Original Article

Formant Frequency Estimation using Computation auditory Scene method for Speaker Identification and Recognition

Aabha Sahu
Assistant Professor, Department of Electrical Engineering, Kalinga University Naya Raipur C.G.
Email: aabha.sahu@kalingauniversity.ac.in
DOI: 10.47750/pnr.2023.14.S02.228

Abstract

Speaker recognition is the process which is used to identify a particular person. A human voice is one of the medium using which any person can be recognized. Human voice has many characteristics which may vary for individuals. Speaker identification or recognition can been done by analyzing individual voice signal hence it is commonly referred as speech recognition. However; voice signal is one of the most common factors which can be copied easily. Copying the voice signal of one person is known as mimic and many mimicry artists can mimic many famous personalities. Mimicry is done for the entertainment purpose and it is popular worldwide. However its dark side may raise various unethical activities. Hence speech recognition is a sensitive matter which gives a broad area to the researchers so that authenticity may be maintained. Many research have been done in this field however existing techniques does not have 100% efficiency and many more research can be done to achieve better results. In this paper formant frequency estimation is done using computation auditory scene technique for the purpose of speaker identification and recognition. The result shows that proposed technique performs better in comparison with the other existing techniques.

INTRODUCTION

In the last few decades many automatic speech and speaker recognition techniques have been introduces which are used to recognize the speaker by using their sample voice signals[1-5]. These techniques can be done in two different domain viz. time or spatial domain and frequency or spectral domain[6-7]. Each method has its own advantages but simultaneously some limitations which restricts the performance of individual methods[8]. Hence it can be say that there is still has a scope to get better technique for the same purpose. Each technique has its two steps. In the first step feature is to be extracted from the sample signal while in the second step features of the signal would be compared[9]. To understand the progress in this field many papers have been studied and found some limitations of the existing techniques[10-14]. In the next section; methodology part have been discussed which illustrates the working of proposed method.

Methodology:

In this part we will discuss computational auditory scene analysis (CASA) which is used to extract the information from data set of human voice signal[13,15-18]. As aforementioned; speech analysis is to be done in two stages:

1. Feature extraction: In this step various feature is to be extracted using various spatial and/or spectral analysis[19-21]. Fig. show the simplified block diagram of feature extraction step. Feature extraction can be done using various methods e. g. hidden Markov Model (HMM), support Vector Mavhines (SVM), Template Matching (TM) etc.

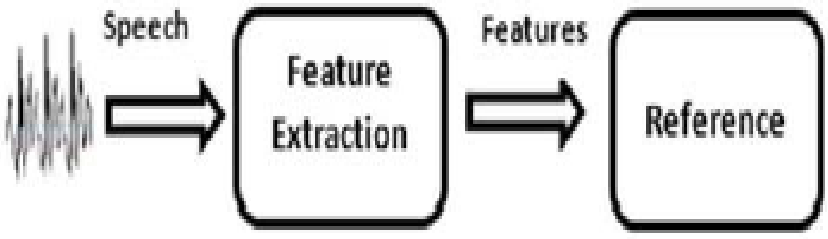

Fig.1: Feature extraction of speech sample

2. Speech analysis: In this technique various features extracted from the sampled signals are to be compared with the authentic person's speech so that speaker can be recognized [2-5, 20-27]. Sppech analysis done in the spectral domain gives better results as compare to that of spatial domain. Wavelet transform based technique gives better results.

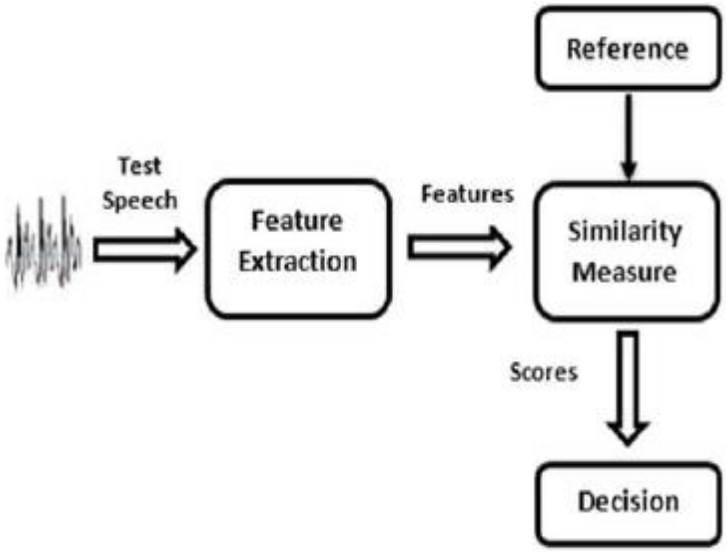

Fig.2: Comparison of various measures to identify authentic speaker

Based on the above two steps proposed system has been introduced. Fig. 3 shows a simplified block diagram of CASA system which uses peripheral for the analysis of acoustic input signal which is the mixer of various audible signals[28-30]. This peripheral converts the given mix signal into its time frequency representation which is done by an auditory filter bank[31,32]. The frequency range of this filter bank ranges from 50 Hz to 8KHz. The next stage is used for the feature extraction.

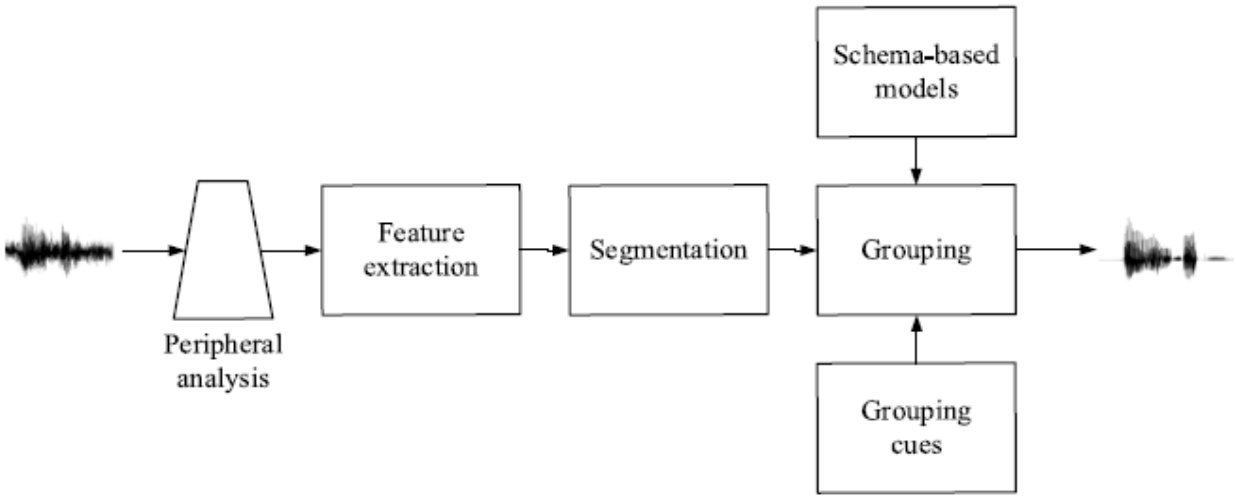

Fig. 3: A typical computational auditory scene analysis system

Results and Discussion:

To know the efficacy of the proposed method various tests have been performed. For this purpose three different noise signals have been considered viz. factory noise, babble noise and bottle noise.

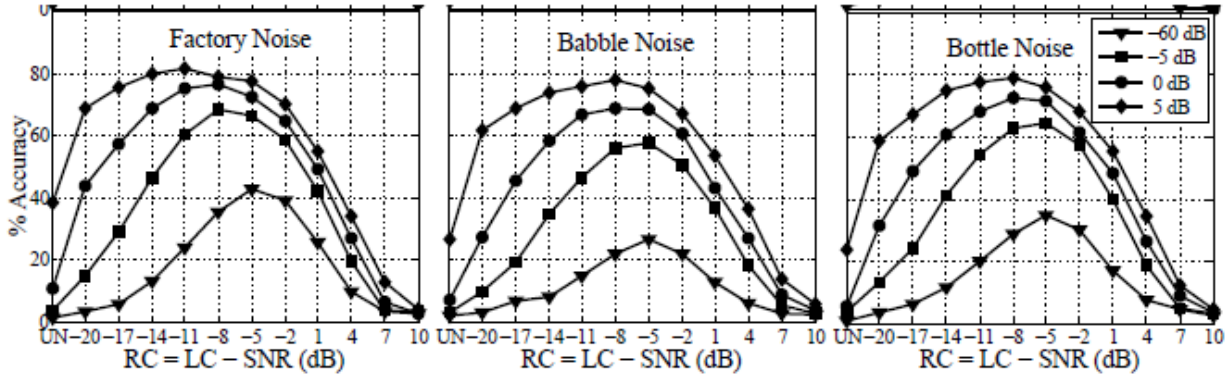

Fig.4: Accuracy v/s SNR for different level of noise singals

Conclusion:

From the results; it is clear that the proposed technique for the speaker recognition gives satisfactory results as compared to the other existing techniques. This method is the integration of acoustic model with enhanced feature which improves the efficacy of the proposed technique. Hence we can say that the proposed technique covers various aspects of speech separation and provides automatic speech recognition system.

REFERENCES

1. V. Sharma. A Review On Speaker Recognition Approaches And Challenges. International Journal of Engineering Research and Technology (IJERT),

2013.
2. H. H. Lei. Structured approaches to data selection for speaker recognition.. Thèse de doctorat. UC Berkeley, 2010.
3. C., Ruchi. Short-term spectral feature extraction and their fusion in text independent speaker recognition: A review. BVICA M's International Journal of Information Technology, vol. 5, no 2, pp. 630-639, 2013.
4. D., Mengistu. Automatic text independent amharic language speaker recognition in noisy environment using hybrid approaches of LPCC, MFCC and GFCC. International Journal of Advanced Studies in Computers, Science and Engineering, vol. 6, no 5, p. 8, 2017.
5. D., Mengistu, & D. M. Alemayehu. Text independent amharic language speaker identification in noisy environments using speech processing techniques. Indonesian Journal of Electrical Engineering and Computer Science, vol. 5, no 1, pp. 109-114, 2017.
6. F. Z., Chelali, & A., Djeradi. Text dependant speaker recognition using MFCC, LPC and DWT. International Journal of Speech Technology, vol. 20, no 3, pp. 725-740, 2017.
7. X., Zhou, D., Garcia-Romero, R., Duraiswami, C., Espy-Wilson, & S. Shamma. Linear versus mel frequency cepstral coefficients for speaker recognition. In : 2011 IEEE Workshop on Automatic Speech Recognition & Understanding. IEEE, pp. 559-564, 2011.
8. P., Kenny, T., Stafylakis, P., Ouellet, V., Gupta, & M. J. Alam. Deep Neural Networks for extracting Baum-Welch statistics for Speaker Recognition. In: Odyssey, pp. 293-298, 2014.
9. V. Vasilakakis, S. Cumani & P. Laface, Speaker recognition by means of Deep Belief Networks, Biometric Technologies in Forensic Science, Nijmegen, 14-15 October 2013.
10. T. Stafylakis, P. Kenny, M. Senoussaoui and P. Dumouchel, Preliminary investigation of Boltzmann machine classifiers for speaker recognition, in Proceedings of Odyssey Speaker and Language Recognition Workshop, 2012.
11. V., Gupta, P., Kenny, P., Ouellet, & T. Stafylakis. I-vector-based speaker adaptation of deep neural networks for french broadcast audio transcription. In: international conference on acoustics, speech and signal processing (ICASSP). pp. 6334-6338, IEEE, 2014.
12. V., Srinivas, & T. Madhu. Neural network based classification for speaker identification. International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 7, no 1, pp. 109-120, 2014.
13. T., Justin, V., Štruc, S., Dobrišek, B., Vesnicer, I., Ipši, & F. Miheli speaker de-identification using diphone recognition and speech synthesis. In: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1-7, IEEE, 2015.
14. S., Pandiaraj, & K. R. Shankar Kumar. Speaker identification using discrete wavelet transform. Journal of Computer Science, vol. 11, no 1, pp. 53–56, 2015. doi:10.3844/jcssp.2015.53.56.
15. M. I., Abdalla, H. A., Abobakr, & T. S. Gaafar. DWT and MFCCs based feature extraction methods for isolated word recognition. International Journal of Computer Applications, vol. 69, no 20, 2013.
16. V., SHARMA. A Review On Speaker Recognition Approaches And Challenges. International Journal of Engineering Research and Technology (IJERT), 2013.
17. Y. Lei, N. Scheffer, L. Ferrer, and M. McLaren, A novel scheme for speaker recognition using a phonetically aware deep neural network. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1695-1699, IEEE, 2014.
18. L. Ferrer, Y. Lei, and M. McLaren. Study of senone-based deep neural network approaches for spoken language recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no 1, p. 105-116, 2015.
19. L. Ferrer, Y. Lei, M. McLaren, and N. Scheffer. Language identification based on senone posteriors, In : Fifteenth Annual Conference of the International Speech Communication Association. 2014.
20. Y. Song, B. Jiang, Y. Bao, S. Wei, and L. Dai. i-Vector representation based on bottleneck features for language identification, Electronics Letters, vol. 49, no. 24, pp. 1569–1570, 2013.
21. P. Matejka, L. Zhang, T. Ng, S.H. Mallidi, O. Glembek, J. Ma, and B. Zhang. Neural network bottleneck features for language identification, in Proc. Speaker Odyssey, 2014.
22. Y. Lei, L. Ferrer, A. Lawson, M. McLaren, and N. Scheffer. Application of convolutional neural networks to language identification in noisy conditions, in Proc. Speaker Odyssey, 2014.
23. M. McLaren, Y. Lei, and L. Ferrer. Advances in deep neural network approaches to speaker recognition. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 4814-4818, IEEE, 2015.
24. M., McLaren, L., Ferrer, & A. Lawson. Exploring the role of phonetic bottleneck features for speaker and language recognition. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 5575-5579, 2016.
25. [56] A., Mahmood, M. Alsulaiman, & G. Muhammad. Automatic Speaker Recognition Using Multi-Directional Local Features (MDLF). Arab J Sci Eng vol 39, no 5, pp 3799–3811, 2014.
26. M., Alsulaiman, A., Mahmood, & G. Muhammad. Speaker recognition based on Arabic phonemes. Speech Communication, vol. 86, pP. 42- 51, 2017
27. I. Shahin, Text-Independent Emirati-Accented Speaker Identification in Emotional Talking Environment. In : 2018 Fifth HCT Information Technology Trends (ITT), pp. 257-262, IEEE, 2018.
28. P., Chen, S. Y., Zhang, C. T., Yeh, J. C., Wang, T., Wang, & C. L. Huang. Speaker Characterization Using TDNN-LSTM Based Speaker Embedding. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6211-6215. IEEE, 2019.
29. L., Eljawad, R., Aljamaeen, M., Alsmadi, et al. Arabic Voice Recognition Using Fuzzy Logic and Neural Network. pp. 651-662, 2019.
30. M., Hamidi, H., Satori, O., Zealouk, K., Satori, & N. Laaidi. Interactive Voice Response Server Voice Network Administration Using Hidden Markov Model Speech Recognition System. In: 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), pp. 16-21, IEEE, 2018.
31. F. Z. Chelali K. Sadeddine & A. Djeradi. Speaker identification system using LPC -Application on Berber language, HDSKD journal, Vol. 01, No. 02, pp. 29-46, December 2015.
32. F.Z. Chelali, & A. Djeradi. Text dependant speaker recognition using MFCC, LPC and DWT. International Journal of Speech Technology. Vol 20, no 3, pp 725–740, 2017