Windowing in Voice Conversion[*]

Matthew Hutchinson

Abstract

How to window a speech signal to prepare it for speech processing.

The characteristics of a speech signal vary with time. For this reason, it is difficult to process an entire phrase as tone, pitch, and other characteristics may have a very large range over the whole of the signal. Thus, it is necessary to split the phrase into many parts, or **window** the speech signal.

There are many different ways to window a signal. The first, and most common, method is to split the signal into equal parts. While this may prove useful for many applications, it is not the best technique for speech processing.

Speech is composed of many strung together segments known as syllables. Since each syllable contains unique characteristics, we find it very useful to split phrases into syllables before processing the signal. Syllable extraction amounts to looking for the breaks in the speech signal. This can be accomplished by performing **envelope extraction** on the absolute value of the speech signal and comparing to a threshold.

Definition 1: Envelope Extraction

Envelope Extraction is the process of obtaining the **evelope**, or general shape of, a signal.

In speech processing, the envelope can be extracted by taking the absolute value of the speech signal and subjecting it to an averager. An averager simply convolves the signal with a boxcar. The signal envelope is then compared to a threshold.

This is the process we employed in our windowing algorithm. A diagram is shown below:

[*] Version 1.4: Dec 21, 2004 5:13 pm -0600

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

Speech Signal Envelope

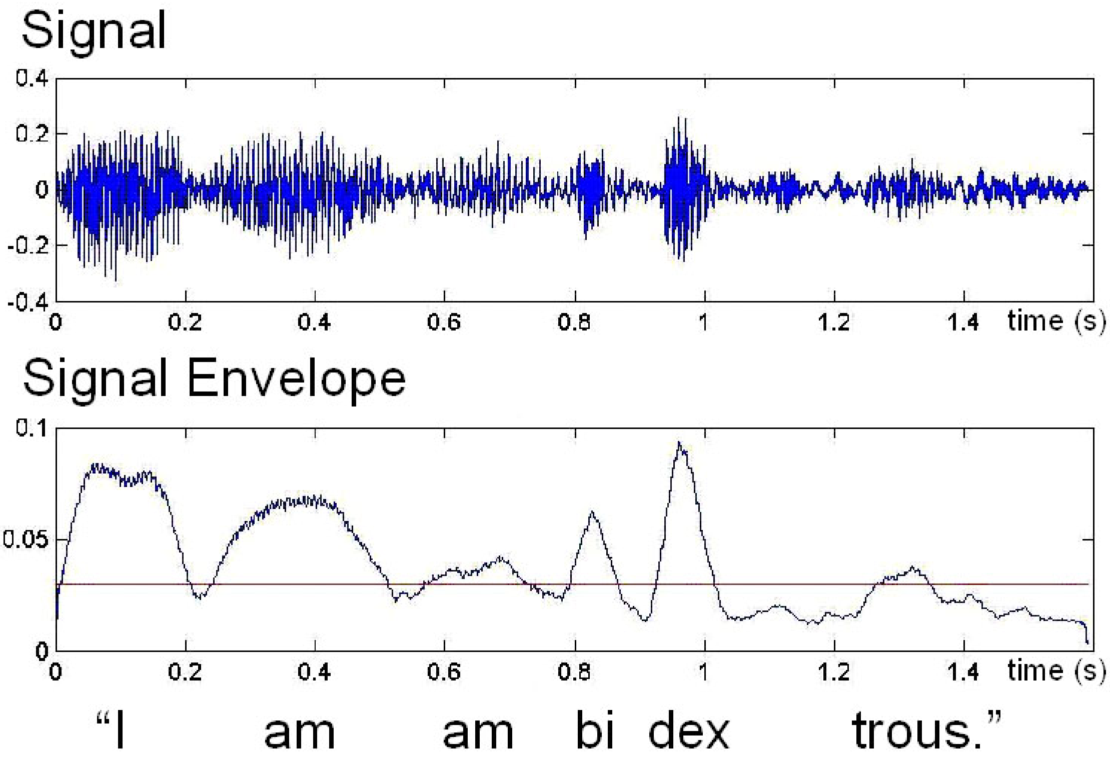

Figure 1: A speech signal and the envelope of its magnitude compared to a threshold.

When the signal envelope falls below the predetermined threshold, the speech signal is assumed to have a "break" in it - the end of one syllable and the beginning of another. Thus, the signal is split and each syllable is sent to the system for processing. After processing, the windows are reassembled to obtain the results for the entire phrase.