E6820 Speech & Audio Processing & Recognition **Homework 8**

Assigned: Thursday 2001-03-22 Due: Tuesday 2001-03-27

Dan Ellis <dpwe@ee.columbia.edu>

Background reading:

Read chapters 22 through 24 in Gold & Morgan.

Reading assignment:

"A comparison of signal processing front ends for automatic word recognition", C. Jankowski, V. Hoang-Doan & R. Lippmann, IEEE Trans. Speech & Audio Proc. 3(4), 286-293, July 1995. This covers contemporary issues in feature design for speech recognition, and is typical of the way this work is presented. Add a summary to your web page, including any questions or aspects you found particularly interesting.

http://www.ee.columbia.edu/~dpwe/e6820/papers/JanVL95-asrfecomp.pdf

Practical assignment:

This week we look at speech compression based on LPC representations.

First we examine the LPC filter coefficients. Analyze a speech segment using the new version of `lpcfit` (Beware! The input and output arguments have been 'improved', so you'll need to follow the usage shown in this week's Matlab diary). Try quantizing the `a` matrix of filter coefficients between -3 and 3 using `quant` (as shown in the diary). How many steps do you need for good quality resynthesis (via the new `lpcsynth`, using the original excitation)? What happens if you use too few steps (quantize too coarsely)? Now try converting the a matrix to LSPs using `lpca2lsp`, quantizing in that domain (between 0 and 0.5), then converting back to filter coefficients with `lpclsp2a` to feed into `lpcsynth`. How does the bitrate of an acceptable LSP representation compare to the direct filter coefficients?

Next, consider the excitation. You can quantize the original residual `e` with `quant`; experiment to find the best range and number of levels. Then try Multi-Pulse Excitation representation via `lpcMPEenc` and `lpcMPEdec` (use `help lpcMPEenc` etc. to get the usage messages). Try using between 16 and 64 pulses per frame. How does the quality vary? What is the bitrate required for transmission, assuming you quantize the pulse magnitudes with the same resolution as you used for the original residual? (The times are already quantized as sample indexes).

Finally try the buzz-hiss imitation of the residual implemented by `lpcBHenc` and `lpcBHdec`. How does the sound quality compare with other encodings for the residual? What about the bitrate?

Problems:

1. One aspect of wide-band coding not mentioned in class is using the particular interdependence of stereo (2-channel) signals. How might compression gains be obtained for stereo signals, distinguishing between redundancy and irrelevancy removal?
2. Question 33.5 from Gold & Morgan.