ISSN 2229-5518

Recognotion of Speaker Useing Mel Frequency Cepstral Coefficient & Vector Quantization for Authentication

Priyanka Mishra, Suyash Agrawal

Abstract— Human Voice is characteristic for an individual. The ability to recognize the speaker by his/her voice can be a valuable biometric tool with enormous commercial as well as academic potential. Commercially, it can be utilized for ensuring secure access to any system. Academically, it can shed light on the speech processing abilities of the brain as well as speech mechanism. In fact, this feature is being used preliminarily along with other biometrics including face and finger print recognition for commercial security products. Speaker recognition is the method of automatically identify who is speaking on the basis of individual information integrated in speech waves. There are two types of speaker recognition systems basically divided into two –classification: speaker identification and speaker verification.

Index Terms— The Sound Wave, a Band Pass Filter of bandwidth, vector quantization techniques, Voice Recognition Algorithm, MFCC & Vector Quantization .

1 INTRODUCTION

Speech is produced through the biological system of a larynx or sound box, which resides in the throat of the human beings The larynx or the sound box is provided with some fibres which are capable to vibration when the air passes through them .The major organs that pump air through the larynx are the lungs attached to the larynx by the windpipe. The laryngeal fibres or vocal cords, as they are properly capable of vibrating at all frequencies. In this particular case, the range goes from the just audible 20 Hz up to about 11000 Hz. The higher frequencies found usually in the children and the women and the lower in the men.

In the study, the effectiveness of combinations of cepstral features, channel compensation techniques, and different local distances in the Dynamic Time Warping (DTW) algorithm is experimentally evaluated in the text-dependent speaker identification task. The training and the testing has been done with noisy telephone speech (short phrases in Bulgarian with length of about 2 seconds) selected from the BG-SRD at corpus. The employed cepstral features are – Linear Predictive Coding derived Cepstrum (LPCC), Mel-Frequency Cepstral Coefficients (MFCC), Adaptive Component Weighted Cepstrum (ACWC), Post-Filtered Cepstrum (PFC) and Perceptually Linear Predictive coding derived Cepstrum (PLPC).

Priyanka mishra Student of M.Tech IV Sem in Computer Technology-Department of Computer Science & Engineering Rungta college of Engineering & Technlogy, Kurud Bhilai, CSVT University, Chhattisgarh,India, PH.919755931266 Email: mishrapriyanka414@gmail.com)
Suyash Agrawal Reader in TechnologyDepartment of Computer Science & Engineering Rungta college of Engineering & Technlogy, Kurud Bhilai, CSVT University, Chhattisgarh,India Email: suyash.agrawal1983@ gmail.com)

2 Procedure for Paper Submission

2.1 Review Stage

Human Voice recognition has been an interesting research field for the last decades, which still yields a number of unsolved problems. This paper aims to present a speaker recognition system which recognizes the speaker as opposed to what is being said by the speaker as in case of speech recognition. The methodology followed in this paper for Speaker recognition is using Feature Extraction process and then Vector Quantization of extracted features is done. At last the speaker is identified by comparing the data from a tested speaker to the codebook of each speaker and then measuring the difference. [14] Speech processing is emerged as one of the important application area of digital signal processing. Various fields for research in speech processing are speech recognition, speaker recognition, speech synthesis, speech coding etc. The objective of automatic speaker recognition is to extract, characterize and recognize the information about speaker identity. Feature extraction is the first step for speaker recognition. [1]

2.2 Final Stage

We studied for many papers in Voice Recognition. The problem identification work depends solely on literature survey. Problem identification is a process of identifying the problem and it define the problem clearly. In the literature survey it is found that various author used various methods for speaker identification but still there is a scope for future work according to my shown method. Now in problem identification we will show the method used in literature survey and how i can improve these methods to get accurate speaker identification using the proposed work. But Speaker recognition is basically divided into two-classification: speaker recognition and speaker identification and it is the method of

automatically identify who is speaking on the basis of individual information integrated in speech waves. Speaker recognition is widely applicable in use of speaker's voice to verify their identity and control access to services such as banking by telephone, database access services, voice dialing telephone shopping, information services, voice mail, security control for secret information areas, and remote access to computer AT and T and TI with Sprint have started field tests and actual application of speaker recognition technology; many customers are already being used by Sprint's Voice Phone Card.

2.3 Figer

At the highest level, all speaker recognition systems contain two main modules feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the voice signal that can later be used to represent each speaker. Feature matching involves the actual procedure to identify the unknown speaker by comparing extracted features from his/her voice input with the ones from a set of known speakers. We will discuss each module in detail in later sections. Although voice authentication appears to be an easy authentication method in both how it is implemented and how it is used, there are some user influences that must be addressed:

• Colds. If the user has a cold which affects his or her voice that will have an effect on the acceptance of the voice-scanning device. Any major difference in the sound of the voice may cause the voice-scanning device to react in a negative way, causing the system to reject the user.

• Expression and volume. If a person is trying to speak with expressions on their face (i.e. smiling at the same time) their voice will sound different. The user of the device must also be able to speak loudly and clearly in order to obtain accurate results.

• Misspoken or misread prompted phrases. If the user is required to authenticate by speaking a prompted phrase and they mispronounce the phrase, they will be rejected by the system.

• Previous user activity may have an impact on the outcome of the voice scanning device. For example, if the user is out of breath and is unable to speak well.

3 Problem Identification

The problem identification work depends solely on literature survey. Problem identification is a process of identifying the problem and it define the problem clearly. In the literature survey it is found that various author used various methods for speaker identification but still there is a scope for future work according to my shown method. Now in problem identification we will show the method used in literature survey and how i can improve these methods to get accurate speaker identification using the proposed work.

Speaker recognition is basically divided into two-classification: speaker recognition and speaker identification and it is the method of automatically identify who is speaking on the basis of individual information integrated in speech waves. Speaker recognition is widely applicable in use of speaker's voice to verify their identity and control access to services such as banking by telephone, database access services, voice dialing telephone shopping, information services, voice mail, security control for secret information areas, and remote access to computer AT and T and TI with Sprint have started field tests and actual application of speaker recognition technology; many customers are already being used by Sprint's Voice Phone Card. Speaker recognition technology is the most potential technology to create new services that will make our everyday lives more secured. Another important application of speaker recognition technology is for forensic purposes. Speaker recognition has been seen an appealing research field for the last decades which still yields a number of unsolved problems.

4 Citations

Techniques of Feature Extraction: The general methodology of audio classification involves extracting discriminatory features from the audio data and feeding them to a pattern classifier. Different approaches and various kinds of audio features were proposed with varying success rates. Some of the audio features that have been successfully used for audio classification include Mel-frequency cepstral coefficients (MFCC), Linear predictive coding (LPC), Local discriminant bases (LDB). Few techniques generate a pattern from the features and use it for classification by the degree of correlation. Few other techniques use the numerical values of the features coupled to statistical classification method.

A. LPC

LPC (Linear Predictive coding) analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal is called the residue. In LPC system, each sample of the signal is expressed as a linear combination of the previous samples. This equation is called a linear predictor and hence it is called as linear predictive coding .The coefficients of the difference equation (the prediction coefficients) characterize the formants.

B. MFCC

MFCC is based on the human peripheral auditory system. The human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency t measured in Hz, a subjective pitch is measured on a scale called the 'Mel Scale' .The mel

frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1kHz.As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels.

C. LDB

LDB is an audio feature extraction and a multi group classification scheme that focuses on identifying discriminatory time-frequency subspaces. Two dissimilarity measures are used in the process of selecting the LDB nodes and extracting features from them. The extracted features are then fed to a linear discriminant analysis based classifier for a multi-level hierarchical classification of audio signals. [2]

5 EQUATIONS

The sound wave under consideration is filtered with a Band Pass Filter of bandwidth 80 Hz-8000 Hz.

Program code:

```
[b,a]= butter(4, [80/22050 8000/22050]);
x=filter(b,a,y);
```

Where

y= wavread('sample.wav')

22050=sampling frequency

Silence Removal Silence present before and after the voiced part is removed to improve the performance of classifier. FILTERING THE SIGNAL The sound wave under consideration is filtered using a Band Pass Filter of bandwidth (80hz-8000hz) to eliminate noise.

VECTOR QUANTIZATION

Vector quantization (VQ) is the process of taking a large set of feature vectors and producing a smaller set of feature vectors that represent the centroids of the distribution, i.e. points spaced so as to minimize the average distance to every other point. We use vector quantization since it would be impractical to store every single feature vector that we generate from the training utterance. While the VQ algorithm does take a while to compute, it saves time during the testing phase, and therefore is a compromise that we can live with. Here, K-means clustering method is used.

6 MEL FREQUENCY SPECTRAL COFFICIENTS (MFCC)

6.1 Figures and Tables

The Mel-frequency cepstral coefficients (MFCCs) are frequently used as a speech parameterization in speech recognizers. Practical applications of speech recognition and dialogue systems bring sometimes a requirement to synthesize or reconstruct the speech from the saved or transmitted MFCCs. [7]

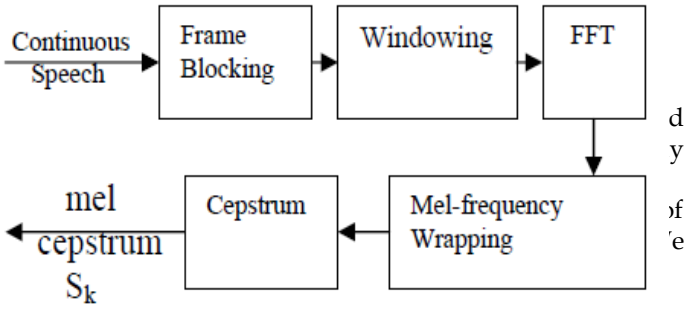

Fig : 6

The speech input is recorded at a sampling rate of 22050Hz. This sampling frequency is chosen to minimize the effects of ***aliasing*** in the analog-to-digital conversion process. In this work, the Mel frequency Cepstrum Coefficient (MFCC) feature has been used for designing a text dependent speaker identification system. The extracted speech features (MFCC's) of a speaker are quantized to a number of centroids using vector quantization algorithm. These centroids constitute the codebook of that speaker. MFCC's are calculated in training phase and again in testing phase. Speakers uttered same words once in a training session and once in a testing session later. The Euclidean distance between the MFCC's of each speaker in training phase to the centroids of individual speaker in testing phase is measured and the speaker is identified according to the minimum Euclidean distance. The code is developed in the MATLAB environment and performs the identification satisfactorily. [2] Speaker recognition is a generic term used for two related problems: speaker identification and verification. In the identification task the goal is to recognize the unknown speaker from a set of N known speakers. In verification, an identity claim (e.g., a username) is given to the recognizer and the goal is to accept or reject the given identity claim. In this work we concentrate on the identification task. The input of a speaker identification system is a sampled speech data, and the output is the index of the identified speaker. There are three important components in a speaker recognition system: the feature extraction component, the speaker models and the matching algorithm. [9]

6.2 Speaker Verification System

A speaker verification system is composed of two distinct phases, a training phase and a test phase. Each of them can be seen as a succession of independent modules.

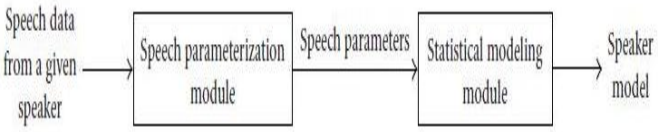

Figure 6.2 Modular Representation of the Training Phase of Speaker Verification System

Figure 6.2 shows a modular representation of the training phase of a speaker verification system. The first step consists in extracting parameters from the speech signal to obtain a representation suitable for statistical modeling as such models are extensively used in most state-of-the-art speaker verification systems. The second step consists in obtaining a statistical model from the parameters. This training scheme is also applied to the training of a background model.

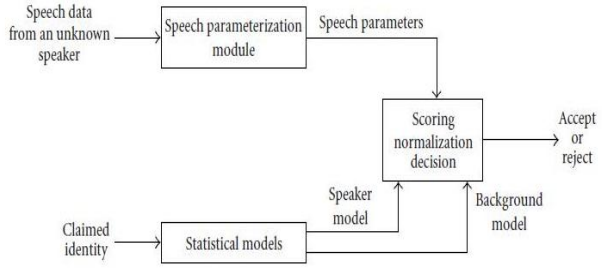

Figure 6.2 Modular Representation of the Test Phase of a Speaker Verification System

modular representation of the test phase of a speaker verification system. The entries of the system are a claimed identity and the speech samples pronounced by an unknown speaker. The purpose of a speaker verification system is to verify if the speech samples correspond to the claimed identity. First, speech parameters are extracted from the speech signal using exactly the same module as for the training phase. Then, the speaker model corresponding to the claimed identity and a background model are extracted from the set of statistical models calculated during the training phase. Finally, using the speech parameters extracted and the two statistical models, the last module computes some scores, normalizes them, and makes an acceptance or a rejection decision. The normalization step requires some score distributions to be estimated during the training phase or/and the test phase. [10]

6.3 TEXT-DEPENDENT VOICE RECOGNITION

The speech-dependent recognition techniques discriminate the users based on the same spoken utterance. Text-dependent recognition methods are usually based on template-matching techniques. Many of them use Dynamic time warping (DTW) algorithms or Hidden Markov models (HMM). Let us consider a sequence of same-speech input vocal utterances to be recognized: {S_1 , ... , S_n }. The feature extraction process is then applied to them, the feature set {V(S_1) ,...,V(S_n) } being obtained. Speaker classification represents the next stage of this pattern recognition process. We use a supervised classifier for our voice identification system, proposing a minimum mean distance classification approach. A set of registered (advised) speakers is set first. Next, a training set is obtained as a collection of spoken utterances, corresponding to the same speech, provided by these speakers and filtered for noise removal. Each speech signal of the training set constitutes a vocal prototype. The feature vectors computed for these prototypes make the feature training set.

6.4 TEXT-INDEPENDENT VOICE RECOGNITION

The speech-independent recognition systems involve impressing volumes of training data ensuring that the entire vocal range is captured. Thus, it is useful for not cooperative subjects, for example like those in the surveillance systems. The most successful speech-independent recognition methods are based on Vector Quantization (VQ) or Gaussian Mixture Model (GMM). The VQ-based methods are parametric approaches which use VQ codebooks consisting of a small number of representative feature vectors, while the GMM-based methods represent non-parametric techniques using K Gaussian distributions. We utilize the same delta mel cepstral analysis for the feature extraction part of this recognition system. [16]

7 End Sections

7.1 SCOPE OF FUTURE WORK

In this proposed work I worked on MFCC . The growth of speech recognition technology in the past five years is amazing. We have come from a market with high-priced products that relied on discrete dictation, or speaking in a r-o-b-o-t-i-c way, to a market where speech recognition technology is common both in the office and at home. People can speak in a natural voice to interact with their computers. This, combined with affordable pricing, and increased consumer demand, is leading to the evolution of transparent computing, where human/machine interaction is to natural that it is almost invisible. In addition to the telephone and mobile devices, speech recognition is making Web-based information accessible,

thanks to recent innovations such as Voice XML. Voice extensible markup language (Voice XML) will open enterprise applications for voice access, in the same way that HTML has enabled the development of graphical user interfaces. For example, Voice XML will let a person use a smart phone to access applications on the network by voice, touch, or key input as appropriate, see the results on the smart phone's display, and hear the results read back. As convenient as a desktop browser, Voice XML lets multiple devices access a company's information because data access is controlled by voice and accessed from a single point.

7.2 References

Human Voice recognition has been an interesting research field for the last decades, which still yields a number of unsolved problems. This paper aims to present a speaker recognition system which recognizes the speaker as opposed to what is being said by the speaker as in case of speech recognition. The methodology followed in this paper for Speaker recognition is using Feature Extraction process and then Vector Quantization of extracted features is done. At last the speaker is identified by comparing the data from a tested speaker to the codebook of each speaker and then measuring the difference. [14] Speech processing is emerged as one of the important application area of digital signal processing. Various fields for research in speech processing are speech recognition, speaker recognition, speech synthesis, speech coding etc. The objective of automatic speaker recognition is to extract, characterize and recognize the information about speaker identity. Feature extraction is the first step for speaker recognition.[1] overview of automatic speaker recognition technology, with an emphasis on text-independent recognition. Speaker recognition has been studied actively for several decades. We give an overview of both the classical and the state-of-the-art methods. We start with the fundamentals of automatic speaker recognition, concerning feature extraction and speaker modeling. We elaborate advanced computational techniques to address robustness and session variability.[2] the effectiveness of combinations of cepstral features, channel compensation techniques, and different local distances in the Dynamic Time Warping (DTW) algorithm is experimentally evaluated in the text-dependent speaker identification task.[4]

4 Conclusion

In this proposed work the methods used Speaker Recognition based on their results are as follows. The methods used for Speaker Recognition are classified into two parts

Training Phase

Testing Phase

This thesis project describes an enhanced Mel frequency Cepstral coefficient (MFCC) Technique for speaker recognition & analysis by the Computer .The computer records the voice pattern of the speaker during training phase. During the testing phase ,a speaker speaks into the microphone, and the computer analyses it by using Mat Lab software Two pattern matching algorithm are used in recognition mode.The algorirhm can be used in security devices that uses voice recognition technology for identification. Most important parts of speaker recognition system are (i) Feature Extraction

(ii) Classification method & Feature matching

The aim of the feature extraction step is to strip unnecessary Information from the sensor data and convert the properties of the signal which are important for the pattern recognition task to a format that simplifies the distinction of the classes. The goal of the classification step is to estimate the general extension of the classes within feature space from a training set. According to Martens there arevarious feature extraction techniques including Linear Predictive Coding (LPC), Perceptual Linear Prediction (PLP) and MFCC, However MFCC is frequently used method for Speaker Recognition & Speaker Verification. In MFCC, the main advantage is that it uses mel-frequency scaling which is Very approximate to the human auditory system.

References

[1]. Vibha Tiwari et.al :" MFCC and Its Applications in Speaker Recognition IJET pp 19-22, 2010.

[2] Tomi Kinnunen, Haizhou Li et.al : "An Overview of Text-Independent Speaker Recognition: from Features to Supervectors" visa_gets_behind_voice_recognition.html Preprint submitted to Speech Communication July 1, 2009.

[4] Atanas Ouzounov et.al.:, " Cepstral Features and Text-Dependent Speaker Identification-A Comparative Study" CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **10**, No 1 Sofia • 2010.

[5] . lindasaiwa Muda, Mumtaj Begam and I. Elamvazuthi "Voice Recognition Algorithms Using Mel Frequency Cepstral Coefficient (Mfcc) and Dynamic Time Warping (Dtw) Techniques Journal of Computing Vol 2, Issue 3 pp 138-143, 2010.

[6]. Md. Rashidul Hasan,Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman et .al . "Speaker Identification Using Mel Frequency Cepstral Coefficients" ICECE pp 565-568, 2004.

[7]]. Comparison of Clustering Algorithms in Speaker Identification Tomi Kinnunen, Teemu Kilpeläinen and Pasi Fräntic.

[8]. A Tutorial on Text-Independent Speaker Verification Fr ´ed´ eric Bimbot, Jean- Franc¸ois Bonastre, Corinne Fredouille, And others Journal on Applied Signal Processing pp 430-451, 2004

[9]. Speaker Recognition Project Report speaker recognition.googlecode.com/files/Finally_version1.2007.

[10]. Text Independent Speaker Recognition Using the Mel Frequency Cepstral Coefficients and A Neural Network Classifier Hassen Seddik AmelRahmouni and Mounir Sayadi IEEE pp 631-634, 2004.

[11]]. A Toolbox For Ann Learnin Poliana Magalhães Reis and Carlos Alberto Ynoguti IEEE pp 130-133, 2005.

[12]. Using Genetic Algorithm to Improve the Performance of Speech Recognition Based on Artificial Neural Network Min-Lun Lan , Shing-Tai Pan , Chih-Chin Lai Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06) 0-7695-2616-0/06, 2006.

[13]. Comparison of Text-Dependent Speaker Identification Methods for Short Distance Telephone Lines Using Artificial Neural Networks Ganesh K Venayagamoorthy and Narend Sundepersadh IEEE pp 253-258, 2000.

[14]. Comparing Various Voice Recognition Techniques Tudor barbu www.sped2009.ro/lucrari/04_Barbu_SpeD...

[15]. Speaker Identification using MFCC-Domain Support Vector Machine S. M. Kamruzzaman, A. N. M. Rezaul Karim, Md. Saiful Islam and Md. Emdadul Haque.