EECS225d Audio Signal Processing in Humans and Machines	Spring 1995
Lecture 25: March 17, 1995	
Lecturer: Professor Ben Gold	*Scribe: Ted Lai*

25.1 Papers of Interest in the Course Reader:

Today – Kang and Coulter

Mon. (3/20) – Klatt (look at list of tape)

Wed. (3/22) – Moorer (review) Chowning

Fri. (3/24) – Dudley, Tarnoczy (review)

25.2 Overview

Today's lecture objectives are based on a 600 bits/sec LPC-derived Vocoder:

(1) How to track formants?

(2) Vector Quantization

(3) DRT (Diagnostic Rhyme Test) – how good is a vocoder or speech system?

25.3 Viewgraphs

(1) A 600-bps voice digitizer:

This viewgraph shows a complete block diagram of the Transmitter and Receiver of the digitizer. The block "Conversion from Linear Predictive Coefficients to Formant Frequencies" is formant frequencies extraction based on LPC parameters. It reduces bandwidth and performs vector quantization. We can do this because human ear and brain don't neet to listen to all possible spectral patterns.

(2) Root locus of the vocal-tract filter as the last filter coefficient approaches unity:

$$z^n - a_1 z^{n-1} - a_2 z^{n-2} - \ldots - a_n = \Pi_{i=1}^{n}(z - z_i)$$

where z_i are poles that move towards unit circle as $a_n \rightarrow 1$, and a_n is the product of all the poles.

(3) Effect of the last predictive coefficient on the frequency response of the vocal-tract filter:

This viewgraph shows that the second formant becomes visible as the last predictive coefficient, k_{10}, approaches to 1.

(4) Table of parameter coding:

This viewgraph shows how people apportion their bits. For example, while Frame rate of a typical 2400-Bit-Per-Second Linear Predictive Encoder (LPE) is $44.444Hz$ compared to $40Hz$ for a 600-Bit-Per-Second Voice Digitizer (VD), LPE's Vocal-tract filter parameters require 40 bits/frame, but the same parameters require only 7 bits/frame for VD. This difference is made up by interpolation.

So far we have covered (1) formant tracking, (2) vector quantization, (3) interpolation, (4) DRT (Diagnostic Rhyme Test), an objective test for vocoders.

(5) Summary of DRT Score at 600 bps and, For Comparison, at 2400 bps:

This viewgraph shows the result of perception test on various features.

We heard a tape that demonstrates a fair degree of intelligibility for the 600-bps Voice Digitizer.

(6) Fig. 1. Speech synthesis model with short and long delay predictors; Fig. 2. Block diagram illustrating the procedure for selecting the optimum innovation sequence:

This viewgraph is taken from Shroeder and Atal, "Code-Excited Linear Prediction (CELP): High-Quality Speech in Low Bit Rate," ICASSP '85.

In Fig. 2., the analyzer uses each innovation sequence (excitation) to produce an artificial speech, which is then compared with the original speech signal. After repeating this procedure for all innovation sequences in the data bank, the optimum innovation sequence is chosen by using least square error method. The address of the optimum innovation sequence is transmitted to the receiver, which looks up and uses that particular innovation sequence to produce speech, as shown in Fig. 1.

(7) Various innovation sequences in a 5-*msec* frame and their corresponding amplitude and phase spectra:

Innovation sequences (excitations) are generated by Gaussian random process.

We heard a tape on a commercial variable rate LPC vocoder.

Next week: The same system for music requires high bps rate, and music cannot be modeled with just a single excitation source.