



Multi Lingual Speaker Identification on Foreign Languages Using Artificial Neural Network with Clustering

Gurpreet kaur,

Department of CSE
Lovely professional university
Jalandhar, Punjab,India

Harjeet kaur

Department of CSE
Lovely professional university
Jalandhar, Punjab,India

Abstract: *The Speech is most prominent and primary mode of communication among of human being. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer. This paper describes a method for text based speech identification system based on Artificial Neural Network that identifies speech in Indian as well as in Foreign Languages. Our database contain one sentence with different foreign languages i.e. French, Finnish, Catalan, Italian, Portuguese, Indonesian, English, Hindi spoken by different speakers in each languages by both male and female. To increase the performance mainly uses the “Fuzzy K-Mean Clustering” algorithm & the average performance of the system is 96%..*

Keywords : *NN- neural network, LPC- linear predictive coding, ANN- artificial neural network, ASI- Automatic speaker identification*

1. Introduction

Speech recognition is a system which recognizes human voice based on certain characteristics with differentiates one person from the other. Speech recognition has found application in access control mechanisms providing security and can also be applied to large number of applications such as speech driven consumer applications [6, 7]. Based on the information extracted from the speech signal, it can have three different recognition systems itself: Speaker Recognition, Language Recognition and Speech Text Recognition [1]. The whole process of speech recognition is as shown in fig1 Identification is further divided into two parts text based dependent identification and text based independent identification. If the same text is used in both training and recognition process the mode of recognition called dependent and if the different text is used for training and recognition the mode is called the independent identification. The speaker recognition is a process which identifies the person who is speaking by the characteristics of their voices and the speaker identification is process which matches the unknown voices. The speaker identification mostly used in the criminal investigation. The system thus developed can be put to use in access control security mechanism in non critical areas or answering FAQs in tourism agency where there is difficulty understanding queries of foreign people by local individuals [1, 4, 6, and 7]. Speech recognition encompasses a large number of complex applications such as speech driven consumer applications, speech commissioning in logistics, checking and recording in quality assurance work etc. Solving the problems of unauthorized use of computer and communications systems and multiple accesses control the use applications such as speaker identification and speaker verification [7].

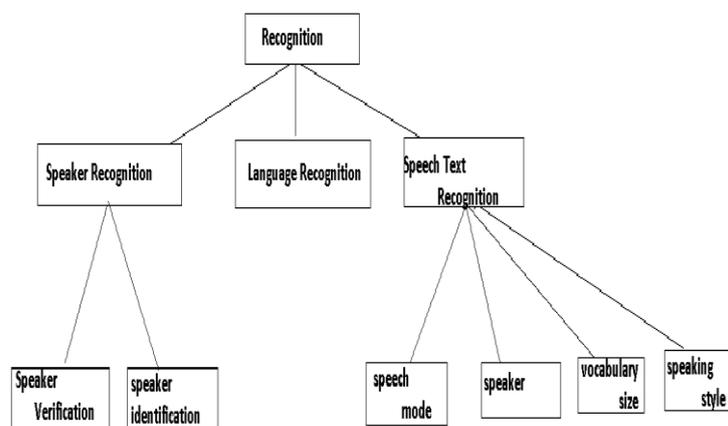


Fig1:Recognition process

1.1 Artificial neural networks

An artificial neural network (ANN) is an information processing system, which simulate the working of biological nervous systems, such as the brain, process information. It comprises of the huge number of highly interconnected processing elements to provide solution for specific problems. ANNs, just like humans do, learn by example. An ANN is configured for a specific applications, such as data classification or pattern recognition, with the help of a learning process .ANN are good in making generalized rules for difficult or complex problems and recognizing pattern but they are not intelligent. An excellent feature of ANNs is that they can be trained that is what makes it more likely to be used in area of research. ANNs are good at generalizing from a set of training data [13].

2. Previous Work

Speech recognition system is a system which recognizes human voices based on certain characteristics which differentiate one person from other. Mostly work is done in speech recognition field on different Indian languages with the help of number of speakers but there is no work done with foreign languages. Most automatic speaker recognition work was used small vocabulary or associated words and recent work is done on sentence with the help of Indian languages. Author [6] purposed the use of multilayer preceptron (MLP), which is trained using back -propagation technique to be the engine of an automated digit recognition system using voice [7]. Example for inputting credit card number, phone number etc. most of the practical applications are of the small vocabulary or isolated word and large vocabulary systems perform well in laboratories but not in real world. Mostly research is done with Indian languages no work has been carried out for speaker recognition with multiple foreign languages [7]. The research has been done for multilingual speaker identification system is by using statistical methods like Hidden Markov Model (HMMs), Harmonic Product Spectrum (HPS) algorithm.

1. General Structure Of A Speech Identification Process

The speech identification process can generally be divided in many different components as shown in following structure [2]. Recordings was taken in multiple languages like English, Hindi, Catalan, French, Finnish, Portuguese, Italian, Indonesian On one sentence 'Now This Time Your Turn' with the help of multiple speakers . Then pre-processed the data as shown in fig3, so that feature extraction become easy, because feature extraction is based on pre- processing. The main purpose of pre-processing is to reduce the noise and make our data smoother. After break-down each word from the different sentences for matching each utterance with other utterance. Our database contains the eight different foreign languages with different speakers.



Fig2: General Structure for Speech Identification

In load database each utterance is loaded one by one when the database is loaded if any utterance is misplace during pre-processing then we come to know in load database. We load each utterance so that we easily identify one person utterance with another person utterance as shown in fig4. In the next block give the input sound file to the system (input), then the feature are extracted with the help of following feature extraction techniques .where in fig5, The row represent the number of speakers and the column represent the number of feature which extract after applying feature extraction techniques.

3. Features of speech

The main aim of Speaker identification is to determine speaker's identity from his/her speech utterances. While speaking every speaker have his own characteristics. These characteristics are called speaker feature which can be extracted from speech utterances [8]. There are six major classes of spectral of analysis algorithms i.e. Digital filter bank, Fourier Transform, Linear Prediction used in speech recognition system, where the LPC gives the best result for the identification.

LPC (LPC analysis): LPC is one of the most powerful speech analysis techniques for extracting good quality features and hence encoding the speech signal at low bit rate [4]. LPC have capability for speech compression, synthesis and as well as identification. LPC is spectral estimation technique because it provides an estimate of the poles of the vocal tract transfer function. The LPC algorithm is a Pth signal is stationary within and zero outside, the analysis window. The autocorrelation solution to equation can be expressed as

$$R(i) = \sum_{j=1}^p a(j)R(i-j) \quad j=1, \dots, p$$

Where, $R(j)$ is an even function and is computed form:

$$R(j) = \frac{1}{\gamma} = \sum_{m=0}^{n-1-j} s(m)s(m+j) \quad j, \dots, p$$

Where γ is normalization factor. Once the auto correlation term $R(j)$ have been calculated, a recursive algorithm named Levinson-Durbin Algorithm used to determine the value of $a(j)$ [9]. Different techniques for feature extraction have been studied but LPC is more powerful so I use LPC for feature extraction in multilingual identification process:

- Temporal analysis techniques involve less computation, ease of implementation. But they are limited to determination simple speech parameter like power, energy and periodicity of speech. For finding vocal parameter we simply require spectral analysis techniques.
- Mel Cepstral analysis has decor relating property of cepstral analysis and also include some aspects of auditions.
- LPC provides the compact representation of vocal configuration by relatively simple computation compared to cepstral analysis.
- LP derived cepstral coefficients have decor relating property of cepstrum and computational ease of LPC analysis.
- Critical band filter bank decomposes the speech signal into discrete set of spectral samples containing information, which is similar to the information, presented to higher levels processing in auditory system.
- Cepstral analysis separates the speech signal into component representing excitation source and a component representing vocal tract impulse response. So it provides information about pitch and vocal tract configuration. But it is computationally more intensive [10].

4. Clustering Algorithm

Clustering algorithm is used for the task of searching groups of data such that the objects in same cluster are of similar kind and objects in different clusters are of different kind. Clustering can be used as an important tool to break the whole high dimensional input space into reduced input space without any change in the dimensionality and also to train the ANN with large amount of data [8]. We used “Fuzzy k- Means Algorithm” for clustering data [6, 9, 10]. As the number of clusters increased it make the solution more localized.

K-mean clustering can improve the correctness of the speaker identification. Various types of techniques are used for identification, the well known technique k-mean clustering algorithm is used which is sufficient algorithm [11]. K-mean clustering algorithm is based on two necessary conditions for optimality: the centroid and the nearest neighbor conditions [12]. When gave the input to the system ,the input sound is match with every utterance present in the database and the clustering result is produce as shown in fig 11 and when the speaker is recognized then the cluster is shown is as shown in fig12. When two utterances match with one another give the highest value than the other matching values. Clustering is applied on each feature independently because it give the better generalization. It provides a method to group the data from multidimensional space into a specific number of different clusters.

Run Neural Network: To train the network gives the number of input utterance to the network and see the result as shown in fig7 . When we gave the input the network give output like the network performance as shown in fig8 and the training of network in fig9.

5. Results:



fig3: Pre processing of the utterances

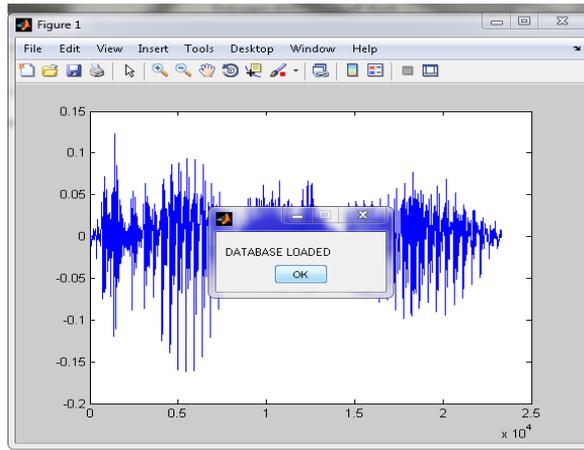


fig4: database loaded

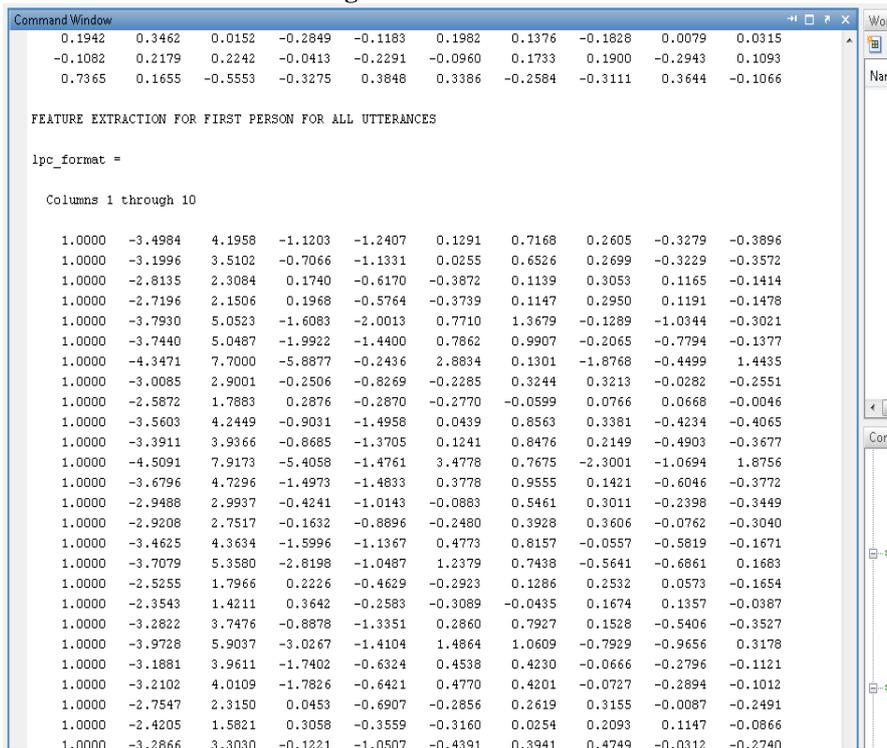


Fig5: Feature Extraction of different speakers from database

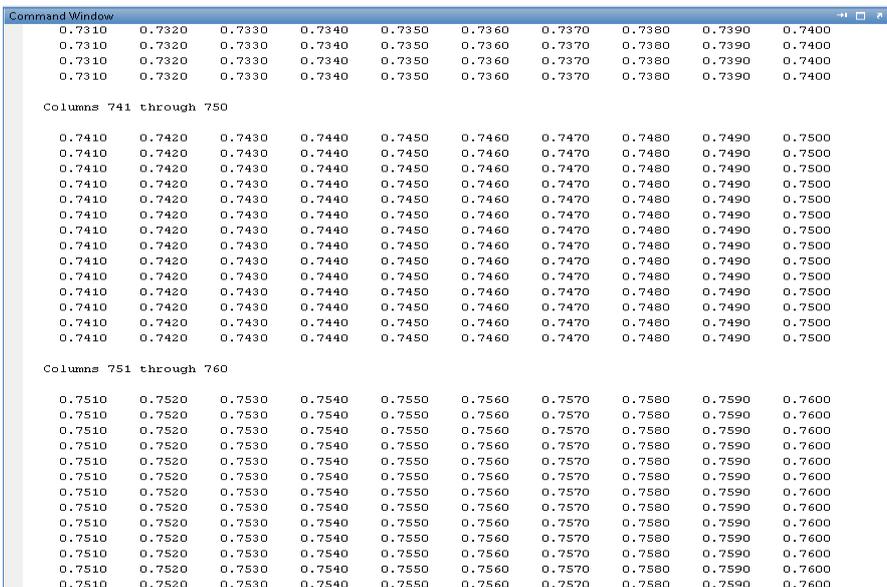


Fig6: Target Matrix of whole speaker utterances

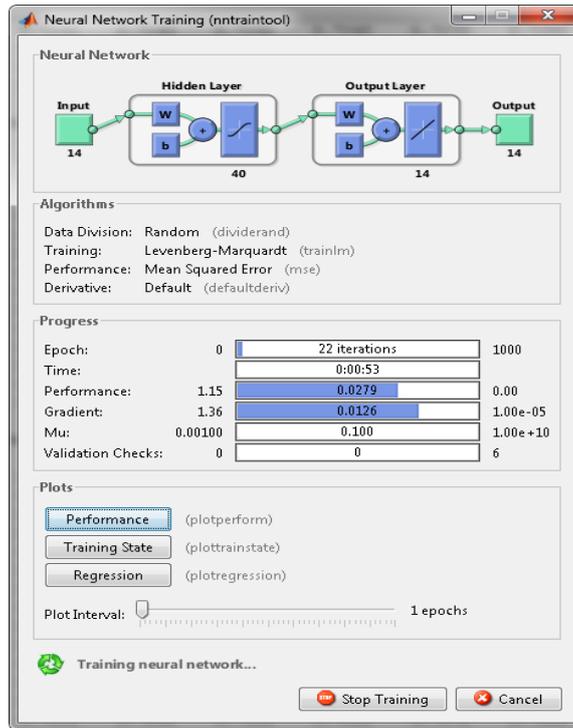


Fig7: Train the Neural Network

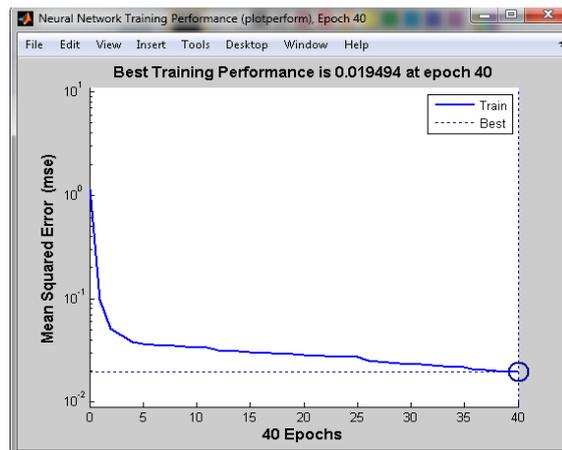


Fig8: Performance of the system

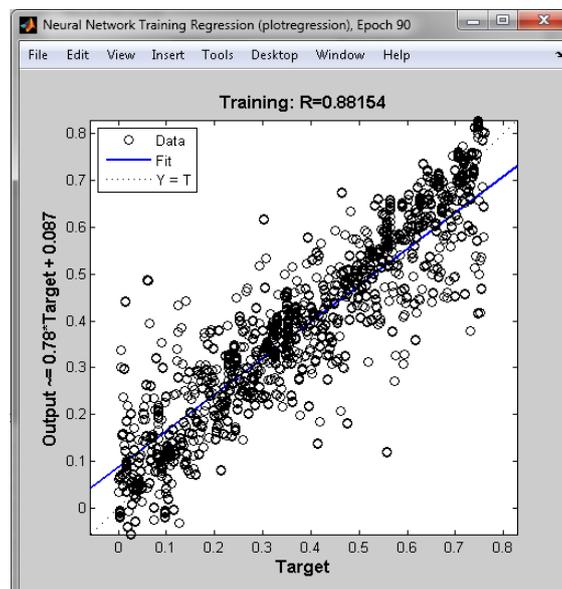


Fig9: Network Training Regression

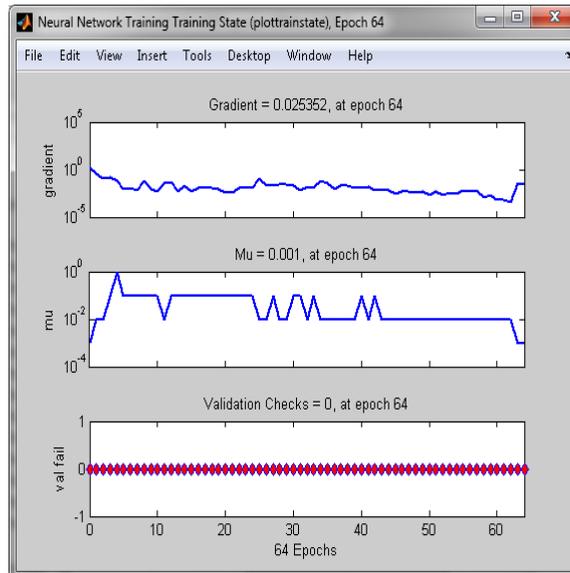


Fig10: Training State of the Network

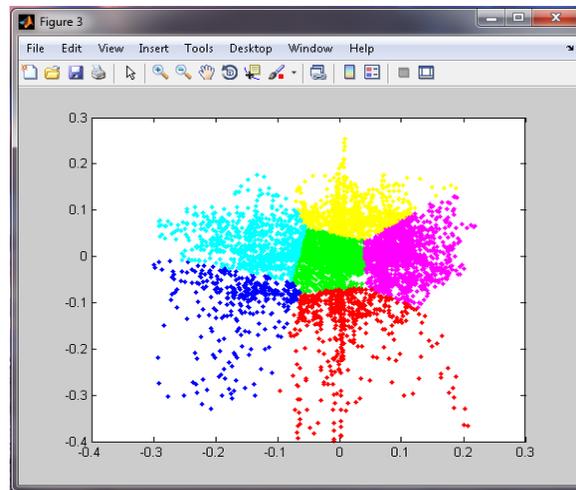


Fig11: Clustering of the Data

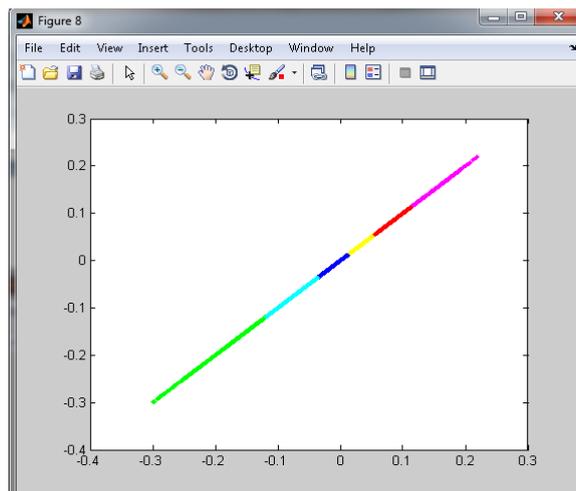


Fig12: After matching (identification) The Clustering Result

6. Conclusion

A series of cluster based multi lingual speaker identification experiments using neural network has been conducted. The result shows that k-mean clustering can be used for multi lingual system. This research focuses on text dependent speaker identification. The minimum performance of the system is 92.08% while the best performance is being reached up to 100%. Overall performance of the system is reached on 96%. This system is unable to identify on very small words. The future work can include small words also. It can increase the performance of this system.

References:

- [1] M.A.A Nusuya, S.K. Katti, Speech Recognition by Machine: a review, International Journal of Computer Science And Information Security, 2009
- [2] Wouter Gevaert Tsenov, Valeeri Mladenov, Neural Networks Used for Speech Recognition, Journal Of Automatic Control, vol2012
- [3] Santosh K.Gaikwad, Bharti W. Gawali, Pravin Yannawar, A Review Of Speech Recognition Technique, International Journal Of Computer Application, nov 2010
- [4] N.Uma Maheswari, A.P.Kabilan, R.Venkatesh, A Hybrid Model Of Neural Network Approach For Speaker Independent Word Recognition, International Journal Of Computer Theory And Engineering, Dec2010
- [5] Oriol Vinyals, Suman V.Ravuri, Daniel Povey, Revisiting Recurrent Neural Networks For Robust ASR, International Computer Science Institute, Berkley, CA,USA
- [6] Parteek Agarwal, Anupam Shukla, Ritu Tiwari, Multi lingual speaker recognition using artificial neural network, , Indian institute of information technology and management.
- [7] Parteek Agarwal, Gurpreet Kaur, Harjit Kaur, Multi lingual Speaker Identification Using ANN, International journal of computer application(IJCA), Nov2012
- [8] YUE Xicai YE Datian, department of electrical engineering and applied electronic technology, Tsinghua University, Beijing 10084, P.R. china. LIU Ming Telecommunication Institute, AIR force Engineering University, Xi'an 710077, P.R. china
- [9] Bhupinder Singh, Rupinder kaur, Nidhi Devgun, Ramandeep kaur, The Process Of Feature Extraction In Automatic Speech Recognition System For Computer Machine Interaction With Humans , International Journal of advanced research in computer science and software engineering, volume2, Feb 2012,
- [10] Manish P.Kesarkar, Feature Extraction For Speech Recognition, M.tech credit seminar report, electronic systems group, E.E. Deptt, IIT Bombay.
- [11] Bing Sun, Wenju Liu, Qiu Hai Zhong, Hierarchical Speaker Identification Using Speaker Clustering, National Laboratory Of Pattern Recognition, Institute Of Automation
- [12] YUE Xicai YE Datian , Department Of Electrical engineering and Applied Electronic Technology, LIU Ming Telecommunication Institute , Air Force Engineering University
- [13] Simon Haykin, Neural Networks and Learning Machines, Third Edition