Comparative Study and Evaluation of Speaker Verification Systems on Various Techniques: A Literature Review

Sneha Sahu[1], Neerja Dharmale[2]

[1]Dept. of Electronics & Telecommunication
Rungta college of Engineering & Technology
Bhilai,India
Sneha.sahu2011@gmail.com

[2]Dept. of Electronics & Telecommunication
Rungta College of Engineering & Technology
Bhilai,India
n.dharmale@rediffmail.com

Abstract: ***Nowadays, speech recognition is become more and more important. Various speech applications are available in the market. Consumer electronic devices can operate through voice. In the proposed system mobile phone is used as a controller to control whole system. In mobile phones CELP (code excited linear prediction) method is used for speech coding. CELP based speaker verification method, the CELP encoding method used for mobile phone voice communication is applied to the encoded voice data to perform speaker verification. A fuzzy c-means algorithm is used, in fuzzy clustering data point can belong to more than one cluster, and associated with each of the points are membership grades which show the degree to which the data points belong to the various cluster. FCM algorithm gives better result for overlapped data set and is comparatively better than k-means algorithm.***

Keywords: Consumer electronic device, mobile phone, CELP, LSP, speaker verification.

1. Introduction

The speech is the most common and primary mode of communication among human beings. In the systems, some speech recognition systems use "training" which is also called enrollment where each speaker reads text or isolated vocabulary. Accuracy is increased in the system by analyzing the person's original voice and uses it for fine-tune the recognition of that person's speech. "Speaker independent" Systems are those that do not use training and the system that use training are called "speaker dependent". The term voice recognition refers identifying the speaker, rather than what they are saying. For security process, recognizing the speaker can simplify the task of translating speech in system that had been trained on a specific person's voice or to authenticate or verify the identity of a speaker it can be used. The basic block diagram of the system is shown in Fig .1.

In the system, first block is controlling device which is used to control the system. Controlling device can be mobile phone or any microcontroller. Use mobile phone to control the system. Input to this is speech command of person who wants to control the device at a distance. After this a transmitting network is present. Over a point-to-point or point-to-multipoint communication channel data transmission, digital transmission or digital communications is the physical transfer of data. Whereas analog transmission is the transfer of a continuously varying analog signal over an analog channel, digital communication is the transfer of discrete messages over a digital or an analog channel. The transmitted signal is received by the receiving network. After this to recognize the speech, speech recognition is present in the system. In mobile phones CELP (code excited linear prediction) method is used for speech coding. For creating minimally redundant representation of a speech signal speech coding is used. The aim of all speech coders is to minimize the disturbance or noise at a given bit rate, or minimize the bit rate to reach a given distortion with high quality [1]. By speech coding process distortions are very less and system becomes more efficient. Accuracy of the system can be more and obtained good result. In speech coding good quality of speech can be obtained by analysis-by-synthesis method. To match the reconstructed speech waveform the excitation signal is chosen by attempting as closely as possible to original speech waveform in the time domain analysis-by-synthesis coder. After speech recognition, finally speech signal is transfer to any device. It can be any electronic device such as bulb, fan, machine, etc. and in this way final output is obtained.

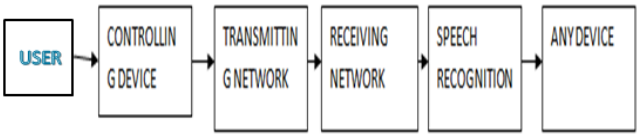

Figure 1: Block diagram of the system

2. Classification of speech

There are many parameters define the capability of a speech recognition system [2].

2.1 Isolated word

Isolated word contain sample windows, it receives single word or single utterances at a time. Isolated utterances may be a better name with this work.

2.2 Connected word

Connected word and isolated word both are of similar type but allow separate utterance to be run together minimum pause between them.

2.3 Continuous speech

In this computer will examine the content and allows the user to speak naturally. Various methods are there to determine utterances boundaries and difficulties occurred in it.

2.4 Spontaneous speech

Spontaneous speech means a system should be capable to handle a variety of natural speech feature such as words being run together.

3. Speech recognition techniques

Nowadays different techniques are available for speech recognition. The target of speech recognition is to characterize, analyze, extract, and recognize information about the speaker identity. Speech recognition techniques are used in different ways. For determining the speech characteristics various techniques are used. In this technique speech is analyzed in different manner and speaker identity is verified. The speech data contains different type of information that shows the speaker identity. This involves speaker specific information due to excitation source, vocal tract, and behavior feature. Speech analysis techniques are used in various purposes. It is very essential to use and the speech analysis stages are of the following three types.

3.1 Segmentation analysis

In segmentation analysis, speech is analyzed using the frame size and shift in the range of 10-30 ms to extract speaker information. Vocal tract information of speaker recognition is extract by this method.

3.2 Sub segmental analysis

Sub segmental analysis is defined as speech analyzed using the frame size and shift in range 3-5 ms. For the excitation state this technique is used to mainly analyze and extract the characteristic.

3.3 Supra segmental analysis

In this work, using the frame size speech is analyzed. This technique is mainly used to analyze the behavior and characteristic of the speaker.

4. CELP- Based speaker verification

In this paper, a CELP based speaker verification method is proposed for electronics devices including mobile phones. In the CELP based speaker verification, the CELP encoding method used for mobile phone voice communication is applied to the encoded voice data to perform speaker verification. A system is proposed here for operating consumer electronic devices by voice using CELP (code excited linear prediction) parameters which are commonly used in speech coding in mobile phones. There is two characteristics of the CELP based speaker verification. First is speaker verification used only the encoded voice data and it does not need the decoding process. So that is can perform on any electronic devices. And the second is when the mobile phone is used to operate the consumer electronics devices it can use the voice encoding function that is built in to the mobile phone itself. No additional data is required to operate the system. Figure 2. Shows the flow diagram of CELP based speaker verification method [3].

4.1 Overview

Speaker verification has two phases' enrollment and verification:

4.1.1 Enrollment

- By using microphone transforms user's voice in to an electronics signal.
- By using CS-ACELP (conjugate structure algebraic code excited linear prediction) encoding, encode the electronic signal. After that extract the LSP's (line spectrum pair) which are used as the full rate standard codec for digital mobile phones.
- By using the power of the speech signal extract the speech interval from the speech signal and by this remove the silence interval from the speech signal.
- After extraction of speech interval number of frames can be reduce. Reduce number of frames by replacing the frame data with cluster centers. Cluster centers can be calculated by using k-means method.
- By using the kernel principal component analysis (KPCA) calculates the basis from the reduced frame data and in terms of the basis represent the enrollment subspace.

4.1.2 Verification

- By using the microphone transform user's voice in to an electronics signal.
- By using the CS-ACELP (conjugate structure algebraic CELP) encode the electronic signal and extract the LSP's. Represent the frame data as a vector concatenating the LSPs extracted from each frame.
- By using the power of the speech signal extract a speech interval from the speech signal and by this remove silence interval from speech signal.
- By using the KPCA calculate the basis from the reduced frame of data and in terms of the basis represent the input subspace.
- Calculate the similarity.
- Compare the similarity with the threshold. If the similarity is greater than the threshold then select yes otherwise select no.

4.2 CELP Coding

In the CELP encoding process, using CELP the input speech data is encoded and the encoding parameters are extracted. To calculate a synthesis filter for quantization a linear predictive coding is used. To calculate an excitation signal for vector quantization analysis by synthesis is used. For highly efficient voice encoding in mobile phones the CELP framework is widely used.

The proposed verification method uses CS-ACELP, which has been standardized as ITU- T G.729. CS-ACELP uses 8-kbit/s encoding. It produces the same voice quality as 32-kbit/s ADPCM (Adaptive differential PCM).

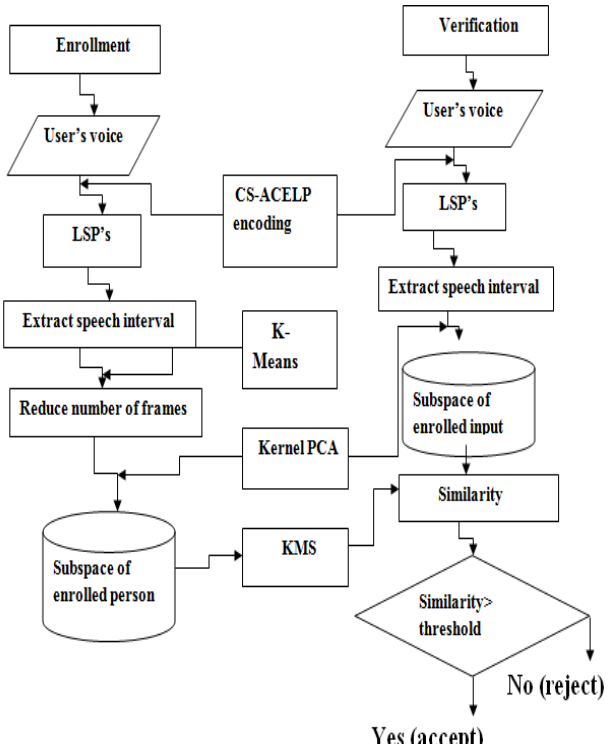

Figure 2: Flow of CELP based speaker verification method

By changes in the shape of the vocal tract LSP is the one that corresponds to speech articulation controlled and it corresponds good to the voice spectral envelope. The spectral envelope is a physical feature that reflects voice quality, and its use as an effective feature for speaker recognition has been demonstrated in previous research.

5. Feature Extraction

There are two types of clustering:

- **Hierarchical Clustering:** Hierarchical tree is defined as a set of nested clusters.
- **Partitioning Clustering:** A division data object into non-overlapping subsets such that each data object is in exactly one subset.

5.1 K-Means Algorithm

K-means clustering is a partitioning method. It partitions data in to k mutually exclusive clusters, and returns the index of the clusters to which it has assigned to each observation. Single level of cluster K means clustering creates a single level of cluster and operates on actual observations. For large amount of data the distinctions mean that k-mean clustering is often more suitable than hierarchical clustering [4].

K-means uses a iterative algorithm that minimizes the sum of distance from each object to its cluster centroid, overall cluster. This algorithm moves object between clusters until the sum cannot decrease further. The result that is obtained is a set of clusters that are well separated as possible. The details of minimization using several optional input parameters to k-means should be control, including ones for the initial values of the cluster centroids, and for the process of maximum number of iterations. With the initial starting point values the input parameters of the clustering algorithm are the various numbers of clusters that are obtained. Using equations, when the initial starting values are given, the distance between each initial starting value and sample data point is found. Then in the cluster associated with the nearest starting point each data point is placed. The new cluster centroids are calculated after all the data point is assigned to a cluster. The new centroid value is then determined for each factor in each cluster. This procedure continues until the centroids no longer moves or no more data point changes.

Limitation of k-means is that it has problems when clusters are of various densities, sizes and non-globular shapes. And when the data contains outliers k-means has problems.

Algorithmic steps for k-means clustering [5] :

- Set k- to choose a number of desired clusters, k.
- Initialization – to choose starting points which are used as initial estimates of the cluster centroids. They are taken as the initial starting values.
- Classification – to examine each point in the dataset and it to the cluster whose centroid is nearest to it.
- Centroid calculation – when each point in the data set is assigned to a cluster, it is needed to recalculate the new k centroids.
- Convergence criteria – the steps of three and four are repeated until the centroids no longer move or no point changes its cluster assignment.

5.2 Fuzzy C-Means Clustering Algorithm

In modern science fuzzy logic becomes more and more important.Clustering involves the task of dividing the data point in to homogeneous classes or clusters so that items in the same type are as similar as possible and items in different types are as dissimilar as possible. In hard clustering data is divided in to crisp cluster, where each data point belongs to exactly one cluster. The data point can belong to more than one cluster in fuzzy clustering, and associated with each of the points belong to one cluster, and associates with each of the points that are membership grades which indicates the degree that is belong to the different clusters. At present there are many different methods of fuzzy clustering nowadays. In MATLAB we review the fuzzy c-means clustering method in

our work. Generally the FCM algorithm is known as the constrained optimization of the square-error distortion

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \|\mathbf{x}_j - \mathbf{v}_i\|_A^2$$

Where U have been the (c*n) partition matrix, V = {v1….vc} have been the set of c cluster centers in R^ d, m>1 is the fuzzification constant, and ‖. ‖A is any inner product A-induced norm. We may only use the Euclidean norm (A = I) in our examples, but there are many examples where the use of another norm-including matrix, e.g., using A=S^-1 is the inverse of the sample covariance matrix, have been shown to be effective [4]. The FCM/AO algorithm produces a solution to using alternating optimization (AO).

In the clustering solution, the weighted FCM model introduces weights that define the relative importance of each object. Therefore, wFCM is defined as the constrained optimization of

$$J_{m\mathbf{w}}(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} w_j u_{ij}^m \|\mathbf{x}_j - \mathbf{v}_i\|_A^2$$

Where w > 0 is a set of predetermined weights that define the purpose of each feature vector. FCM iteratively moves the cluster centers to the right location within a dataset. In k-means algorithm, specific introducing the fuzzy logic is the fuzzy c-means algorithm in general. In fact fuzzy c-means techniques are based on fuzzy behavior and they provide a technique which is natural for providing a clustering where membership weights have been a natural interpretation but not probabilistic at all. This algorithm is basically same type in structure to k-means algorithm and it also behaves in simple fashion.

6. Modeling technique

For creating speaker models, modeling technique is used for the specific feature. The speaker modeling technique is basically are of two types one is speaker recognition and other is speaker identification. Again speaker recognition can be divided into two methods. One is Text- dependent and the other is text independent methods. In text dependent method the speaker speaks key words or sentences for both testing and training trials having the same text whereas text independent did not rely on a specific texts that have been spoken. Following are the methods used in speech recognition process are as follows:

6.1 Pattern Recognition approach

A speech pattern can be introduces in the form of a speech template or a statistical model (e.g., a HIDDEN MARKOVMODEL or HMM) [6] and can be applied to a sound, a word, or a phrase. In various practical problems, pattern recognition has been developed for two decades and received much attention over the system.

6.2 The acoustic-phonetic approach

This method have been studied and used for more than 40 years. This approach introduces a theory which is based upon acoustic phonetics and postulates.

6.3 Learning based approaches

Machine learning methods were introduced in neural networks and genetic algorithm learning based approaches had been introduced to overcome the disadvantages of the HMMs.

6.4 Knowledge based approaches

The guidance should be taken from an expert knowledge about variations of speech is hand coded in to a system. This approach has been given the advantages of explicit modeling but this situation is much difficult to found and can't used successfully.

6.5 Artificial intelligence approach

According to the person who applied it, the artificial intelligence approach coordinates the recognition procedure. The intelligence of a person such as analyzing synthesizing ,visualizing, etc are used for making a decision on the measured acoustic features. For the acoustic phonetic approach and pattern recognition approach, artificial intelligence is a hybrid.

7. Conclusion and future work

CELP (code excited linear prediction) method is proposed. In mobile phones, CELP method is used for speech coding. By voice system can be remotely operate consumer electronic devices using the code excited linear prediction methods that have been used for speech coding in mobile phones was proposed. To protect private information and separate user's voice from other person nearby also speaking a speaker verification function has been introduced. A fuzzy c-means algorithm is used. This algorithm works by assigning membership to every data point corresponds to every cluster center on the basis of distance between the data point and cluster center. FCM Algorithm gives best result for overlapped dataset and is comparatively better than k-means Algorithm.

References

[1] Nimisha Susan Jacob, Ancy S. Anselam, Sankuntala S. Pillai "performance analysis of CS-ACELP speech coder" IJEAT, vol.4, pp. 191-195, June 2015.

[2] Shikha Gupta, Amit Pathak, Achal Saraf "a study on speech recognition system"IJSETR, vol.3, pp. 2192-2196,August 2014.

[3] Masatsugu Ichino, Yasushi Yamazaki, Hiroshi Yoshiura "speaker verification method for operation system of the consumer electronics devices" IEEE, vol.61, February 2015.

[4] TimothyC.Havens, Christopher Leckie, Lawrence O. Hall " Fuzzy C-means clustering Algorithms for very large data" IEEE, vol. 20, pp. 1130-1145, December 2012

. [5] Soumi Ghosh, Sanjay Kumar Dubey " Comparative study of k-means and fuzzy c-means algorithms" IJACSA, vol. 4, pp. 35-39, 2013

[6] Tzuu-Hseng S. Li, Min Chi Kao, Ping-Huan Kuo"recognition system for Home-Service-Related Sign Language using Entropy-Based K-means algorithm and ABC-Based HMM",IEEE,,pp. 1-13.

.