Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 5, May 2015, pg.394 – 400

RESEARCH ARTICLE

A MFCC Integrated Vector Quantization Model for Speaker Recognition

Suraina Kinha
Student, Dept. of ECE, ITM University Gurgaon
Email id: Suraina0403@gmail.com

Dr. Rekha Vig
Guide, Dept. of ECE, ITM University Gurgaon
Email id: rekhavig@itmindia.edu

Abstract: Speaker recognition is one of the cost effective biometric authentication method adaptive by many online and offline applications. The mobile phone integrated passwords, voice commanders have increased the use of speaker authentication. But these recognitions are having various challenges in terms of background noise, instrumentation noise etc. To provide the robust system authentication, it is required to provide the impurity preserved recognition. In this work, an improved layered model is presented. In first stage of this mode, the feature exploration is achieved using LMS approach. The feature set adaptation is considered as the second stage applied using MFCC method. Finally, vector quantization is applied to perform the recognition. The experimentation results show that the work has provided effective speech signal recognition. The work is presented as an online application.

Keywords: LMS, Vector Quantization, Speaker Recognition, MFCC

I. INTRODUCTION

Speaker Recognition is the biometric feature that can captured from most easy means such as mic, mobile, telephone etc. It can be used to provide online as well as offline authentication. Speech recognition itself having various challenges and applications. The challenges are in terms of various inclusive disturbance and noise vectors. As the capturing devices are cheap and available in different forms, these devices also affect the capturing quality. If some cheaper instrument is used to capturing the voice, it can include instrumentation noise or device interference as the noise vector included in the signal itself. Another challenge to the recognition system is the voice frequency of the speaker. The person sound or voice frequency at two different time instances can be different. Because of this, it is required to analyze the speaker's sound pitch and frequency and based on it perform the robust recognition. Another problem associated with this recognition model is background noise. As the surrounding recording environment can also include different noises such as traffic noise, fan noise etc. These noises also affect the quality of input voice so that the overall recognition process can be affected. To perform the recognition there are number of associated vectors. The speech recognition also have various inclusive methods and applications. These applications include speech transformation, speech compression, data hiding using speech vector etc. These all speech processing applications and methods are shown in figure 1

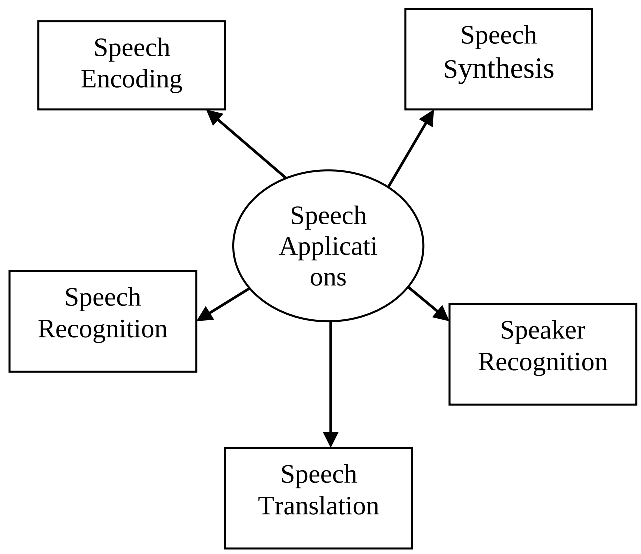

Figure 1: Applications of Speech

Here figure 1 showing the various applications of speech processing. One of foremost application is speech encoding. This encoding stage is define to transform the input to some specific form so that the information processing can be done. The encoding itself includes the process to transform the input to some secure form so that safer transmission will be obtained. The signal effective cryptography and data hiding techniques can be applied to achieve speech encoding. Another application area is speech synthesis. It basically converts information from once source to other so that effective device specific transformation will be done. The input mediums are defined to perform the speech generation and to provide the effective speech communication. To perform this speech transformation, the work is to reduce the speech distortion. The device specific transformation is applied to achieve the effective information transmission.

Speaker recognition is the type of biometric authentication application model used to identify the person based on speech signal estimation. The voice commanders can be used to recognize the person or the person voice to recognize the speech. This kind of system includes speech content analysis also. If the speech contents are also analyzed and identified, it comes under speech recognition. Speech translation is also used the globalization approach used to convert the speech to various other information forms. These forms include the textual information or content representation, English speech to hindi speech translation. These kinds of systems are adaptable to mobile communication such as SMS translation etc.

In this paper, a feature adaptive Vector quantization recognition model is presented. The work is divided in two main stages. In first stage, feature exploration and feature extraction is done. To improve the signal features LMS adaptive approach is applied. In second stage, MFCC is applied to extract the signal features. Finally, vector quantization is applied to perform the recognition and classification of signal. In this section, an exploration to speech recognition model is defined. This work include various speech processing stages. In section II, the work defined by earlier researchers is defined. In section III, the proposed research methodology is described. In section IV, the results obtained from work are described. In section V, the results obtained from the work are presented.

II. EXISTING WORK

Speech processing and recognition is common biometric application area adapted by different researchers. Some of the contribution of earlier researchers is described in this section. Tatsuhiko Kinjo[1] has presented a HMM integrated approach for complex speech signal recognition. Author defined a statistical feature based approach to transform the speech in feature using HMM approach. Once the adaptive signal form is captured, the estimation to the accurate speech form is done using speech spectrum analysis and low frequency signal derivation. This kind of analysis is performed to generate the speech feature that is been used to identify the signal from the pool. The distance adaptive mapping is applied to recognize the speech signal. The results shows that the work has provided effective speech recognition.A.M. Peinado[2] has provided a work on vector quantization under HMM integration to provide speech recognition. Author used the multi model derivation with semi continuous speech signal mapping to provide the recognition under derivative mapping. The lesser computation is applied in this integrated vector quantization approach. Author used SCMVQ modelling to achieve the more effective and accurate results. GongiunLi[3] has provided a HMM based approach to provided the probability effective mapping to provide the proportionate contribution with speech samples. The parameter based mapping is applied to obtain the most mapped speech sample so that the geometrically effective recognition will be obtained. S. Kwong[4] has provided a work on parallel genetic algorithm under HMM integrated approach. Author used HMM as the statistical measure to generate the featureset. Later on J3MM model is applied to describe the utterance is the signal features. Finally, the genetic approach is applied to recognize the signal. Tetsuya Takiguchi[5] has provided a separation effective speech signal recognition so that the distance speaker will provide the effective recognition rate. The acoustic measure based recognition under HM separation and maximum likelihood mapping is applied for data adaptive recognition. Author reduced the signal noise and provided the accurate signal recognition and transition. Author defined the HMM composition to provide the relative signal recognition.

Wei Han[6] has provided a table lookup based modelling to perform HMM integrated speech recognition. Author employed the design measurement and recognition and word isolation so that the effective word recognition will be obtained. Shingo Yoshizawa[7] has defined the scalable architecture under word processing so that the accurate speech recognition will be done. Author defined a acoustic model under phoneme level analysis and word level analysis to provide recognition. Author defined the vocabularies and computation time effective recognition. Sven E. Kruger[8] has provided a SVM based model for HMM integrated speech recognition. Author defined a parallel system model for improving the recognition. Author used the speech integration to improve the recognition. Panikos Heracleous[9] has provided a work on vowel and consonant recognition using HMM based modelling. Author defined a cued system modelling to provide the speech recognition for French. Author defined the integrated character modelling to reduce the problems including the spoken impurities and noise. Author applied work on different experimentation so that the recognition will be improved.

Wushour Silamu[10] has provided a work on Uyghur continuous recognition model so that the recognition process will be improved. Author defined the agglutinative language formation based work on language criteria mapping for improving the recognition modelling. Author defined a work on bulding the recognition process under continuous speech dataset. Cong-Thanh Do[11] has provided a work on spectral signal analysis based feature extraction approach. The feature extraction is performed using MFCC and HMM modelling and later on coefficient driven modelling is applied to improve the recognition. The sub band temporal model is here defined to generate the acoustic information processing so that the recognition rate will be improved. Panikos Heracleous[12] has provided a work HMM modelling and speech signal processing so that the visual information will be obtained from the signal set. The spectral space analysis and distance preserved signal transformation is applied to perform the recognition. Yeh-Huann Goh[13] has provided an adaptive framing on speech signal processing. Author defined a work on spectral signal derivation so that the pitch controlled signal processing will be done. Author improved the signal and the wavelet transformed rates are obtained.

III. RESEARCH METHODOLOGY

In this present work, a three stage model is presented to perform the speaker recognition. The presented model is feature adaptive that is able to use the obtained statistical features to perform the accurate recognition. This model is shown in figure 2. As shown in the model, the model is having a filtered dataset. The dataset is here present in the form of integrated signal features available in the form of feature set. This dataset will be processed for any of input speech signal. To process this speech signal, the raw speech signal is taken in .wav format. Once the input is taken, it can have different impurities discussed previously. Because of this, there is the requirement to improve the signal features. To improve the signal features adaptive LMS approach is defined. This approach is based on the sample signal set to reduce the signal noise and to enhance the signal feature. Once the features are improved, the next work is to extract the effective signal features. In this work, the MFCC is applied to extract the signal features. The feature extraction is defined to generate the feature vector from the input signal. This featured signal form is finally compared with database signal set. To perform this mapping vector quantization approach is suggested in this work. This vector quanitization has performed the feature mapping to recognize the speaker. The work stages of this model are described in this section

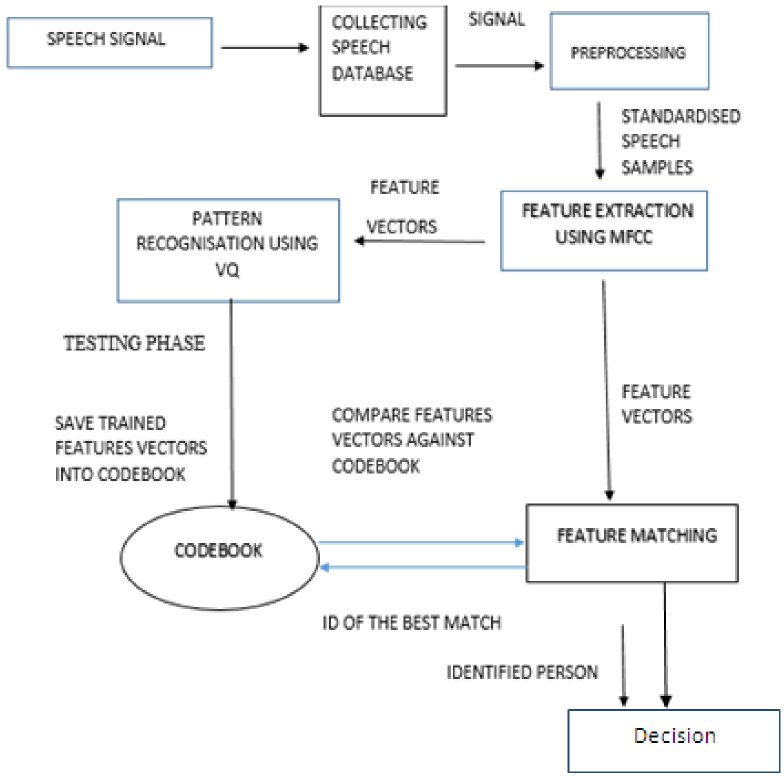

Figure 2 : Proposed Model

A) Feature Extraction

Feature extraction is basically adapted in this work to transform the waveform to featureset. The recognition will be here applied using feature adaptive mapping. The characteristics analysis is here performed under speech segment generation and relative feature characterization. Mel-Frequency Cepstrum (MFC) is used in this work as the coefficient vector to generate the feature set. This method is the derived represented to process the spectrum and to derive the signal processing under frequency adaptive mapping. The frequency band formation and spaced scaled signal processing is done to obtain the linear space features. The normal spectrum is analyzed to obtain the information gain that is different for different speakers so that the relatively effective recognition will be performed. The word specific descriptive analysis is defined to perform the recognition.

B) Matching

Once the features are extracted, the next work is to perform the recognition using feature adaptive mapping. The pattern recognition is performed on multiple classified objects to obtain the object class. The interest adaptive pattern analysis is performed to generate the speech sequence so that the relative effective recognition will be performed. The extracted signal processing is done to generate the feature class so that the supervised pattern recognition and mapping will be done. This adaptive mapping process is here defined under the specification of relative training set so that the classification derived information processing will be done to obtain the classification algorithm so that the adaptive recognition process will be obtained over the set.

In this work, vector quantification process is applied to perform the recognition for known and unknown speaker. To provide the recognition over the unknown speakers, the dataset featureset inclusion is also implied in this work. The recognition process is here defined under two processes. In first stage, the LBG algorithm is applied to separate the feature vectors under code book vectors and mapping with training set. Later on Euclidian distance mapping is applied on test set to perform the recognition.

The obtained results show that the work has provided the effective and accurate recognition. The recognition experimentation is described in next section.

IV. RESULTS

The presented work is experimented in real time environment in which the input speech signal is captured from user. The dataset is adaptive and dynamic in which new signals can be included for improving the accuracy of this work. The properties of this experimentation is given here under

Table 1 : Signal Set properties

Property	Values
Dataset	Dynamic (20 Signal Initially)
Dataset Form	MFCC Extracted Feature Set
Input signal	Raw Speech Signal
Format	.wav
Input form	Dynamic and Real time
Noise	Yes

The experimentation results are given here under

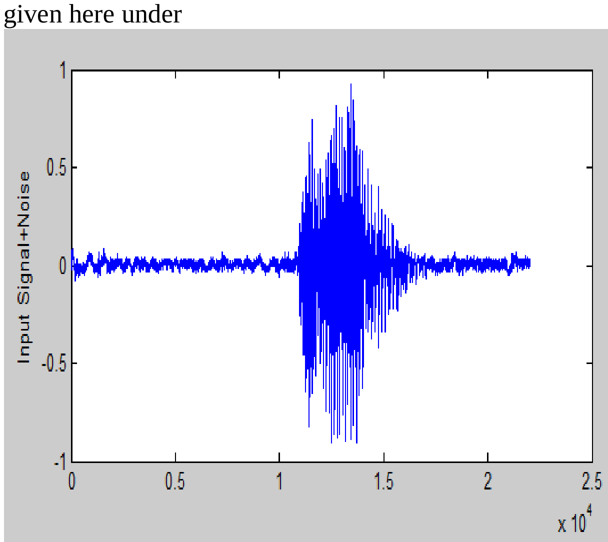

Figure 3 : Input signal

Here figure 3 is showing the input signal. Here x axis showing the signal length and y axis showing the corresponding amplitude value. The first stage of work is to apply LMS method to improve the signal features. The improved signal form is shown in figure 4

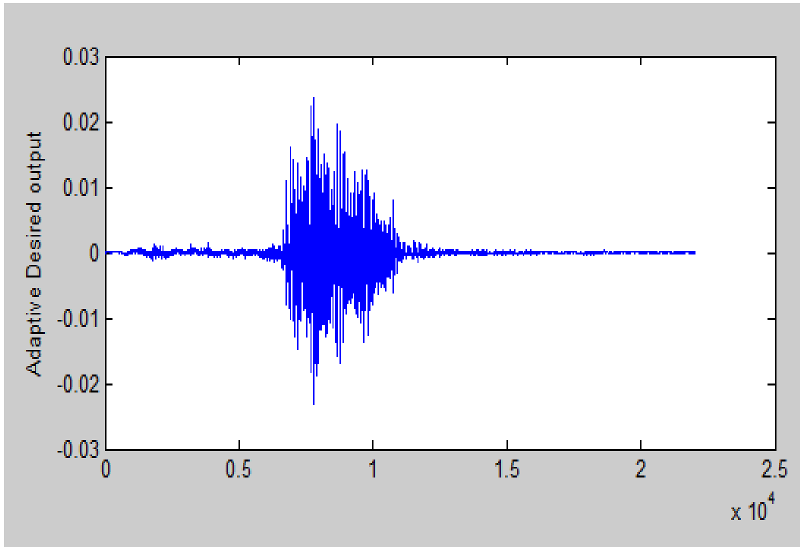

Figure 4 : Improved Filtered Signal

Here figure 4 is showing the improved signal form, the noise vector is removed from the signal. The noise reduction has explored the signal features.

Once the explored signal form is obtained, the final stage is to recognize the signal. The experimentation is here applied on about 10 different signal values that provided the accurate recognition for 8 to 9 signal. It shows that the work has provided the accuracy rate upto 90%.

V. CONCLUSION

In this present work, a three stage adaptive model is defined to perform speaker recognition on real time speech signal. The speech signal is here obtained in real-time environment. The work model has defined feature exploration as first stage. The feature enhancement is here done using LMS feature. In second stage, MFCC is applied to generate the signal features. In final stage, vector quantization is applied to perform the recognition. The experimentation results shows that the work has improved the accuracy and effectiveness of this recognition model.

REFERENCES

[1] Tatsuhiko KINJO," On Hmm Speech Recognition Based on Complex Speech Analysis", 1-4244-0136-4/06@'2006 IEEE

[2] A.M. Peinado," Use of multiple vector quantisation for semicontinuous-HMM speech recognition", IEE Proc-Vis. Image Signal Process 1994

[3] GongiunLi, "An Improved Trqi"G Algorithm in Hmm-Based Speech Recognition".

[4] S. Kwong, "Analysis of Parallel Genetic Algorithms on Hmm Based Speech Recognition System", 0098 3063/97@ 1997 IEEE

[5] Tetsuya Takiguchi," HMM-Separation-Based Speech Recognition for a Distant Moving Speaker", IEEE Transactions On Speech And Audio Processing 1063–6676/01 © 2001 IEEE

[6] Wei HAN, "An HMM-Based Speech Recognition IC", 0-7803-7761-3/031@2003 IEEE

[7] Shingo YOSHIZAWA,"Scalable Architecture for Word HMM-Based Speech Recognition".

[8] Sven E. Kruger," Mixture of Support Vector Machines for HMM based Speech Recognition", The 18th International Conference on Pattern Recognition (ICPR'06) 0-7695-2521-0/06 © 2006 IEEE

[9] Panikos Heracleous," HMM-based Vowel and Consonant Automatic Recognition in Cued Speech for French", VECIMS 2009 - International Conference on Virtual Environments, Human-Computer Interfaces and Measurements Systems

[10] Wushour Silamu," HMM-Based Uyghur Continuous Speech Recognition System", 2009 World Congress on Computer Science and Information Engineering 978-0-7695-3507-4/08 © 2008 IEEE
[11] Cong-Thanh Do," On the Recognition of Cochlear Implant-Like Spectrally Reduced Speech With MFCC and HMM-Based ASR", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING 1558-7916 © 2010 IEEE
[12] Panikos Heracleous," Analysis and Recognition of NAM Speech Using HMM Distances and Visual Information", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING 1558-7916 © 2010 IEEE
[13] Yeh-Huann Goh," HMM-Based Speech Recognition Using Adaptive Framing", TENCON 2009 978-1-4244-4547-9/09©2009 IEEE

