E1-17: Human voice and speech: Creating a simple speech recognizer for the Sinhala language

MP Warusawithana[1], DUJ Sonnadara[1,*], MK Jayananda[2], Asiri Nanayakkara[2]
[1]Department of Physics, University of Colombo
[2]Department of Physics, University of Sri Jayewardenepura

A Speech Recognizer which can recognize spoken words and convert them to text has numerous applications in commercial industry, in academic institutions as well as in our homes. Since the standard computer keyboard is designed to process the English language, constructing such a system for the Sinhala language is of particular importance to us.

The work described in this paper is directed towards building a simple speech recognizer that can convert voice signals to text.

First, the raw data corresponding to the speech signals were extracted from the pre-recorded sound files. The entire data set was chopped into windows of a finite number of samples. A Hamming window was used to reduce the errors due to discontinuities in the boundaries. By applying a Fast Fourier Transform (FFT), frequencies corresponding to these sound files were extracted. To limit the number of frequencies, a Mel scaled filter bank was applied on the extracted frequency spectrums. Finally a lookup table was constructed with the mean values of the selected frequencies for basic sounds in the Sinhala language.

The comparison of the test sounds with the lookup table produced quite remarkable results. With a limited set of sounds this technique can be used to produce a speaker independent speech recognizer. A direct industry application to such a system would be a voice command process control system. In building a speech recognizer, the accuracy can be further improved by using a higher number of frequencies or by increasing the sampling frequency.