Project 1: Cepstral Analysis

Professor Deepa Kundur

Objectives of this project

• To expose you to the concepts of cepstral analysis and homomorphic deconvolution.

• To give you the opportunity to be creative and play around with audio signal processing applications.

Please note that this is an *individual* project and each person can talk to others, but must ultimately conduct their own simulations and write their own report.

Background

Deconvolution

As discussed in the lectures, *deconvolution* is a process of undoing the effects of convolution. Ideally, if $g(n) = f(n) * h(n)$ where $f(n)$ is a signal of interest such as a sound signal and its convolution with $h(n)$ represents system distortion from, say, the recording process, then it is necessary to restore $f(n)$ from $g(n)$. This can be conducted by inverting the convolution process with $h(n)$ through a number of techniques such as inverse filtering.

However, in the presence of noise, an inverse filtering approach has practical disadvantages of noise amplification. Another strategy to deconvolution involves directly separating the components $f(n)$ and $h(n)$ from $g(n)$. One approach, which is the focus of this project, makes use of *cepstrum* domain processing.

The Cepstrum Domain

The cepstrum is a common transform used to gain information from a person's speech signal. It can be used to separate the excitation signal (which contains the words and the pitch) and the transfer function (which contains the voice quality). The word cepstrum has the first syllable "ceps" which is "spec" with the letters rearranged in a different order. Thus, the words *ceps*trum and *spec*trum are related through an interesting naming convention that will use again. Figure 1 provides a block diagram showing how a signal would be converted to the cepstral domain.

$$x(n) \longrightarrow \boxed{FFT} \longrightarrow \boxed{abs()} \longrightarrow \boxed{log()} \longrightarrow \boxed{IFFT} \longrightarrow c_x(n)$$

Figure 1. Computation of the cepstrum of a signal $x(n)$ requires the stages shown.

Speech Models

Within human speech, there are two methods employed to form our words. These sounds are categorized as *voiced* and *unvoiced*. For the voiced part, our throat acts like a transfer function. The vowel sounds are included in this category. The unvoiced part describes the noise-like sounds of speech. These are the sounds made with our mouth and tongue (as opposed to our throat), such as the "f" sound, the "s" sound, and the "th" sound. This way of looking at speech as two separable parts is known as the source filter model of speech or the linear separable equivalent circuit model. Mathematically, they are described in the time domain (assuming a discrete-time formulation) as:

$$g(n) = f(n) * h(n) = \sum_{k=0}^{n} f(k)h(n-k)$$

Since convolution in the time domain is multiplication in the frequency domain, taking the Fourier transform of both sides becomes:

$$G(\omega) = F(\omega) \cdot H(\omega)$$

An effective way to mathematically separate two multiplied functions is to take the logarithm of the magnitude of both sides to obtain:

$$\log |G(\omega)| = \log |F(\omega)| + \log |H(\omega)|$$

Computing the inverse Fourier transform of this equation into the cepstrum domain:

$$\Im^{-1}\{\log |G(\omega)|\} = \Im^{-1}\{\log |F(\omega)|\} + \Im^{-1}\{\log |H(\omega)|\}$$

where the independent variable resulting from the inverse Fourier transform is called "quefrency" and is in units of *time*. Therefore, quefrency is the horizontal axis of the cepstrum speech signal; typically areas of interest for speech are from 0 ms to 10 ms. Thus taking a given speech signal into the cepstral domain has advantages when trying to separate to convolved speech components.

The Use of the Cepstrum for Speech Processing

Often after having calculated the cepstrum, we want to *lifter* (i.e., *filter* in the cepstral domain). When we lifter, we are separating the transfer function and the excitation signal. The transfer function usually appears as a steep slant at the beginning of the plot. The excitation appears as periodic peaks occurring after around 3 to 9 ms. Figure 2 shows examples of cepstrum plots for vowel samples of male and female speech.
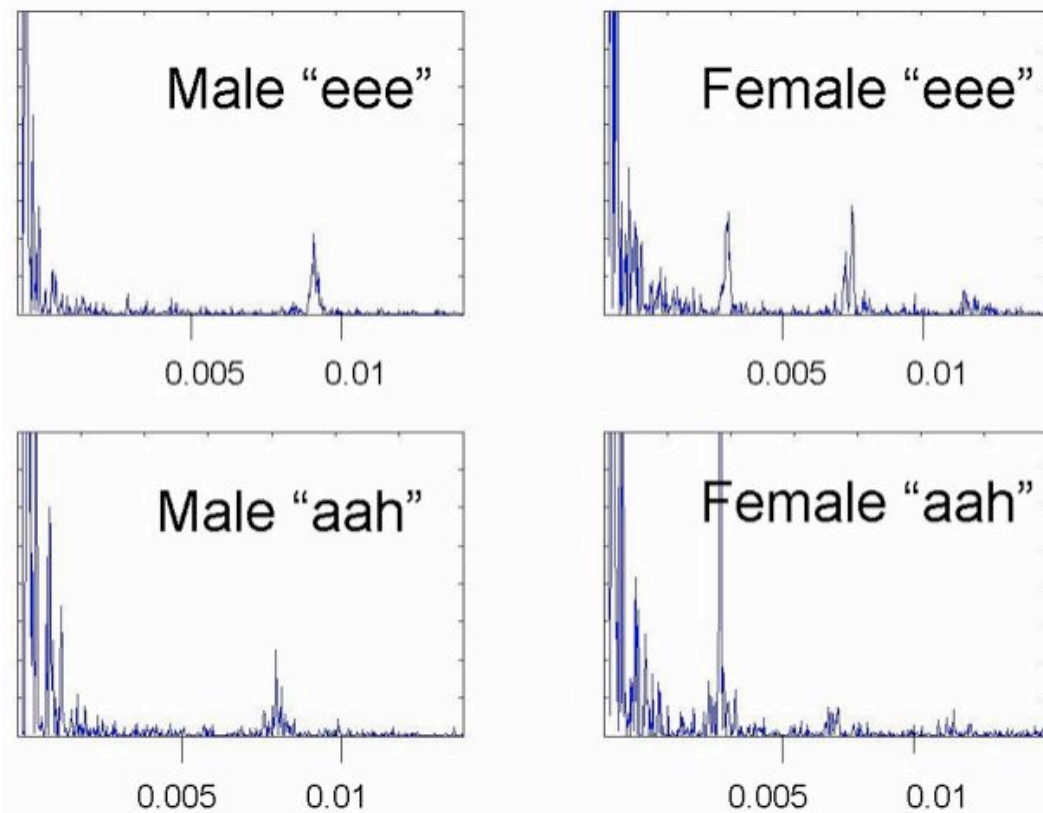
Figure 2. Vowel samples for a male and female test subject. The horizontal-axis is in units of milliseconds and the vertical axis is the magnitude of the cepstrum. The sounds "aah" and "eee" for both speakers are shown.

Project Guidelines

In this project you have to obtain samples of male and female voice signals for all five different vowel sounds. Many computers have built in microphones to do this and it is up to you how you will obtain your 10 speech samples. Then, you must compute the cepstrum of the sound samples and determine any differences you notice between male and females voice patterns and amongst the different vowels themselves. Last, you must lifter the signals to remove the transfer function (using an appropriate window length) and compute the excitation signal by taking the inverse cepstrum to get back to the time domain. Specific instructions follow in the next section.

Also, *please please please* note that **the project results must be original and your own work**. We have caught people in the past who have taken things off of the WWW for this project and copied from other people. As a result they have faced serious academic consequences.

Instructions

1. Acquire voice samples. In this part, please either record or find at least 10 voice samples of a *male* and *female* individual making the five vowel sounds – "a", "e", "i", "o", "u". If you are going to record them yourself or using a friend, then exaggerate the sounds a little and keep your voice extended for a while. It is also okay to work with other groups and use their voice samples,

but please credit them. Please make note of the conditions used to obtain the voice samples (e.g., what computer, what type of speaker – built-in or microphone, using which software program, or where or from whom the files were obtained). You should have 10 files in the end. If you want to be keen and impressive, you can get more than one male and female voice to obtain a better understanding of differences between both signals in the cepstral domain. Have fun with this.

2. Compute the cepstrum of each voice signal and discuss any difference qualitatively and quantitatively amongst male and female voices in general and amongst the different vowel sounds. This is an important component of the project, so please be creative and as comprehensive as possible. Your report should provide figures with original time-domain signals as well as cepstrum signals. Female voices should generally have more peaks than male voices in the cepstrum domain. You should discuss why you think this would be the case.

3. Lifter the cepstrum domain signals. Design a window (length is an important design parameter and you should discuss how and what you select – it can be the same or different for each speech sample depending on what you would like to experiment with) to remove the transfer function dependency. Then, compute the time domain signal of the corresponding windowed result to obtain the deconvolved signal. Plot the deconvolved result. Is there anything you can say about the signal and its difference from the original time domain recorded sample? Again, your discussion is an important part of the report.

Deliverables

Your results should be presented in a written report. There should be one report per person. You should submit your project report with the following components:

- A cover page with the following information: your name, student number, title of this course and project.

- Executive summary – you should briefly summarize your findings in this lab and convey the main points of interest to you.

- Introduction – discuss how you obtained the speech signals or where you obtained them from giving appropriate credit.

- Simulations and Lifter Design – discuss how you implemented your cepstral simulations, what program (e.g., MATLAB) and what steps you took to compute it. Discuss how you have designed you lifter lengths, etc. Please provide any simulation code you used in the Appendix section.

- Results – provide plots of results in as organized a manner as possible highlighting points of interest in the simulations to justify some of your discussion in the next section.

- Analysis and Discussion – discuss any insights you've found about the differences in vowel sounds and male/female voices in the cepstral domain. You can feel free to do research if you like and provide any references that you may have used to justify your statements.

- References – provide a list of references or links to websites of interest.

- Appendix – list the code for your simulations here. Please make it in small font (if needed) to avoid this section being very long.

Grading of the reports will be based on the quality of your report: insights presented, accuracy, organization, attention to detail. The length of the report can vary greatly from person to person because the results may take a lot of room or be compact depending on how one wants to present things, but if a range is needed at all, I would suggest (assuming 12-point font, 1" margins, single-spacing) the analyses and discussion parts be 2-5 pages.

Credits

This project is based on the following module available at the Connexions educational content repository:

Van Osdol, Brian. "Cepstrum." Connexions. December 21, 2004. http://cnx.org/content/m12469/1.4/.