Cepstrum*

Brian Van Osdol

Abstract

Use and application of the cepstrum domain for speech analysis.

1 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). It is similar to a channel vocoder or LPC[1] in its applications, but using the cepstrum as a spectral analyzer is a completely different process. (It is also worth pointing out that cepstrum is "spectrum" with the first syllable flipped... we will encounter several words with this naming convention.) Before describing the details of the cepstrum, a little background in speech models is needed.

2 Background Information

Within human speech, there are two methods employed to form our words. These sounds are categorized into the **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 "noisy" 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 seperable parts is known as the Source Filter Model of Speech [2] or the Linear Separable Equivalent Circuit Model. Mathematically, they are described in the time domain as:

$$x(t) = \int_0^t g(\tau)\, h(t-\tau)\, d\tau$$

Since convolution in the time domain is multiplication in the frequency domain, this become:

$$X(\omega) = G(\omega)\, H(\omega)$$

There is a mathematical process with which we are familiar that can separate two multiplied variables. If we take the log of the magnitude of both sides of this transform, we reach:

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

*Version 1.4: Dec 21, 2004 5:10 pm US/Central

†http://creativecommons.org/licenses/by/1.0

[1] "Linear Predictive Coding in Voice Conversion" <http://cnx.org/content/m12473/latest/>

[2] "The Source Filter Model of Speech" <http://cnx.org/content/m12470/latest/>

Computing the inverse Fourier Transform of this equation brings us into the realm of "quefrency."

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

Quefrency is the x-axis of the cepstrum. Its units are in time. Typically the areas of intest are from 0ms to around 10ms. See figure 1 below for the full process.

Cepstrum Block Diagram

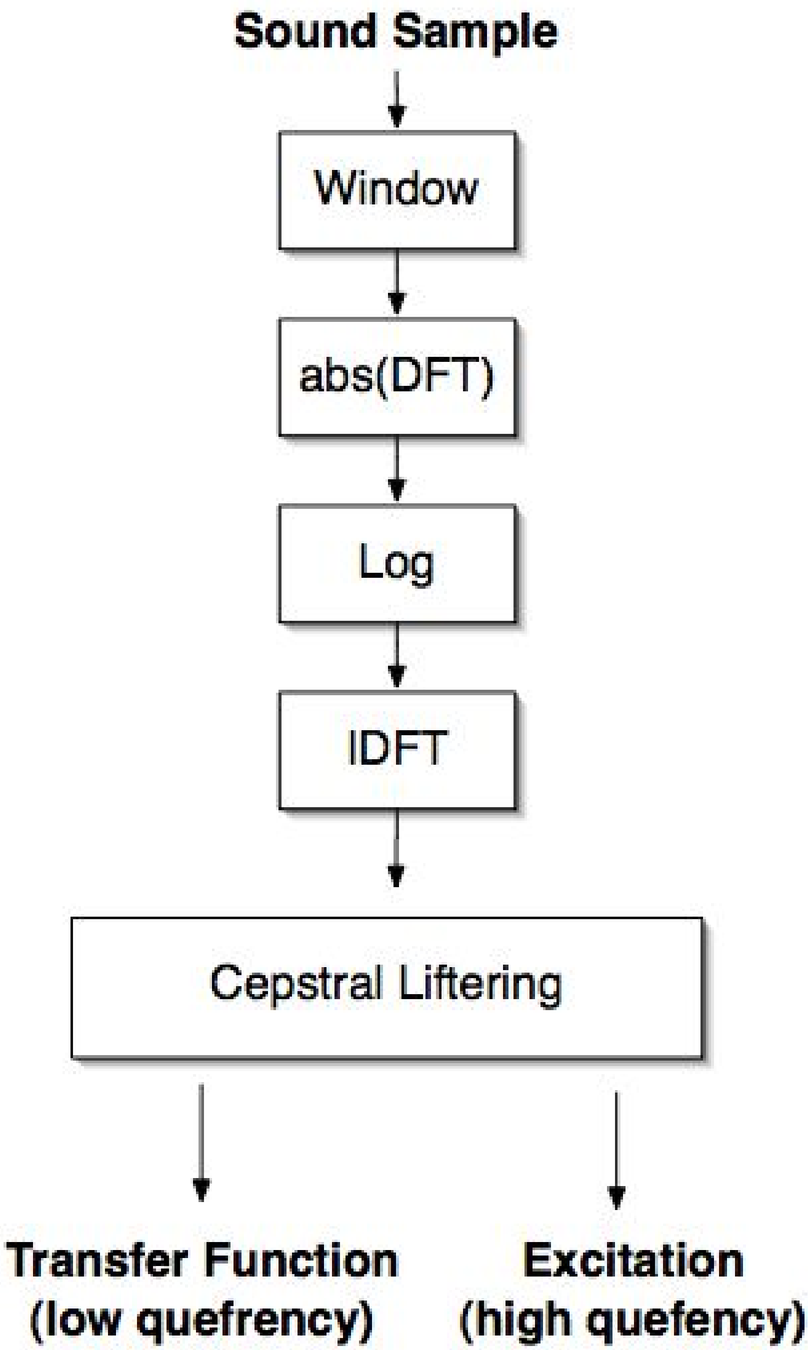

Figure 1: This is the process used to compute the cepstrum

3 The Use of the Cepstrum

We have now seen the process by which we calculate the cepstrum of a signal. It is now time to dicuss some the uses of the cepstrum. Often after having calculated the cepstrum, we will want to "lifter" the signal. (Once again the naming scheme has been used. This time on the word filter) When we lifter, we are seperating the **transfer function** (the spectral envelope) 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 5ms. Below we can see examples of several cepstrum plots. Note how the female voice has peaks occurring more often then in the male's cepstrum. This is due to the higher pitch of a female voice.

Cepstrum Samples

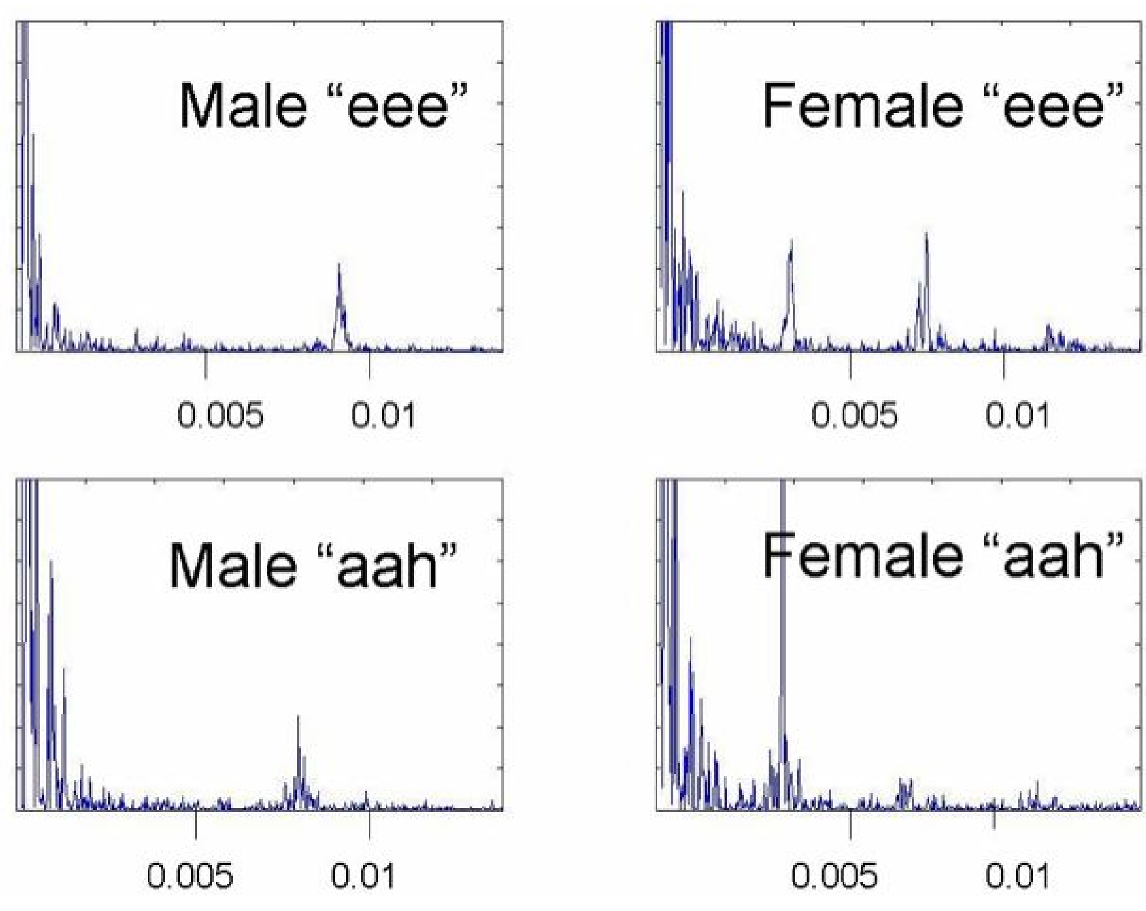

Figure 2: These are vowel samples of a male and female. The x-axis is in milliseconds and the y-axis is the magnitude. The sounds "aah" and "eee" for both speakers are shown.

4 Reference

Furui, Sadaoki. Digital Speech Processing, Synthesis, and Recognition. Marcel Deccer, Inc.: New York.

Gold, Ben and Nelson Morgan. Speech and Audio Signal Processing: Processing and Perception of Speech and Music. John Wiley and Sons, Inc: New York. 2000.