Design and Implementation of a Speaker Identification System

Caleb Tenberge and Sunny Raheem

Advisor: Dr. Nidhal Bouaynaya

Department of Systems Engineering, University of Arkansas at Little Rock

Problem Overview and Applications

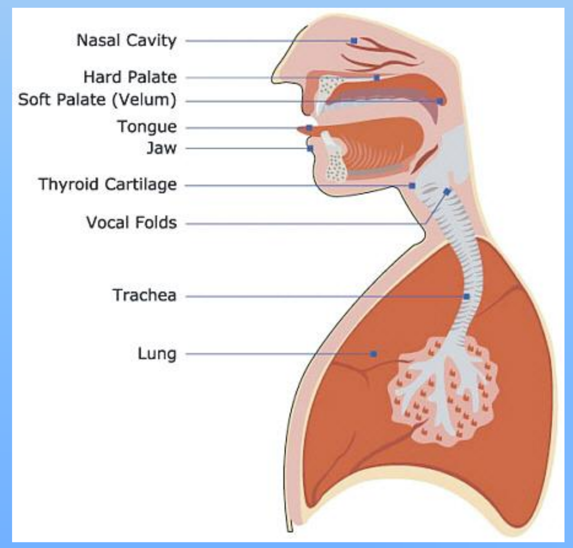
- The goal of this project is to design and implement a speaker identification system using signal processing techniques.
- Specifically, we propose to automatically recognize users known to the system (called claimants) and deny speakers unknown to the system who are posing as known speakers (called imposters).
- A speaker identification system, implemented in MATLAB, validates and/or detects a user's identity from their voice characteristics.
- This speaker identification system has applications in:
  1. Security (e.g., speech sample for authentication)
  2. Voice control and command (e.g., voice control of an autonomous vehicle).
  3. Biometric recognition systems, e.g., voice transmission technology allowing recognition over long distances via ordinary telephones (wired or wireless).

Required Digital Signal Processing Skills

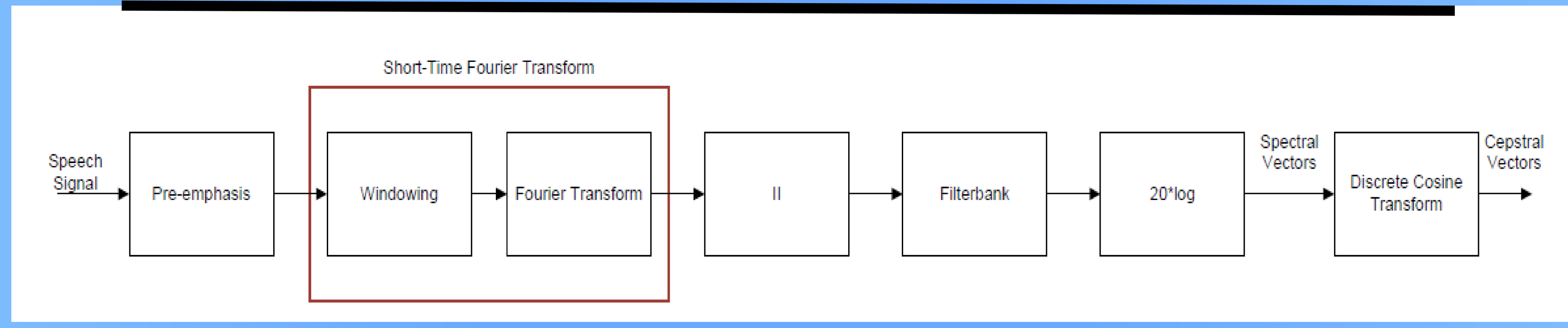
- Sampling Theory (Nyquist Theorem)
- Short-Time Fourier Transform (STFT)
- Filterbank
- Filter Design

Acoustic Theory of Speech Production

- Speech is an acoustic signal with having structure which has many levels such as, syllables, words, phrases, sentences, etc.
- Current speech recognition equipment is mainly concerned with decoding the words. These may convey the speaker's emotional or physical state, grammatical structure, or help to control the dialogue between multiple speakers.
- The science of Phonetics has concentrated on the processes between articulation and the acoustic Signal.



Speaker Identification System: Flow Block Diagram

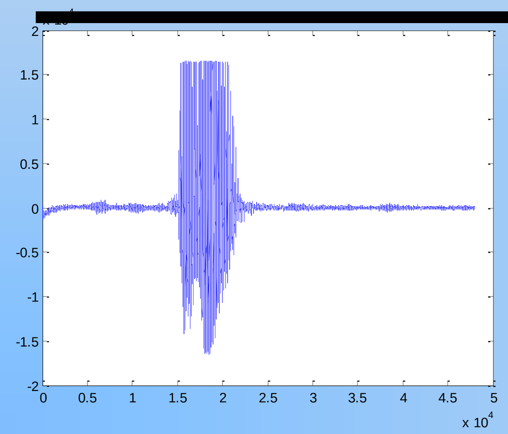


The flow block diagram is a modular representation of a cepstral-based speaker identification system. This diagram is used as a sketch of the components that needed to be implemented into a MATLAB program.

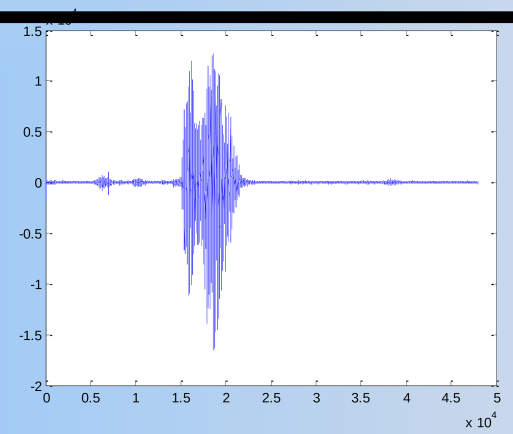
MATLAB Program

- **Speech Signal Generation** – The human speech spectrum is less than 4000 Hz. The sampling frequency of the recorded signal is 16000 samples/second or 16bit/sample. The speech signal generation is done by speaking into a computer recording program.
- **Pre-emphasis Filter** – The goal of this filter is to enhance the high frequencies of the spectrum, which is reduced by the speech production process. xp(t) = x(t) – ax(t-1) is programmed into MATLAB to apply pre-emphasis filter. Generally, *a* takes a value in the interval [0.95, 0.98].
- **Short-Time Fourier Transform** – An application of Hamming and Hanning window is applied to create the short time speech signal needed. The Hamming and Hanning windows taper the sides of the original signal, causing a reduction in the side effects. This window is first applied to the beginning of the signal and then moved further down the signal until it has reached the end of the signal. After the Fast Fourier Transform, each speech signal window application portion provides a spectral vector. Length of the window is either 20 or 30 milliseconds is set for the length of the window. The other quantity needed for the window is the shift amount between two consecutive windows, which is 10 milliseconds. By extracting this, the power spectrum is obtained. Since the power spectrum is symmetric, only half the points contain all the information that is necessary.
- **Filterbank** – Since the power spectrum has a lot of flexibility in it, the focus is on the envelop of the spectrum. By multiplying the spectrum the envelop is obtained by a Filterbank. Filterbank is defined by the shape of the filters and by their Frequency localization. Five Chebychev fourth-order band-pass filters ranging from 50Hz to 1500Hz are implemented. [B,A] = cheby2(2,20,[f0,f1] is the function used in MATLAB. The fingerprint of the speech is equal to the square of the output of each filter. By taking the log and multiplying each coefficient by 20 of this spectral envelop obtains the envelop in dB form. Now the spectral vectors can observed.
- **Discrete Cosine Transform** – Is applied to the spectral vectors by $c_n = \sum_{k=1}^{k} S_k \cos\left(n\left(k-\frac{1}{2}\right)\frac{3.14}{K}\right)$ n = 1,2,..L. K is the number of log-spectral coefficients calculated previously. Sk are the log-spectral coeffcients, and L is the number of cepstral coefficients that we want to calculate. Now the cepstral vectors are obtained for each window.

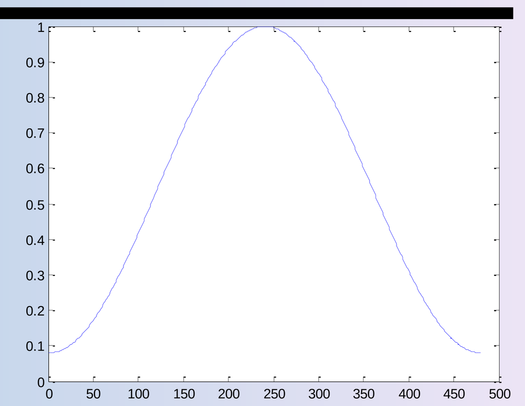
Simulation Results



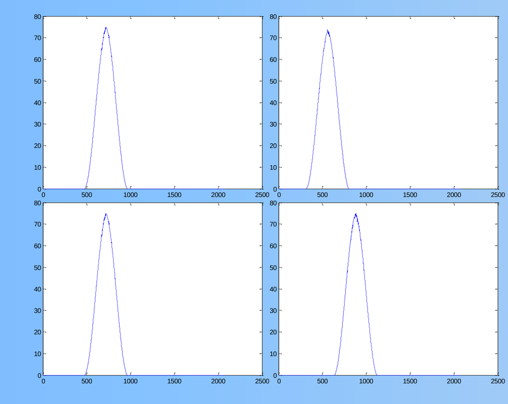
Input Speech Signal



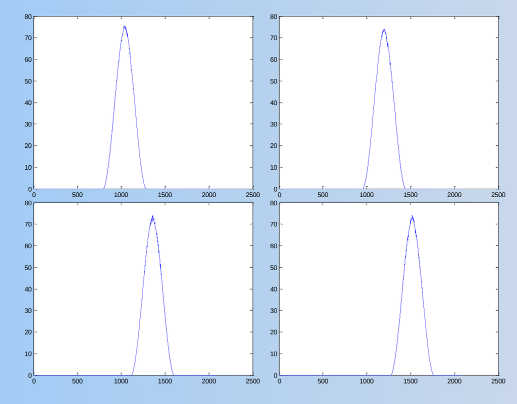
Enhanced Input Speech after pre-emphasis filter.



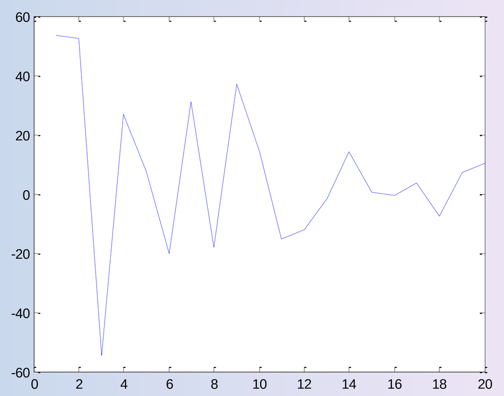
Hanning window we used to window the signal.



This is the output after the window was a applied.



The follow windowed sections.



After the DCT, a vector of the cepstral coefficients is obtained with which we can compare to others to find a match.

Conclusions

- Speaker identification is definitely possible by finding certain characteristics within the speech to which we can compare to others.
- This system could be used for many important applications, such as: command and control, security, user verification, etc.
- There are other characteristics in speech that could be used as an identifiers to compare the speech.

References

[1] Stevens, Acoustic Phonetics, MIT Press, 1998
[2] Allan South. Acoustic Theory of Speech Production, Chapter 2, April 2001.
[3] Rabiner & Schafer, *Digital Processing of Speech Signals, Prentice-Hall, 1978*

Acknowledgment

We would like to acknowledge Ms. Nidhal Bouaynaya, Ph. D., an Assistant Professor at the University of Arkansas at Little Rock, for her guidance and support on this project. Also, we acknowledge Mr. Ahmad Abdulsadda, a graduate student in the department for helping with the MATLAB code

UNIVERSITY OF ARKANSAS AT LITTLE ROCK