Indian Journal of Science and Technology, Vol 10(30), DOI: 10.17485/ijst/2017/v10i30/115518, August 2017
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Issues and Challenges of Voice Recognition in Pervasive Environment

Sonali Goyal* and Neera Batra

Maharishi Markandeshwar University, Mullana - 133207, Haryana, India; sonaliguglani21@gmail.com, batraneera1@gmail.com

Abstract

Objectives: To provide detailed study of all the algorithms and analyse them in a way to conclude that an integration approach describes the best combination for voice recognition in terms of recognition rate. **Methods:** Voice processing is a process in which words of a speaker are recognized by the information of the waves. There are number of algorithms used for voice recognition named as Perceptual Linear Prediction, Linear Predictive Code, Mel Frequency Cepstral Coefficient, Dynamic Time Warping etc. **Findings:** Graph is used to depict the recognition rate of all the voice recognition techniques with different types of classifiers named as HMM (Hidden Markov Model), DTW (Dynamic Time Wrapping), VQ (Vector Quantization), SVM (Support Vector Machine) etc. which clearly describes that the hybrid approach may provide better results as compared to individual methods. Performance and recognition rate is not so good by using individual techniques because it does not provide better recognition rate while taking into consideration the security of an individual living alone at home. After the comparative analysis, it is concluded that there is a need to develop a better combined approach which may provide better recognition rate as compared to individual methods. This paper will help the researchers to know the basic difference between the explained feature extraction techniques. **Application:** Main application we are using is Voice Recognition in order to provide security to an individual living alone at home.

Keywords: Feature Extraction, LPC, MFCC, PLP, Voice Recognition

1. Introduction

Speech is the ability of a program to receive and identify the spoken words[1,2]. One of the most important parts of voice recognition is extraction of features as shown in Figure 1 Any individual voice recognition algorithm can be used and implemented to provide security but it has been observed in this analysis that there is a need to develop combined approach which may provide better recognition rate as compared to individual methods.

Our primary focus is on voice recognition in order to provide a secure environment for an individual living alone at home.

In[3,4] it describes a comparative study of feature extraction techniques for a voice recognition system. It defines number of techniques which can be used named as PLP, LPC, MFCC, DTW, and RASTA, ZCPA etc in which RASTA with MFCC tells about the nature of speech and on the other hand, LPC defines the features of future which are based on past features. It shows a comparison table in which individual methods are described like PLP, LPC, and MFCC etc which gives different recognition rate in different applications. As our work concerns with face recognition using PCA algorithm so on that application LPC gives 83% recognition rate. In MFCC, application for security purpose provides 86% recognition rate. And it is also used for continuous speech using PCA with 85.3% recognition rate.

In[5,6] it presented MFCC technique of feature extraction with its applications. In this paper, it is said that MFCC is used for design a text dependant speaker system of identification. This paper describes the modifications in MFCC in order to enhance the accuracy of complete system. To calculate MFCC, It includes Mel frequency

*Author for correspondence

wrapping and Cepstrum. This paper describes the comparison of different implementations of MFCC by taking different number of filters. It means MFCC performance is affected by changing number of filters. If we take 12 filters it will give 75% efficiency but as we raise up to 42 filters, efficiency goes to 80%.

These papers[7,8] presented an efficient voice recognition system using different feature extraction techniques like MFCC, PLP, VQ, HMM (Hidden Markov Model). This paper defines how speaker recognition followed by speech recognition is best to recognise the voice efficiently. Two ways are used to complete the process. First is by using MFCC which extracts the input characteristics with respect to word spoken by speaker and then HMM is used to recognise that work. Results are calculated in terms of recognition rate.

Kuldeep[9,10] presented a Hindi speech recognition system using an application named as HTK (Hidden markov model toolkit). For recognition, it uses HMM. It takes vocabulary of 102 words and then system is trained to recognise the words. In initial step, MFCC is used to extract the features then HMM is used to train the network.

The rest of the paper is divided as follows: section 2 describes voice recognition techniques and section 3 presented the conclusion.

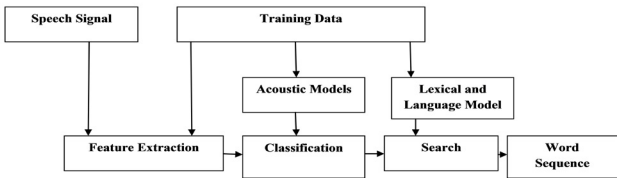

Figure 1. Voice Recognition Framework.

2. Types of Feature Extraction Techniques

2.1 LPC (Linear Predictive Code)

It is one of the feature extraction technique used to extract the features. It predicts the future features which are based on previous features. It is mainly required to compress signals for efficient transmission and for storage purpose shown in Figure 2. It works best when medium or low bit coder is used in[11,12] .When we pass the signals from filter in order to remove the redundant values, an error is generated as an output. Best way to transfer the signals is to send residual error with parameters of speech if we want to generate original signal. This technique is named as Linear Prediction as it is totally based on least mean square error theory.

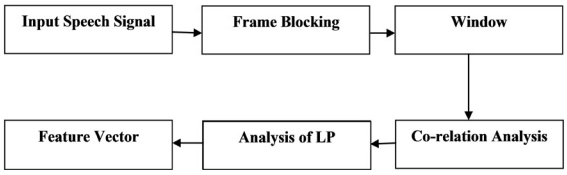

Figure 2. LPC (Linear Predictive Code) Model.

2.2 PLP (Perceptual Linear Prediction)

It is developed by hermansky. It is based on the concept of psychophysics of hearing as shown in Figure 3 This is the technique which improves the recognition rate and eliminates the irrelevant information provided by[13,14] . PLP and LPC are almost same except in PLP some special characteristics are transformed to match characteristics of auditory system.

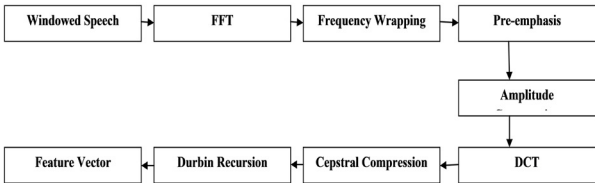

Figure 3. PLP (Perceptual Linear Prediction) Model.

2.3 MFCC (Mel Frequency Cepstral Coefficient)

As we know the first step in voice recognition is feature extraction or we can say to identify the variables of audio voice. It takes human perception sensitivity with respect to frequencies into consideration and therefore it is best for speech recognition[15]. It is based on the idea of cestrum. Firstly, we convert analogy representations to digital form. Analogy to digital conversion consists of two steps. First is Sampling: It is calculated by the number of samples per second and other is Quantization: It is the process of representing real valued numbers as integers. Prior using this technique, LPC was there to predict the features properly as shown in Figure 4.

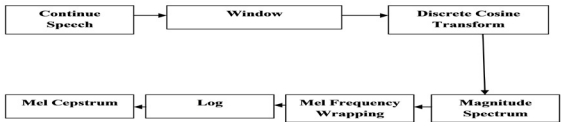

Figure 4. MFCC (Mel Frequency Cepstral Coefficient) Model.

2.4 ZCPA (Zero Crossing with Peak Amplitudes)

This technique[16,17] is used for feature extraction and it is based on Human Auditory System. To represent signal frequency information, it uses zero crossing method. Figure 5 represents diagram for detection block of zero crossing. It defines Band Pass Filters, Zero Crossing Detector, Peak Detector, Interval Histogram and non linear compression. Its output feature vector can easily represent the pitch and features of speech.

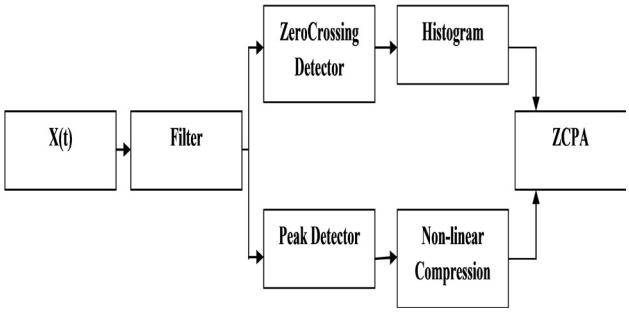

Figure 5. ZCPA (Zero Crossing with Peak Amplitudes) Model.

2.5 DTW (Dynamic Time Warping)

This technique[17,18] is used to measure the similarity index between two time series based on Dynamic Programming. It is also used to find the proper alignment between two time series.

3. Conclusion

Based on the analysis of comparison done between different types of feature extraction techniques, we draw a table named as Table 1 in which recognition rate is shown in correspondence with application to find out better technique for continuous speech recognition.

Table 1 shows individual as well as integrated speech recognition systems along with their recognition rate. From this table, it may be concluded that individual methods don't give good performance and recognition rate. So, an integrated approach may be used to provide the better results.

Table 1. Recognition Rate Analysis

Technique used	Recognition Rate
LPC	82.3%
MFCC	86.67%
ZCPA	38.5%
DTW	75.19%
MFCC+LPC	85%
MFCC+PLP	83%

We also attempted to analyse the above mentioned parameter on the basis of neural networks as a classifier as shown in table 2. A performance graph depicting the recognition rate of all the voice recognition techniques with different types of classifiers named as HMM (Hidden Markov Model), DTW (Dynamic Time Wrapping), VQ (Vector Quantization), SVM (Support Vector Machine) and Euclidean distance is shown in Figure 6 for vector quantization method in which x is the input vector, x_j is the j^{th} component of input vector and y_{ij} is the j^{th} component of vector y_i which clearly states that the hybrid approach may provide better results as compared to individual methods.

Table 2. Techniques with Different Classifiers

Technique	HMM	DTW	VQ	SVM	Euclidean distance	Recognition Rate
MFCC	88	91	96	86	32	86
PLP	79	87	79	69	28	84
LPC	80	76	66	83	25	82
MFCC+PLP	98	92	91	52	70	89

In this paper, we have studied about number of feature extraction techniques for voice recognition. It was noticed that performance and recognition rate is not so good by using individual techniques because it does not provide better recognition rate while taking into consideration the security of an individual living alone at home. After the comparative analysis, it is concluded that there is a need to develop a better combined approach which may provide better recognition rate as compared to individual methods. This paper will help the researchers to know the basic difference between the explained feature extraction techniques.

4. References

1. Saxena A, AmitSinha, Chakrawati S, Charu S. Speech Recognition using Mat lab. International Journal of Advances in Computer Science and Cloud Computing. 2013 Nov; 1(2):1–5.
2. Vibha. MFCC and its Applications in Speaker Recognition. International Journal on Emerging Technologies. 2010 Feb; 1(1):19–22.
3. Kurzekar PK, Deshmukh RR, Vishal B, Shrishrimal PP. A Comparitive study of Feature Extraction Techniques for Speech Recognition System. International Journal of Innovative Research, Engineering and Technology. 2014 Dec; 3(12):1–11.
4. Prithvi P, Kishore T. Comparative Analysis of MFCC LFCC RASTA-LPP. International Journal of Scientific Engineering and Research. 2016 May; 4(5):1–4.
5. Patil RP, Mulla MR. A Review Design and Implementation of Image Acquisition and Voice Based Security System. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. 2015 Mar; 4(3):1–6.
6. Dave N. Extraction Methods LPC and MFCC in Speech Recognition. International Journal for Advance Research in Engineering and Technology. 2013 Jul; 1(6):1–5.
7. Sinha S, Jain A. Spectral and Prosodic features based Speech Pattern Classification. International Journal Applied Pattern Recognition. 2015; 2(1):96–109. Crossref
8. Swamy S, Ramakrishnan KV. An Efficient Speech Recognition System .Computer Science and Engineering an International Journal. 2013; 3(4):1–7. Crossref
9. Kumar K, Jain A. A Hindi Speech Recognition System for Connected Words using HTK. International Journal Computational Systems Engineering. 2012; 1(1):1–8. Crossref
10. Franke T, Tong C, Ashe MC. The Secrets of Highly Active Older Adults. Journal of Aging Studies. 2013; 27(4):398–409. Crossref PMid:24300060
11. Kumar J, Prabhakar OP, Sahu NK. Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification systems a Review. International Journal of Innovative Research in Computer and Communication Engineering. 2014; 2(1):2760–9.
12. Kavya BM, Chakrasali SV. Performance Analysis of MFCC and LPC techniques in Kannada Phoneme Recognition. International Journal of Advances in Electrical Power System and Information Technology. 2015; 1(2):21–5.
13. Singh V, Kumar V, Tripathy N. A Comparative Study of Feature Extraction Techniques for Language Identification. International Journal of Engineering Research and General Science. 2004; 2(3):286–91.
14. Honig F, Stemmer G, Hecker C, Brugnara F. Revising perceptual linear prediction. Interspeech. 2005; 2997–3000.
15. Abdalla MI, Abokar MH, Gaafar ST. DWT and MFCCs Based Feature Extraction Methods for Isolated Word Recognition. International Journal of Computer Applications. 2013; 69(20):21–6. Crossref
16. Madan A, Gupta D. Speech Feature Extraction and Classification A Comparative Review. International Journal of Computer Applications. 2014; 90(9):20–5. Crossref
17. Kepuska VZ, Elharati HA. Robust Speech Recognition System using Conventional and Hybrid Features of MFCC LPC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions. Journal of Computer and Communications. 2015; 3(6):1–9. Crossref
18. Gamit MR, Dhameliya K. Isolated Words Recognition Using MFCC, LPC and Neural Network. International Journal of Research in Engineering and Technology. 2015; 4(6):146–9. Crossref