magnitude of the Fourier transform of the windowed signal. What similarities and differences do you observe?

5. Using your estimates of the predictor coefficients from above, compute the prediction error sequence associated with this speech segment and plot it. From the prediction error sequence, what conclusions can you draw about the model (i.e., all-pole/impulse-train-driven) and estimation accuracy?

6. Using your estimates of the predictor coefficients, compute an estimate of the vocal tract impulse response.

7. Using the prediction error sequence you computed above, estimate the average pitch period of the voiced speech segment.

8. Using your results from 6 & 7, synthesize a 250-ms estimate of the unwindowed speech segment from `assgn2.wav`, assuming that the excitation source is a perfectly period train of ideal impulses. How does your waveform estimate differ from the original? Consider the minimum-phase nature of the impulse response estimate.

9. Using the MATLAB number generator function `randn()`, synthesize the "whispered" counterpart to your voiced synthesis of 8 above. Using the MATLAB function `soundsc()`, listen to your two estimates and compare to the original.

Dr. Ian C. Bruce   March 4, 2004