Automatic Speaker Recognition using LPCC and MFCC

Mr. P, Kumar[#1], Dr. S. L. Lahudkar[*2]

[#] *Department of electronics and Tele-communication & University of Pune*
JSPM's ICOER, Wagholi, Pune, Maharashtra, India

[*] *Department of Electronics and Tele-Communication & University of Pune*
JSPM's ICOER, Wagholi, Pune, Maharashtra, India

[1]*pappukumar1991@gmail.com*
[2]*swapnillahudkar@gmail.com*

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 LPCC is use.

C. Mel-Frequency Cepstral Coefficients

The third method to be evaluated is the MFCC derived voice features. MFCC are coefficients that represent sound based on human perception.. MFCC are derived by taking the Fourier Transform of the signal, warping it to by using a Mel-filter bank that closely mimic the Mel-scale, the final step is to perform Discrete Cosine Transform on the logarithm power of the speech frame from the Mel-scale output. From figure 2.4, the effect of varying the order of the MFCC does not seemed to have much effect on the recognition rate. The recognition rate increases from 80% (MFCC8) and stays stagnant at 93.33% (MFCC12, MFCC16 and MFCC20). The results of MFCC using codebook size of 32 shows that MFCC function better than LPCC and LPC when using smaller size codebooks. This might be due to the MFCC being more immune to noise that affects the LPC and LPCC. The results obtained from figure 5.3 shows a recognition rate of 93.33% across all the orders of the MFCC used. From the results, varying the orders of the MFCC does not show any effect of increasing or decreasing the recognition rate. The recognition rate peaks at the 93.33%. Overall, the MFCC recognition rate is better when compared to the LPC and LPCC. The findings are consistent with MFCC being known to be more robust to noise and spectral estimation errors 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.

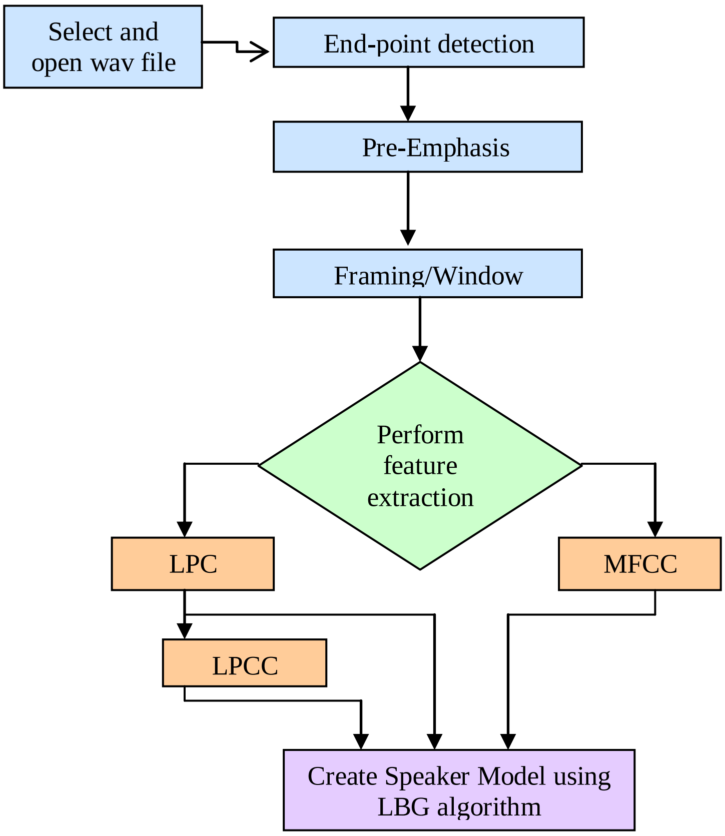

Figure: Algorithm for speaker recognition.

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.

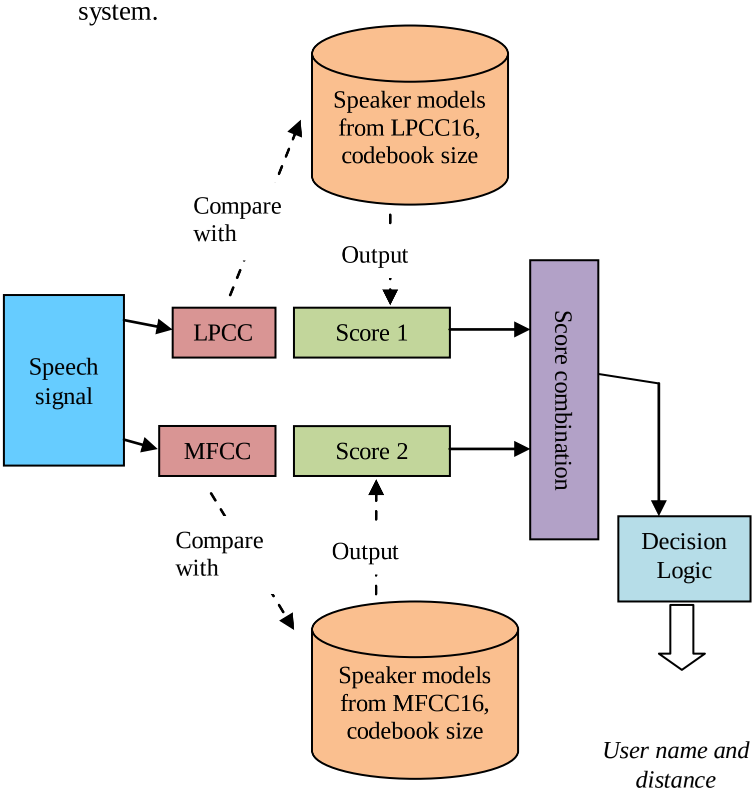

Figure: Block diagram of proposed

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

[1] **Reynolds, D.A.,.** "An overview of automatic speaker recognition technology,". Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on , vol.4, no., pp. IV-4072-IV-4075 vol.4, 2002.

[2] **Ayaz Keerio, Bhargav Kumar Mitra, Philip Birch, Rupert Young, and Chris Chatwin.** "On Preprocessing of Speech Signals". *International Journal of Signal Processing ; Vol.5 No.3 2009 [Page 216].*

[3] **Campbell, J. P.**"*Speaker Recognition*",. 1999. Technical report, Department of Defence,Fort Meade..

[4] **Al-Akaidi, Marwan.** "Introduction to speech processing". *Fractal Speech Processing.* s.l. : Cambridge University Press The Edinburgh Building, Cambridge CB2 2RU, UK, 2004.

[5] **Saha. G., Chakroborty. S., and Senapati. S,.** “A new Silence Removal and End Point Detection Algorithm for Speech and Speaker Recognition Applications” . *in Proc. of Eleventh*

National Conference on Communications (NCC), IITKharagpur, India, January 28-30, 20.

[6] **Kinghorn, M. Greenwood and A.***"Suving: Automaticsilence/unvoiced/voiced classification of speech,".* Departmentof Computer Science, The University ofSheffield, 1999.

[7] **Long, Hai-Nan and Cui-Gai Zhang.** "An improved method for robust speech endpoint detection," . Machine Learning and Cybernetics, 2009 International Conference on , vol.4, no., pp.2067-2071, 12-15 July 2009.

[8] **Liu, Li, He, Jialong and Palm, G.,.** "Signal modeling for speaker identification,". Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on , vol.2, no., pp.665-668 vol. 2, 7-10 May 1996.

[9] **Picone, J.W.,.** "Signal modeling techniques in speech recognition," . *Proceedings of the IEEE , vol.81, no.9, pp.1215-1247, Sep 1993.*

[10] **Jayanna H S, Mahadeva Prasanna S R.** "Analysis, Feature Extraction, Modeling and Testing Techniques for Speaker Recognition". *IETE Tech Rev 2009;26:181-90.*

[11] **Schroeder, M.,.** "Linear prediction, entropy and signal analysis," . *ASSP Magazine, IEEE , vol.1, no.3, pp. 3-11, Jul 1984.*

[12] **Schroeder, M.,..** "Linear predictive coding of speech: Review and current directions,". *Communications Magazine, IEEE , vol.23, no.8, pp. 54-61, Aug 1985.*

[13] **Kwong, S. and Nui, P.T.,.** "Design and implementation of a parametric speech coder,". *Consumer Electronics, IEEE Transactions on , vol.44, no.1, pp.163-169, Feb 1998.*

[14] **Gupta, V., Bryan, J. and Gowdy, J.,.** "A speaker-independent speech-recognition system based on linear prediction,". *Acoustics, Speech and Signal Processing, IEEE Transactions on , vol.26, no.1, pp. 27-33, Feb 1978.*

[15] **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.*

[16] **Wong, E. and Sridharan, S.** "Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification,. *"Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on, vol., no., pp.95-98, 2001.* 2001.

[17] **Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani, Md. Saifur Rahman.** "Speaker Identification using Mel Frequency cepstral coefficients". *3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh.*

[18] **Vergin, R.,.** "An algorithm for robust signal modelling in speech recognition," . Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on , vol.2, no., pp.969-972 vol.2, 12-15 May 1998.

[19] **Vergin, R., O'Shaughnessy, D. and Farhat, A.,.** "Generalized mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition," . *Speech and Audio Processing, IEEE Transactions on , vol.7, no.5, pp.525-532, Sep 1999.*

[20] **Buchanan, C.R.** "Informatics Reseach Proposal - Modeling the Semantics of sound".2005.

[21] **Seddik, H., Rahmouni, A. and Sayadi, M.,.**"Text independent speaker recognition using the Mel frequency cepstral coefficients and a neural network classifier"., *Communications and Signal Processing, 2004. First International Symposium on , vol., no., pp. 631-634, 2004.* 2004.

[22] **Molau, S., et al.** "Computing Mel-frequency cepstral coefficients on the power spectrum," . Acoustics, Speech, and Signal Processing, 2001.Proceedings. (ICASSP '01). 2001 IEEE International Conference on , vol.1, no., pp.73-76 vol.

[23] **Reynolds, D.A.,.**"Experimental evaluation of features for robust speaker identification,". *Speech and Audio Processing, IEEE Transactions on , vol.2, no.4, pp.639-643, Oct 1994.*

[24] **Zhonghua, Fu and Zhao Rongchun.** "An overview of modeling technology of speaker recognition," .Neural Networks and Signal Processing, 2003. Proceedings of the 2003International Conference on , vol.2, no., pp. 887-891 Vol.2, 14-17 Dec. 2003.

[25] **Y. Linde, A. Buzo, and R.M. Gray,.**"An algorithm for vector quantizer design,".*IEEE Trans. Communications, vol. COM-28(1), pp. 84-96, Jan. 1980.*

[26] www.mathworks.com/exchange files