Gender Classification by Speech Analysis

BhagyaLaxmi Jena[1], Abhishek Majhi[2], Beda Prakash Panigrahi[3]
[1]Asst. Professor, Electronics & Tele-communication Dept., Silicon Institute of Technology
[2,3]Students of Electronics & Tele-communication Branch, Silicon Institute of Technology

Abstract

This paper is about a comparative investigation on speech signals to devise a gender classifier. Gender classification by speech analysis basically aims to predict the gender of the speaker by analyzing different parameters of the voice sample. This comparative investigation mainly concentrates on short-time analysis of the speech signals. The analysis includes comparison of short-time average magnitude, short-time energy, short-time zero crossing rate and short-time auto-correlation values of male and female voice samples. This quantitative comparison is implemented through MATLAB programming. A database consisting of voice samples collected from many students, both male and female, of our college was created. The short-time analysis was performed on all the collected voice samples and the parameters were compared to establish a working principle for the gender classifier from speech.

Keywords: short-time average magnitude, short-time energy, short-time zero crossing rate, short-time auto-correlation

1. Introduction

At a linguistic level, speech can be viewed as a sequence of basic sound units called phonemes. A phoneme is a sound or group of different sounds perceived to have the same function by the speakers of a language. An example of a phoneme is */k/* sound in the words *kit* and *skill*. The same phoneme may give rise to many different sounds or allophones at the acoustic level, depending on the phonemes which surround it. Different speakers producing the same string of phonemes convey the same information yet sound different as a result of differences in dialect and vocal tract length and shape.

1.1 Speech Analysis

The techniques used to process speech signals that can be broadly classified as either time-domain or frequency-domain analysis. In time-domain analysis, the measurements are performed directly on the speech signal to extract information. In frequency-domain analysis, the information is extracted after the frequency content of the speech signal computed to form the spectrum.

1.2 Gender Classifier

Gender Classifier from speech is a part of automatic speech recognition system to enhance speaker adaptability and a part of automatic speaker recognition system. The need for gender classification from speech also arises in several situations such as sorting telephone calls by gender for gender sensitive surveys. It is also a part of modern voice password technology.

2. Short-Time Analysis

Properties of a speech signal changes relatively slowly with time. Thus allows examination of a short-time duration of speech to extract parameters that are assumed to remain same for that time duration. This forms the basis of the short-time analysis. The speech signal is divided into many sub-signals of short-time duration by means of "windowing" technique. After splitting the large signal into many analysis frames with use of appropriate "windows", each frame is analyzed and then a cumulative result is obtained.

2.1 Short-Time Average Magnitude

Short-Time Average Magnitude (STAM) [7] is used for detecting the start point and end point of the speech signal. Short-Time Average Magnitude of a speech signal is given by

$$M_n = \sum_{m=-\infty}^{\infty} |x(m)| w(n-m)$$

where,
M_n = Short-Time Average Magnitude
x = Speech Signal
w = Window

2.2 Short-Time Energy

Short-Time Energy (STE) [4] is the energy associated with the signal in time domain. Short-Time Energy of a speech signal is given by

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2$$

where,
E_n = Short-Time Energy
x = Speech Signal
w = Window

2.3 Short-Time Zero Crossing Rate

Zero Crossing Rate (ZCR) [4] is the rate of sign-changes along a signal. Short-Time Zero Crossing Rate for a speech signal is given by

$$Z_n = \sum_{m=-\infty}^{\infty} \frac{|sgn[x(m)] - sgn[x(m-1)]|}{w(n-m)}$$

where,
Z_n = Short-Time Zero Crossing Rate
x = Speech Signal
w = Window

$$sgn[x(m)] = \begin{cases} 1, x(m) \geq 0 \\ -1, x(m) < 0 \end{cases}$$

2.4 Short-Time Auto-correlation

Short-Time Auto-correlation [1] for a speech signal is given by

$$R_n(k) = \sum_{m=-\infty}^{\infty} \frac{[x(m)w(n-m)]}{[x(m+k)w(n-m-k)]}$$

where,
$R_n(k)$ = Short-Time Auto-correlation
x = Speech Signal
w = Window
k = Sample time at which auto-correlation was calculated

3. Simulation & Results

In this comparative investigation, we have used voice samples from our database that contains voice samples of many students, both male and female, of our college. Each of the voice samples collected contain one predefined sentence (Oh My God), spoken by only one speaker, speaking in English, with no other background sounds. The comparative short-time analysis of male and female voice samples was done using MATLAB software.

It was observed that the average short-time energy value for female voice samples was greater than that of male voice samples for almost all the voice samples in our database. The STE plots for a randomly selected female and male voice sample are shown in figure 1.
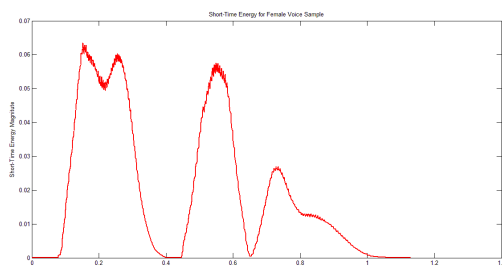

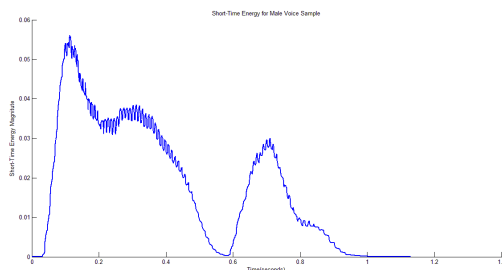Figure-1(a): Short-Time Energy plot of female voice sample

Figure-1(b): Short-Time Energy plot of male voice sample

It was also observed that the average short-time zero crossing rate value for female voice samples was higher than that of male voice samples throughout our database. The short-time ZCR plots for a randomly selected female and male voice sample are shown in figure 2.
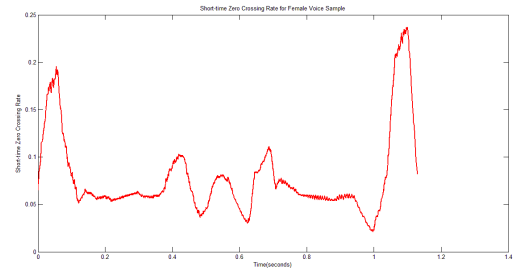

Figure-2(a): Short-Time Zero Crossing Rate plot of female voice sample
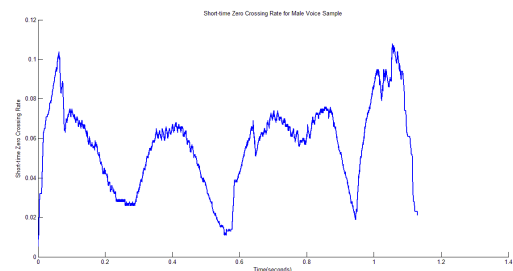

Figure-2(b): Short-Time Zero Crossing Rate plot of male voice sample

There was a significant difference in the short-time average magnitude and short-time auto-correlation plots of the male and female voice samples. The STAM plots for a randomly selected female and male voice sample are shown in figure 3.
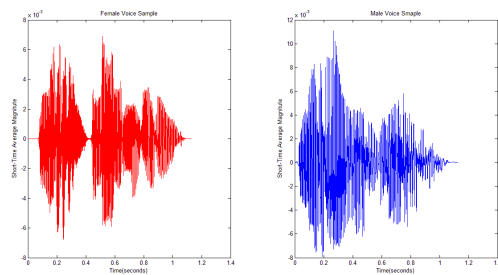
Figure-3: Short-Time Average Magnitude plots of female and male voice sample.
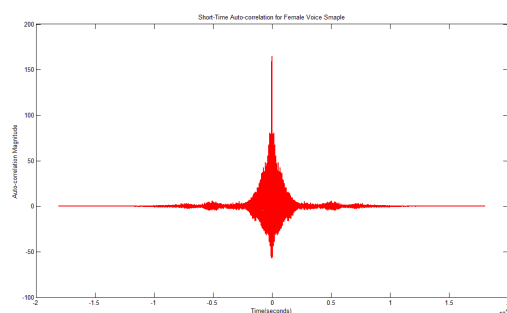

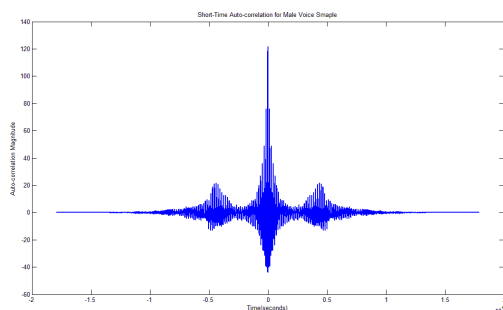Figure-4(a): Short-Time Autocorrelation plot of female voice sample

Figure-4(b): Short-Time Autocorrelation plot of male voice sample

All the convolutions computed during this analysis were based on FFT/IFFT algorithm [6] implemented in MATLAB software. Appropriate rectangular windows [3] were designed and used for the analysis.

Conclusion

By comparing the parameters obtained by short-time analysis of the male and female voice samples, it is observed that there is sufficient difference between the parameters. This difference in parameters can be used as the working principle of a Gender Classifier which predicts the gender of the speaker in a voice signal by analyzing it.

Our long term goal is to implement a gender classifier that can automatically predict the gender of the speaker based on the above investigation.

References

[1] H. Harb, L. Chen, J. Auloge, "*Speech/ Music/ Silence and Gender Detection Algorithm*"

[2] Vinay K. Ingle, John G. Prokakis, "*Digital Signal Processing Using MATLAB*"

[3] Chiu Ying Lay, Ng Hian James, "*Gender Classification from Speech*"

[4] Thomas F. Quatieri, "*Discrete-Time Speech Signal Processing*"

[5] John G. Proakis, Dimitris G. Manolakis, D. Sharma, "*Digital signal Processing, Principles, Algorithms and Applications*"

[6] Douglas O'shaughnessy, "*Speech Communications: Human and Machine*"

[7] Lawrence R. Rabiner, Biing-Hwang Juang, "*Fundamentals of Speech Recongnition*"

[8] Joseph Mariani, "*Language and Speech Processing*"

[9] Tanja Schultz, "*Speaker Characteristics*"

[10] Thomas F. Quatieri, "*Discrete-Time Speech Signal Processing*"

[11] Christian Muller, "*Speaker Classification: Fundamentals, Features and Methods*"

[12] E. Parris, M. Carey, "*Language Independent Gender Indentification*"