Available online at http://www.nrjp.co.in

Journal of Management Science, Operations & Strategies
Vol. 3, Issue, 03. 94-97p, Dec. 2019

National Research & Journal Publication

Review Article

Voice Recognition: A Novel Technical Step in Neural Network

Manjari Gangwar

Abstract

The proposed system first is trained using eight different sample of a person voice and then feature extraction is applied. Feature extraction is a process of extracting various different features of voice. Then feature matching is performed. In this method, we give the system the input of that person voice and ask system to grant access to another system after matching the input with the stored voices.
Key words: *feature, Study, method, matching, stored voices.*

INTRODUCTION

Speaker recognition is the identification of the person who is speaking by characteristics of their voices also called as voice recognition [1, 2]. Voice conversion is the process that automatically transforms the source speaker's voice to that of target speakers [3]. From the past 20 years various attempts had made for the voice conversion [3]. To get an acceptable quality of voice transformation researchers tried to transform only the filter features [4, 5]. But it proved the need for transformation of excitation features to attain an effective voice morphing system. For transforming spectral characteristics and excitation characteristics many voice conversions currently are using GMM or ANN [6, 7].

Different techniques are used such as mapping code books, artificial neural networks, dynamic frequency warping and Gaussian Mixture Model (GMM) [7]. In successive conversion of different voice conversion system various different techniques are used, but all these system share voice modeling and training components [3]. As we use GMM and ANN, but out of those GMM mapping technique is widely used but in this paper we are going to exploit mapping abilities of ANN [8]. GMM are like kernel density estimates, but with a small number of components (rather than one component per data point) [9, 10].

Various Methods Used For Voice Recognition Radial Basis Function

Radial basis function is a method used for voice recognition system. It includes the following features: It consists of two-layer feed-forward networks. RBF is implemented by the hidden layers. It has a faster training/ learning capacity. RBF networks are excellent for the purpose of interpolation.

Ease of Use

i. It has a faster convergence.
ii. Smaller extrapolation errors.
iii. It offers a very high reliability.
iv. It provides a better theoretical analysis.

Demerits

i. In the noisy environment, it performs badly.
ii. They are not efficient computationally.

MULTILAYER PERCEPTRON

Multilayer perceptron is another method for voice recognition system. It includes the following features: It consists of multiple layers of nodes in a directed graph, and each node is connected to the further one. Each node has a nonlinear activation function. It utilizes back propagation, which is a supervised learning technique for training the network.

Ease of Use

i. Generalization.
ii. Fault tolerance.
iii. It is less reliable as compared to radial basic function.
iv. The accuracy of Multilayer Perceptron is 93%.

Demerits

i. It does have the scaling problem.
ii. It does not produce the guaranteed solution.
iii. It is a very expensive process.

RELATED WORK

Year	Research Paper(IEEE)	Methods	Conclusion
1995	Text independent speaker recognition using Neural Techniques	Multilayer Perception(MLP), Radial Basic Function(RBF)	RBF is more efficient then MLP for both multi-speaker and single-speaker configuration
1997	Robust speech recognition technique using a RBF neural network for mobile app	RBF,noise reduction technique, Speech modeling, system evaluated by NOISEX-92	Real time noise robust realization is then acceptable using RBF based neural networks for VAD
2000	Speakerrecognition using ANN based vowel phonemes	ACW cepstrum card, LP cepstrum	ACW spectrum is more efficient then Linear Predictive (LP) cepstrum
2004	Speech Recognition and its application in based robot system	Linear predicted coefficient(LPC),Pattern comparison technique, Dynamic time warping(DTW)	The correctness of the recognition is more than 90% but the recognition in pc is not portable to improve flexibility recognition can be realized in DSP instead
2012	Duration modeling in voice conversion using artificial neural networks	Artificial neural network, DTW is used to align MCEP vector of source and target speakers	Segmental duration transform can be done in the baseline voice conversion
2012	Speaker identification using neural network on an FPGA	Programmable Array (FPGA),LPC,ANN	98% accuracy using FPGA which can be made 100% using mat lab
2013	Voice biometrics using linear Gaussian model	Linear Gaussian modeling method, PLDA model	It provides a more robust approach than PLDA and 2COV systems
2014	NavEye: Smart Guide for Blind Students	QR code and censors, JOSM (Java Open Street Map Editor), Eclipse	The System communicates with the user through voice command and it can read QR Code to identify the current location
2014	Speech Biometric Based Attendance System	Interactive Voice Response (IVR), MFCC features ,I vector based Speaker modeling	The system performance in terms of recognition rate is found to be 94.2 %and the average response time of the system for a test data of Duration 50 seconds is noted to be 26 seconds

GAUSSIAN MIXTURE MODEL

Gaussian mixture model contains the following features: GMM is a supervised learning classification algorithm. GMM is used for data clustering. It converges to a local optimum using an iterative algorithm. This method is also called as soft clustering method.

Ease of Use

i. It is very easy to use and important.
ii. For the non-temporal pattern recognition it is a good algorithm.
iii. The accuracy of GMM is 94.2%.

Demerits

i. For high level of problem it can fail to work.
ii. The user should set the number of mixture models which the algorithm will use.

FIELD PROGRAMMABLE GATE ARRAY

Field programmable gate array has the following main features: We use it for wide range of logic gate. To change the logic function they can be reconfigured.

Ease of Use

i. It consists of inexpensive logic design because it has short design cycle.
ii. It consists of powerful design, programming and syntheses tools.
iii. Accuracy of FPGA is 98%.

Demerits

i. For multi FPGA system routine is easily blocked.
ii. No natural hierarchical extension

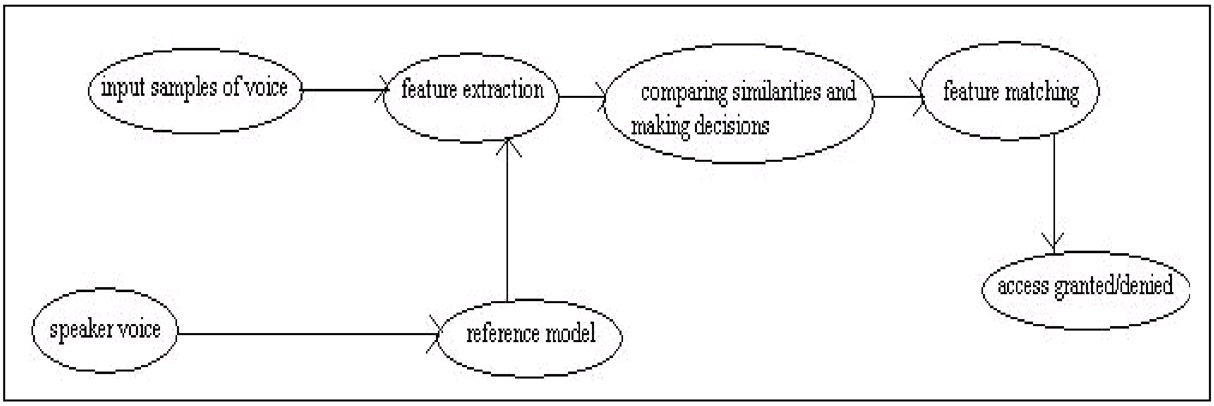

Fig. 1: *The Proposed System.*

PROPOSED SYSTEM

Our system is voice recognition system. This system consists of two modules, which include feature extraction and feature matching. It aims to identify as well as verify the speaker. Speaker identification means determining the speaker who is already registered and Speaker verification means to verify the registered speaker.

The proposed system first is trained using eight different sample of a person voice and then feature extraction is applied. Feature extraction is a process of extracting various different features of voice. Then feature matching is performed. In this method, we give the system the input of that person voice and ask system to grant access to another system after matching the input with the stored voices. The proposed system uses the Radial Basis Algorithm for implementation. This system is used as a login application for several other applications.

CONCLUSION

The different methods used for the voice recognition system are Radial Basis Function, Multilayer Perceptron, Gaussian Mixture Model, Field Programmable Array etc. Comparing all the methods the accuracy of Multilayer Perceptron is 93% and 94.2% accuracy is of Gaussian Mixture Model and Field Programmable Array has accuracy of 98%. On the basis of these methods much research has been done.

These research papers tell us which methods are more efficient. RBF is more efficient than MLP for the single as well as multi speaker. Using RBF Real time noise robust realization can be acceptable. 98% of accuracy of FPGA can be made 100% by using MATLAB.

REFERENCES

1. Luo Zhizeng, Zhao. *Speech, Recognition and Its Application in Voice-based Robot Control System Jinghing*. Robot research laboratory. Zhejiang Province, China: Hangzhou Institute of Electronic Engineering Hangzhou; 2010.
2. F-Romero, S. O. Caballero-Morales. *Speaker identification using Neural Networks on an FPGA*. División de Estudios de Posgrado.
3. Ronanki Srikanth, Bajibabu B, Kishore Prahallad. *Duration Modelling In Voice Conversion Using Artificial Neural Networks*. International Institute of Information Technology, Hyderabad, India. http://ravi.iiit.ac.in/~speech/publications/2012_Conf_P003.pdf
4. Hamza A. AlAbri, Ahmed M. AlWesti, Mohammed A. AlMaawali, *et al. NavEye: Smart Guide for Blind Students Qaboos University*.
5. Fredrickson, Tarassenk. *Text-Independent Speaker Recognition Using Neural Network 'Techniques*. Oxford University, UK.
6. Ravi Sankar, Netoo Singh Sethi. *Robust Speech Recognition Techniques Using a Radial Basis function Neural Network for Mobile Applications*, Department of Electrical Engineering. University of South Florida Tampa, Florida.
7. Hai Yang, Yunfei Xu, Houjun Huang, *et al*. *Voice biometrics using linear Gaussian model*. 2014
8. *Laboratory of Speech Acoustics and Content Understanding*, Chinese Academy of Sciences, 21 Beisihuan XiLu, Beijing.
9. Subhadeep Dey, Sujit Barman, Ramesh K. Bhukya, *et al*. *Speech Biometric Based Attendance System Department of Electronics and Electrical Engineering* Indian Institute of Technology, Guwahati, India.
10. Trujillo-Romero F, Caballero-Morales SO. *Speaker identification using Neural Networks on an FPGA*. División de Estudios de Posgrado. Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2012.