GENDER RECOGNITION SYSTEM USING SPEECH SIGNAL

Md. Sadek Ali[1], Md. Shariful Islam[1] and Md. Alamgir Hossain[1]

[1] Dept. of Information & Communication Engineering
Islamic University, Kushtia 7003, Bangladesh.
E-mail :{ sadek_ice, afmsi76, alamgir_ict}@yahoo.com

ABSTRACT

In this paper, a system, developed for speech encoding, analysis, synthesis and gender identification is presented. A typical gender recognition system can be divided into front-end system and back-end system. The task of the front-end system is to extract the gender related information from a speech signal and represents it by a set of vectors called feature. Features like power spectrum density, frequency at maximum power carry speaker information. The feature is extracted using First Fourier Transform (FFT) algorithm. The task of the back-end system (also called classifier) is to create a gender model to recognize the gender from his/her speech signal in recognition phase. This paper also presents the digital processing of a speech signals (pronounced "A" and "B") which are taken from 10 persons, 5 of them are Male and the rest of them are Female. Power Spectrum Estimation of the signal is examined .The frequency at maximum power of the English Phonemes is extracted from the estimated power spectrum. The system uses threshold technique as identification tool. The recognition accuracy of this system is 80% on average.

KEYWORDS

Gender Recognition, Feature Extraction, First Fourier Transform (FFT), Font-end, Back-end.

1. INTRODUCTION

"Speech" according to Webster's Dictionary is the "communication or expression of throughout in speaker words". Speech signal not only caries the information that is need to communicate among people but also contents the information regarding the particular speaker. The nonlinguistic characteristics of a speaker help to classify speaker (male or female). Features like power spectrum density, frequency at maximum power carry speaker information. These speaker features can be tracked well varying the frequency characteristics of the vocal tract and the variation in the excitation. The speech signal also carry the information of the particular speaker including social factors, affective factor and the properties of the physical voices production apparatus for which human being are able to recognize whether the speaker is a male or a female easily, during telephone conversation or any hidden condition of the speaker [1][2]. With the current concern of security worldwide gender classification has received great deal of attention among of the speech researchers. Also a rapidly developing environment of computerization, one of the most important issues in the developing world is gender recognition.

Gender recognition, which can be classified into two different tasks: *Gender identification* and *Gender verification*. In the identification task, or *1: N matching*, an unknown speaker is compared against a database of *N known* speakers, and the best matching speaker is returned as the recognition decision. The verification task, or *1:1 matching*, consists of making a decision whether a given voice sample is produced by a claimed speaker. An *identity claim* (e.g., a PIN

DOI : 10.5121/ijcseit.2012.2101

code) is given to the system, and the unknown speaker's voice sample is compared against the claimed speaker's voice template. If the similarity degree between the voice sample and the template exceeds a predefined *decision threshold*, the speaker is accepted, and otherwise rejected.

The rest of the paper is organized as follows. In section II shows the related works. Section III describes the mathematical tools and techniques for gender recognition systems. Section IV describes speech recording and feature extraction process. Computations of power and frequency spectrum are described in section V. System implementation is detailed in section VI. Recognition and experimental results are given in section VII. Finally, section VIII concludes the paper.

2. RELATED WORKS

Gender recognition is a task of recognizing the gender from his or her voice. With the current concern of security worldwide speaker identification has received great deal of attention among of the speech researchers. Also a rapidly developing environment of computerization, one of the most important issues in the developing world is speaker identification. Speech processing based several types of research work have been continuing from a few decade ago as a field of digital signal processing (DSP). The most efficient related work is "Speaker recognition in a multi-speaker environment" was submitted in *Proc. 7th European Conference on Speech Communication and Technology (Eurospeech 2001)* (Aalborg, Denmark, 2001), pp. 787–90[3]. Another related work is "Spectral Feature for Automatic Voice-independent Speaker Recognition" was developed in the department of Computer Science, Joensuu University, Finland 2003 [4]. From the study of different previous research works it was observed that among the different features the power spectrum results in best classification rate. Based on the power spectrum, we have computed frequency spectrum from maximum power of speech signal. We have implemented a complete gender recognition system to identify particular gender (male/female) using frequency component. In addition to description, theoretical and experimental analysis, we provide implementation details as well.

3. MATHEMATICAL TOOLS AND TECHNIQUES

For digital communication or digital signal synthesis, it is necessary to convey the analog signal such as speech, music etc. as a sequence of digitized number, which is commonly done by sampling the speech signal denoted by X_n (t) periodically to produce the sequence

$$X(n) = X_n(nt) \qquad \alpha < n < \alpha \quad \ldots\ldots(1)$$

Where n_0 have only integer value. In this paper, we have used pulse code modulation (PCM) technique to digitize speech signal.

The sampled data are operated to find out the different parameter. Discrete Fourier Transform (DFT) computes the frequency information of the equivalent time domain signal [5]. Since a speech signal contains only real point values, we can make use of this fact and use a real-point Fast Fourier Transform (FFT) for increased efficiency. The resulting output contains both the magnitude and phase information of the original time domain signal. The Short time Fourier analysis of windowed speech signal can produce a reasonable feature space for recognition [6]. The Fourier Transform for a discreet time signal $f(kT)$ is given by

$$F(n)=\sum_{k=0}^{N-1} f(kT)e^{-j2\pi/N nk} \quad \text{..... (2)}$$

Equation (2) can be written as

$$F(n)=\sum_{k=0}^{N-1} f(k)W_N^{-nk} \quad \text{..... (3)}$$

Where $f(k) = f(kT)$ and $W_N = e^{j2\pi/N}$, W_N is usually referred to as the *kernel* of the transform. There are several algorithms that can considerably reduce the number of computations in a DFT. DFT implemented using such schemes is referred to as Fast Fourier Transform (FFT). Among the FFT algorithms there are two popular algorithms: decimation-in-time and decimation- in frequency. Through out this paper, the DFT is computed using decimation-in-time algorithm.

4. SPEECH RECORDING AND FEATURE EXTRACTION

Human speech signal contains significant amount of energy within 2.5 KHz. So, we have taken the sampling rate of speech signal as 8 KHz, 8 bit mono, which is sufficient for representing signals up to 4 KHz without aliasing effect. To record speech signal we have used Intel(r) Integrated sound card, a normal microphone and windows default sound recorder software. Speech was recorded in room environment. The recorded sound was stored in PCM (.wav) sound file format. The file header is stored at the beginning of the PCM file and occupied 44 bytes [7]. We also know that the actual wave data are stored after 58 bytes from the beginning. So, to extract wave data, we first discard 58 bytes from the beginning of the wave file and then read wave data as character. This data are stored in a text file (.txt) as integer data.

Feature extraction is the process of converting the original speech signal to a parametric representation that gives a set of meaningful features useful for recognition. Feature extractions is the combination of some signal processing steps including the computation of drive data from wave sound, computation of Fast Fourier Transform (FFT), Power spectrum, the sample point at maximum power and finally compute the frequency.

5. COMPUTATION OF POWER AND FREQUENCY

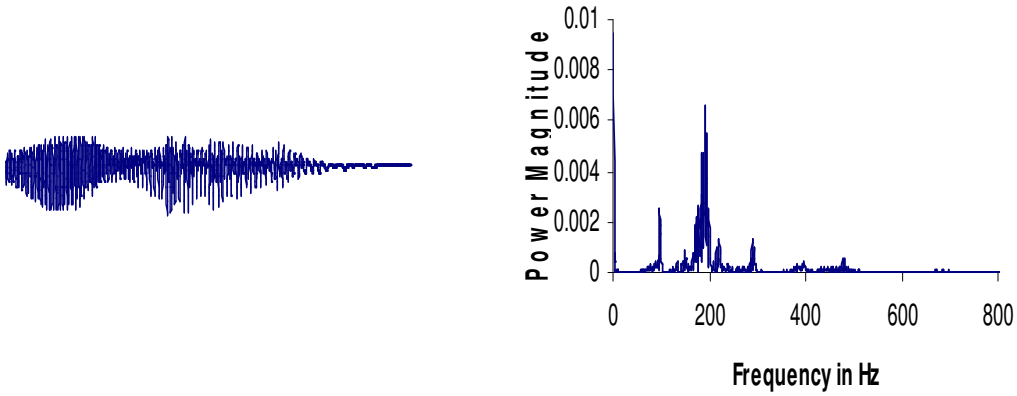
Power spectrum estimation uses an estimator called peridogram [5][6]. The power spectrum is defined at N/2+1 frequency as

$$P(f_k) = 1/N^2\left[|F_k|^2+|F_{N-k}|^2\right]---------k = 1,2,.....,\left(N/2-1\right)$$

Where f_k is defined only for the zero and positive frequencies

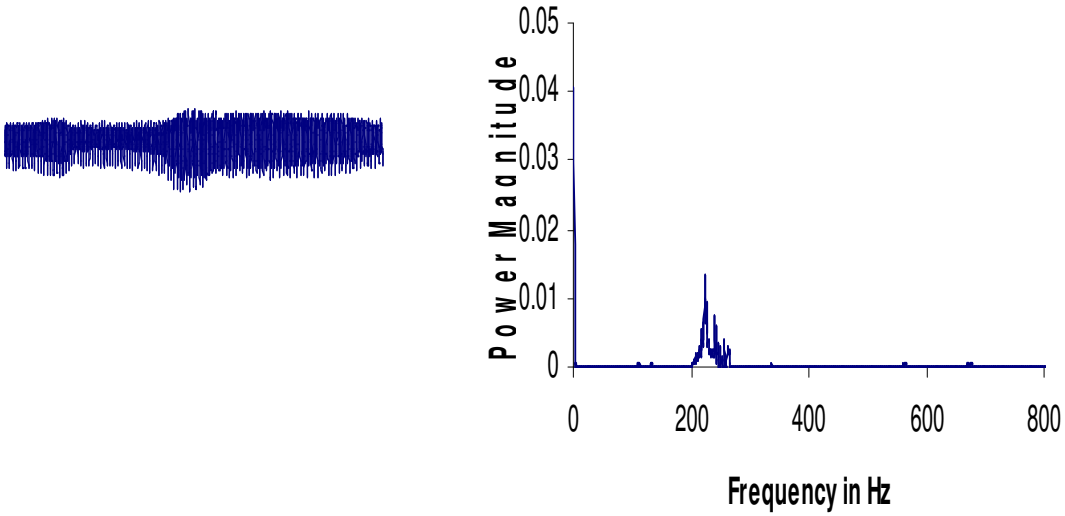
$$f_k \equiv k/N\Delta = 2f_c\, k/N -------k = 0,1,....,N/2$$

To compute the power spectrum of speech, the speech data is segmented into K segments of N=2M points. Where N is the length of a window, taken as a power of 2 for the convenient computation of FFT. Each segment is FFTed separately and the resulting K periodograms are averaged together to obtain a Power Spectrum estimate at M frequency values between 0 and f_c. The figure 1 and 2 shows the signal waveform (fig1.a and fig2.a), power spectrum (fig1.b and fig2.b) of male and female speaker for phoneme "**A**".



(a) (b)

Figure 1. The signal waveform and power spectrum of a male speaker for Phoneme “A”.



(a) (b)

Figure 2. The signal waveform and power spectrum of a female speaker for phoneme “**A”**.

The computation steps for frequency can be summarized as shown in figure 3.

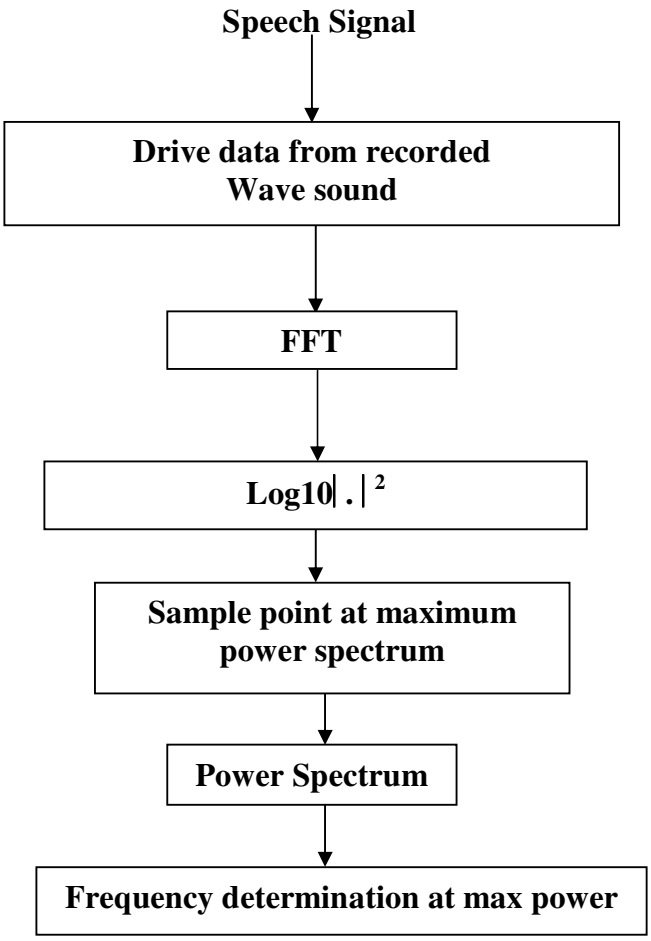


Figure 3. The sequence of operations in converting a speech signal into Frequency features

6. SYSTEM IMPLEMENTATION

The following figure 4 shows the abstraction of a gender recognition system. Regardless of the type of the task (classification or verification), gender recognition system operates in two modes: *training* and *recognition* modes. In the training mode, a new gender person's voice is recorded and analysis. The recognition mode, an unknown gender person gives a speech input and the system makes a decision about the speaker's identity. Both the training and the recognition modes include *feature extraction*, sometimes called the *front-end* of the system. The feature extractor converts the digital speech signal into a sequence of numerical descriptors, called *feature vectors*. The features provide a more stable, robust, and compact representation than the raw input signal. Feature extraction can be considered as a data reduction process that attempts to capture the essential characteristics of the speaker with a small data rate.

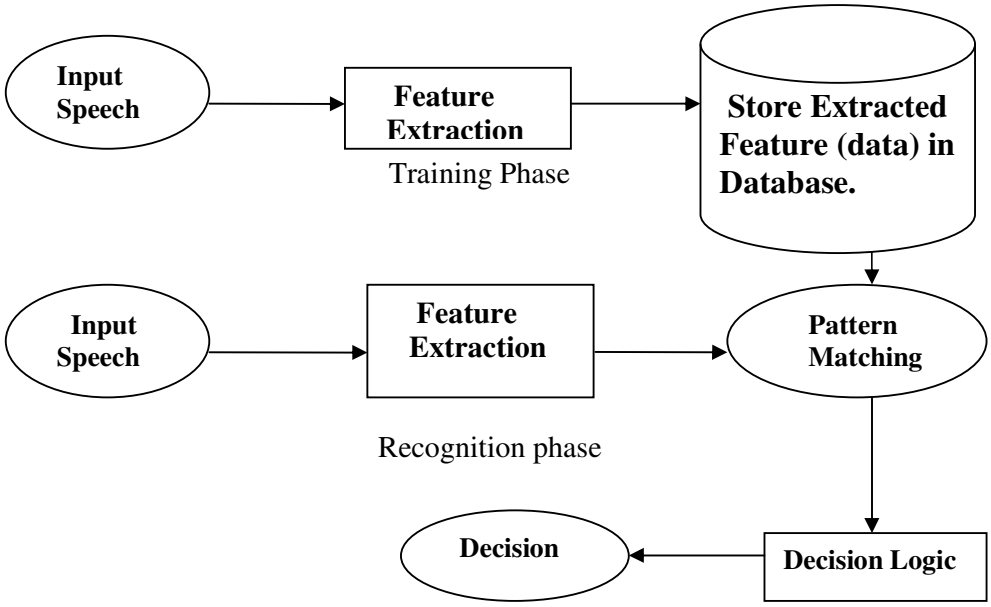


Figure 4. Block diagram of the Gender Recognition System

The software system contains the following steps:

1. To code and record the speech data as a disk file

2. To compute the Fast Fourier Transform of the data from recorded file

3. To compute the power spectrum from transformer data of step 2.

4. To compute the frequency from power spectrum of step 3.

5. To identify the Speaker as a male person or as female person or give a message that indicate the speaker is neither a male person nor a female person.

The developed system is applied to store and analyze English phonemes "**A**" and "**B**". The speech data are recorded for male and female speaker, FFT and power spectrum is calculated using these data files by the developed system. The results are shown in tabular form in Table1:

Table 1. Frequency at maximum power of some male and female corresponding English phonemes "**A**" and "**B**".

Speaker ID	Phonemes	No.of sample	Frequency at maximum power (Male voice)	Frequency at maximum power (Female voice)
1	A	1	570	672
2	A	1	588	720
3	A	1	498	672
4	A	1	462	729
5	A	1	432	687
1	B	2	432	618
2	B	2	342	402
3	B	2	543	666
4	B	2	366	612
5	B	2	510	414

7. SYSTEM RESULTS

7.1. RECOGNITION RESULTS

The detailed recognition results are shown in Table 2.

Table 2. For sample #1

No. of speaker	No. of accuracy of Gender	Recognize Percentage (%)
1	Male	100
2	Male	100
3	Male	100
4	Male	100
5	Male	100
6	Female	100
7	Female	100
8	Female	100
9	Female	100
10	Female	100

Table 2. For sample #2

No. of speaker	No. of accuracy of Gender	Recognize Percentage (%)
1	Male	100
2	Male	0
3	Male	100
4	Male	0
5	Male	100
6	Female	100
7	Female	0
8	Female	100
9	Female	100
10	Female	0

7.2. EXPERIMENTAL RESULTS

The system was tested with 10 speakers (5 male people and 5 female people).The speech (utterance) of word ("A" and "B") of 10 speakers was recorded separately. The threshold technique was used in recognition. The objective of the experiment was to study the recognition performance of the system with different speakers. The percentage of accuracy rate of recognition has been calculated using following equations:

$$\text{Recognition Accuracy (\%)} = \frac{\text{No. of accurately recognized gender}}{\text{No. of correct gender expected}} \times 100 \quad \text{..... (4)}$$

From recognition result, number of accurately recognized gender 16, number of correct gender expected 20.The recognition accuracy is = 16*100/20 = 80%.

8. CONCLUSION

In this paper the main goal was to develop a gender recognition system using speech signal. The feature selection is one of the most important factors in designing a gender recognition system. From the study of different previous research works it was observed that among the different features the power spectrum results in best classification rate. Thus the power spectrum has been selected as the feature for classification. Among the different technique, the statistical analysis and threshold technique is simple in computation and produces very good results. This is why this method was selected for pattern comparison in recognition process to obtain improved performance. The average recognition accuracy is 80 %. From experimental result, it can be seen that recognition rate decreases as the number of speaker increases. Therefore, the system efficiency is decreases, if the number of reference speakers in the speaker database increases. For recognition constant thresholds have been used. If we could use dynamic threshold for recognition it might produce more accurate and better recognition results.

REFERENCES

[1]. Prabhakar, S., Pankanti, S., and Jain, A. *"Biometric recognition: security and privacy concerns"* IEEE Security and Privacy Magazine 1(2003), 33-42.

[2]. Huang X., Acero, A., and Hon, H.-W. *"Spoken Language Processing: a Guideto Theory, Algorithm and System Development"* prentice-Hall, New Jersey, 2001.

[3]. Martin, A., and Przybocki, M. *Speaker recognition in a multi-speaker environment*.In Proc. 7th European Conference on Speech Communication and Technology (Eurospeech 2001) (Aalborg, Denmark, 2001), pp. 787–90.

[4]. Tomi Kinnunen *" Spectral Feature for Automatic Voice-independent Speaker Recognition"* Depertment of Computer Science, Joensuu University,Finland. December 21, 2003.

[5]. John R. Deller, John G Proakis and John H. L. Hansen, *"Discrete- Time Processing of Speech Signals"* Macmillan Publishing company, 866 Third avenue, New York 10022.

[6]. Rabiner Lawrence, Juang Bing-Hwang, *"Fundamentals of Speech Recognitions"*, Prentice Hall New Jersey, 1993, ISBN 0-13-015157-2.

[7]. Md. Saidur Rahman, *"Small Vocabulary Speech Recognition in Bangla Language"*, M.Sc. Thesis, Dept. of Computer Science & Engineering, Islamic University, Kushtia-7003, July-2004.

Biography

Md. Sadek Ali received the Bachelor's and Master's Degree in the Department of Information and Communication Engineering (ICE) from Islamic University, Kushtia, in 2004 and 2005 respectively. He is currently a Lecturer in the department of ICE, Islamic University, Kushtia-Bangladesh. Since 2003, he has been working as a Research Scientist at the Signal and Communication Research Laboratory, Department of ICE, Islamic University, Kushtia, where he belongs to the spread-spectrum research group. He has three published paper in international and one national journal in the same areas. He has also two published paper in international and one national conference in the same areas. His areas of interest include Signal processing, Wireless Communications, optical fiber communication, Spread Spectrum and mobile communication.

Md. Shariful Islam received the Bachelor's and Master's Degree in Applied Physics, Electronics & Communication Engineering from Islamic University, Kushtia, Bangladesh in 1999, and 2000 respectively. He is currently Assistant Professor in the department of ICE, Islamic University, Kushtia-7003 Bangladesh. He has five published papers in international and national journals. His areas of interest include signal processing & mobile communication.

Md. Alamgir Hossain received the Bachelor's and Master's Degree in the Dept. of Information and Communication Engineering (ICE) from Islamic University, Kushtia, in 2003 and 2004, respectively. He is currently Lecturer in the department of ICE, Islamic University, Kushtia-7003, and Bangladesh. He was a lecturer in the Dept. of Computer Science & Engineering from Institute of Science Trade & Technology (Under National University), Dhaka, Bangladesh. from 23th October, 2007 to 18th April 2010. Since 2003, he has been working as a Research Scientist at the Communication Reasearch Laboratory, Department of ICE, Islamic University, Kushtia, where he belongs to the spread-spectrum research group. He has two published paper in international and one national journal in the same areas. His current research interests include Wireless Communications, Spread Spectrum and mobile communication, Signal processing, Data Ming.