Automatic Speaker Recognition using LPCC and MFCC

Mr. P. Kumar¹, Dr. S. L. Lahudkar²

¹ Department of electronics and Tele-communication & University of Pune
JSPI’s ICoER, Wagholi, Pune, Maharashtra, India
² Department of Electronics and Tele-Communication & University of Pune
JSPI’s ICoER, Wagholi, Pune, Maharashtra, India

Abstract—A person’s voice contains various parameters that convey information such as emotion, gender, attitude, health and identity. This report talks about speaker recognition which deals with the subject of identifying a person based on their unique voiceprint present in their speech data. Pre-processing of the speech signal is performed before voice feature extraction. This process ensures the voice feature extraction contains accurate information that conveys the identity of the speaker. Voice feature extraction methods such as Linear Predictive Coding (LPC), Linear Predictive Cepstral Coefficients (LPCC) and Mel-Frequency Cepstral Coefficients (MFCC) are analysed and evaluated for their suitability for use in speaker recognition tasks. A new method which combined LPCC and MFCC (LPCC+MFCC) using fusion output was proposed and evaluated together with the different voice feature extraction methods. The speaker model for all the methods was computed using Vector Quantization- Linde, Buzo and Gray (VQ-LBG) method. Individual modelling and comparison for LPCC and MFCC is used for the LPCC+MFCC method. The similarity scores for both methods are then combined for identification decision. The results show that this method is better or at least comparable to the traditional methods such as LPCC and MFCC.

Keywords—LPC, MFCC, VQ, LPCC, ASR.

I. INTRODUCTION

In daily life, there is a need for controlled access to certain information/places for security. Typically such secure identification system requires a person to use a card system (something that the user has) or pin system (something that the user knows) in to gain access to the system. However, the two methods mentioned above have some shortcomings as the access control used can be stolen, lost, misused.

The desire for a more secure identification system (whereby the physical human self is the key to access the system) which leads to the research in the of biometric recognition systems. There are two main properties of biometric features. Behavioural characteristics such as voice, signature are the result of body part movements.

In the case of voice it merely shows the physical properties of the voice production organs. The articulatory process and the subsequent speech produced are never exactly same even when the same person utters the same sentence. Physiological characteristics refer to the actual physical properties of a person such as fingerprint, iris and hand geometry measurement.

Some of the possible applications of biometric systems include user-interface customisation and access control such as airport check in, building access control, telephone banking or remote credit card purchases. Speech technology offers many possibilities for personal identification that is natural and non-intrusive. Besides that, speech technology offers the capability to verify the identity of a person remotely over long distance by using a normal telephone.

A conversation between people contains a lot of information besides just the communication of ideas. Speech also conveys information such as gender, emotion, attitude, health situation and identity of a speaker. The topic of this thesis deals with speaker recognition that refers to the task of recognising people by their voices.

II. LITERATURE REVIEW

The captivation with employing voice for the many purposes in daily life has driven engineers and scientist to conduct massive amount of research and development in this field. The idea of an “Automatic speaker recognition” (ASR) which aims to build a machine that can identify a person by recognizing voice characteristics or features that are unique to each person.

The performance of modern recognition systems has improved significantly due to the various improvements of the algorithm and techniques involved in this field. As of this moment, ASR is still a great interest to researchers and engineers worldwide and the efficiency level of ASR is still improving. This is to highlight some of the important techniques, algorithm and research that are relevant to this report.

This is to highlight some of the important techniques, algorithm and research that are relevant to this report. Various types of typical pre-processing techniques, feature extraction and speaker modelling techniques will be covered in this report.
An overview of the advantages and typical applications of the techniques and algorithm in the speaker recognition system will be provided. Lastly, an overview of the comparison of the speaker recognition systems using algorithms and techniques that are explained in this report will be presented at the end of the Section II.

A. Linear Predictive Coefficients

The first method to be evaluated is the LPC derived voice features. LPC is seldom used by itself for speaker recognition in modern day ASR but in this project it will serve as a basis for comparison for the other methods.

There is an effect of varying the order of LPC. It is observed that LPC using 8 coefficients has a better recognition rate than other LPC coefficients for codebook of size 32. The results however are not unexpected. The two most significant factors that affect the recognition results are the quality of the speech signal together with the size of the codebook. Increasing the size of the codebook and LPC coefficients increases the effect of noise on the signal, as the signal will contain more information where noise can be present.

The results obtained for LPC using codebook size of 64 are pretty much similar to those using codebook sizes of 32. The recognition rate decreases from 66.67%, to hovering around 40% to 53.33%, as the number of coefficients used increases.

B. Linear Predictive Cepstral Coefficients

The second method to be evaluated is the LPCC derived voice features. LPCC is computed from LPC and is one of the most popular used for speaker recognition in modern day ASR.

There is an effect in varying the order of LPCC. It is observed that the recognition rate increases when the order of LPCC increases using codebook of size 32. The recognition rate increases from 73.33% (LPCC8) to 93.33% (LPCC12 & LPCC16) and drops to 86.67% (LPCC20). From that finding, we can see that the recognition rate does not increase all the time just by increasing the order of the LPCC. In fact, LPCC experiences a drop in the recognition rate when higher order coefficients are used. This tally with the study by Reynolds [23] where the LPCC recognition rate averages at 90% and drops when higher order coefficients are used. (Recognition rates maintained for higher orders. (>12)).

III. DEVELOPMENT OF SPEAKER RECOGNITION SYSTEM

All The main purpose of the prototype system is to compare the recognition rate in order to determine the suitability of the different types of features to be use in a speaker recognition system. This section will describe in detail the techniques used for the pre-processing and voice feature extraction stages.

![Algorithm for speaker recognition.](image)
III. LPCC+MFCC

Based on the above results retrieved for the different voice features, this method aims to combine the two features LPCC and MFCC to achieve better recognition rate by considering supplementary information sources. This is accomplished by using output fusion that model individual data separately and combining them at the output to give the overall matching score. The figure below shows the structure of the proposed system.

**Figure**: Block diagram of proposed

The speech signal of the unknown speaker will be processed individually using LPCC (16th order, codebook size 32) and MFCC (16th order, codebook size 32) and compared with the corresponding codebooks of the known speaker database. The choice of codebook size and order used are based on the following reason:

1. The size of the codebook determines the complexity of the computation and based on the results achieved, Codebook size of 32 managed to achieve 93.33% for both LPCC16 and MFCC16.
2. The extra computational time required for implementing such a system is negligible as compared to other methods when running in typical home PC setup using Pentium Core2 dual.

The corresponding matching scores that indicate the degree of similarity between the users will be generated and combined. The reason for this is due to the fact that the results for the show that LPCC and MFCC have equal recognition rates. The user with the lowest score (highest degree of similarity) for the combined scores will be returned as the identity of the unknown speaker.

IV. CONCLUSIONS

This paper has presented the analysis for voiceprint analysis for speaker recognition. Various pre-processing stages prior to feature extraction were studied and implemented for the prototype ASR. The prototype was developed to analyse and evaluate various voice feature extraction methods such as LPC, LPCC and MFCC for their suitability in ASR. In addition, a new method (LPCC16+MFCC16) was proposed to enhance the recognition rate of ASR by using fusion output.

The results obtained have shown that LPCC and MFCC perform relatively well in speaker recognition tasks. LPCC using an order of 16 with codebook size of 128 achieved the best recognition rate of 100%. However, utilizing a codebook of 128 requires much computational processes that affect the performance of the system. LPCC also performs poorly when insufficient order is used. MFCC is more consistent than the LPCC in performing recognition task as it is less susceptible to noise and due to the fact that it is modelled after the human perception of sound.

An evaluation of the performance of the fusion method using LPCC16 and MFCC16 achieved 84% accuracy using a group of 20+ speakers. The result indicates that by using multiple features sets, it is possible to achieve high recognition rate using smaller size codebooks.

Acknowledgment

I ‘Pappu Kumar’, P.G Student at JSPM’s ICOER, Dept. of electronics and tele-communication co-ordinally thanks to my guide Dr. S. L. Lahudkar who encouraged me at each and every state. I also thanks to my friends to work in the field of speech processing.

REFERENCES


Atal, B. "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification.”. J. Acoust. Soc. Am. 55 (6), 1304-1312.


www.mathworks.com/exchange files