In this project you will build a residual-excited LPC (RELP) vocoder based on vector quantization of the vocal tract parameters. You will perform subjective evaluation by listening to the synthesized speech and quantitative evaluation by measuring the distortion.

**The Residual-Excited Vocoder**

A residual-excited LPC vocoder uses a coded representation of the LPC error signal as the excitation function in an LPC synthesis system. Since you will use this project to examine the effect of vector quantization on the vocal tract parameters, you will use the error signal as computed rather than encoding it to achieve a lower data rate.

For each frame of length 30 ms (480 samples at a sampling rate of 16 kHz) in the input speech, first obtain the LPC coefficients ($b_i$'s) using autocorrelation method LPC on the Hamming windowed frame. For the excitation, the vocoder will compute the error signal for that frame:

$$e_n = s_n + \sum_{k=1}^{p} b_k s_{n-k} \quad \text{for } 0 \leq n \leq N-1$$

where $s_n$ is the original (unwindowed) signal and where a sample outside frame $i$, $i > 0$, (i.e., $s_j$ for $j < 0$) is assumed to be sample $s_{N-j}$ from frame $i-1$. In computing $e_n$ for frame 0, assume that $s_j = 0$ for $j < 0$.

Note that, although the LPC analysis is done using Hamming windowed frames, in computing the error you are assuming that the resulting $H(z)$ is the LPC filter for the unwindowed frame. This allows the use of non-overlapping windows.

(Alternatively, you may use overlapping windows (for example, 30 ms windows with 15 ms overlap), compute $e_n$ using the Hamming windowed input signal with out-of-frame samples assumed to be 0, use the LPC equations to generate the output samples for the overlapping frames, then sum the corresponding points from the overlapped frames in order to generate the final output speech samples).

**Vector Quantization**

The LPC filters used to synthesize speech in the vocoder will use coefficients derived from various vector quantizer codebooks. For each analysis frame of length 30 ms, you will convert $p=20$ LPC coefficients to $p=20$ cepstral coefficients (not including $c_0$, which represents energy):

$$c_k = -b_k - \sum_{i=1}^{k-1} j c_j b_{k-i}$$

(For the case when $i=1$ (so the sum goes from $i=1$ to 0), the sum is empty, so $c[1]=-b[1]$.)

Use a Euclidean distance on the cepstral coefficients to select the appropriate codeword from one of the given VQ codebooks. Using table lookup, convert the codeword’s cepstral coefficients to LPC coefficients. Finally, use the LPC coefficients and the LPC error signal (the residual excitation) to generate the synthesized speech signal.

**Summary of the Procedure:**

Given a VQ codebook and an input sentence:

1. Perform a 20-pole LPC analysis on the Hamming windowed frame.
2. Create the error (excitation) signal for the frame.
3. Convert the 20 LPC inverse filter coefficients to 20 cepstral coefficients $c_1, c_2, \ldots, c_{20}$ and use a Euclidean distortion measure on the cepstral coefficients to select the VQ codeword.
that best represents the frame. Use table lookup to convert the codeword’s cepstral coefficients to LPC coefficients.

4. Generate the output samples for the frame using the LPC coefficients corresponding to the VQ codeword (step 3) and the excitation signal (step 2).

Summary of Data and Experiments
Details about the data are given in the Project 3 Specifications sheet. There are two input (test) sentences, one from a male speaker and one from a female speaker. There are 40 VQ codebooks: 10 VQ codebooks of ranging in size from 2 to 1024 vectors, created by training on four different speaker sets:
- trained on the same male individual as the test sentence (10 codebooks)
- trained on the same female individual as the test sentence (10 codebooks)
- trained on a set of male speakers (10 codebooks)
- trained on a set of female speakers (10 codebooks)

Evaluation
Test the given input sentences using each of the provided VQ codebooks. Your evaluation should consist of two parts:

1. A quantitative evaluation, in which you analyze the average sentence distortion as a function of the codebook type (i.e., number and type of speakers from whom the codebook was generated) and codebook size. Average sentence distortion is defined as the sum of the distortion per frame over all the frames of the input sentence, divided by the number of frames in the input sentence. The distortion measure for each frame should be the Euclidean distance computed between the LPC-derived cepstral coefficients for the input speech and the LPC-derived cepstral coefficients in the codebook. You should report on the average sentence distortions for each of your experiments and discuss the results.

2. A subjective evaluation, in which you discuss the intelligibility and quality of the synthesized speech. Use the quantitative results to decide which sentence-codebook pairs to evaluate subjectively. Information on how to turn in selected output speech files will be posted on the EE649 Projects web page (http://shay.ecn.purdue.edu/~ee649/projects).

Finally, you should discuss the relation between your observations from the quantitative and subjective evaluations.

Suggestions
1. In order to verify your synthesis module, first test your residual-excited LPC vocoder using the LPC coefficients and error signal from analysis of the input speech, without going through the vector quantizer.

2. Test your conversion from LPC to cepstral coefficients by comparing the cepstral coefficients to those you obtain by doing a standard cepstral analysis.

3. Listen to e⁻ⁿ.

Individual or team projects
You may work with a partner or team of up to three people on this project. If you work on a team, all team members will be responsible for understanding the entire project. If you work individually, you may choose to do your evaluation on either the male speech (the input sentence from the male speaker with all 20 of the male codebooks) or the female speech (the input sentence from the female speaker with all 20 of the female codebooks). If you work with a partner or team, your group must evaluate both input sentences (male and female) with all of the codebooks (i.e., evaluate each sentence with all 40 codebooks, for a total of 80 evaluations).