Proceedings of the First

University Journal of Research and Innovation

December, 2019

Organized by
University of Computer Studies (Pakokku)

Proceeding of

The First University Journal of Research and Innovation 2019

December , 2019

Organized by

University of Computer Studies (Pakokku)

Department of Higher Education ,

Ministry of Education , Myanmar

University Journal of Research and Innovation

Volume 1, Issue 1 2019

University Journal of Research and Innovation 2019

Volume 1 , Issue 1 , 2019

This journal and individual papers published at www.ucspkku.edu.mm.

Copies:110

All research papers in this journal have undergone rigorous peer-reviewed which is published annually. Full papers submitted for publication are refereed by the Associate Editorial Board through an anonymous referee process.

The authors of the paper bear the responsibility for their content.

Papers presented at the First University Journal of Research and Innovation(UJRI), University of Computer Studies (Pakokku), December 2019.

UJRI 2019 Editorial Board

UJRI 2019 Editorial Board

Proceedings of

The First University Journal of

Information and Computing Science 2019

December, 2019

Contents

Artificial Intelligence & Machine Learning

Natural Language Processing

Big Data Analysis

Parallel & Distributed Computing

Image Processing

Human Computer Interaction

Database Management System & Information Retrieval

Network & Security

Data Mining & Machine Learning

Digital Business Management

Electronics

Embedded System

Cloud Computing

Software Engineering and Web Engineering

Digital Signal Processing

Theoretical Nuclear Physics

Material Science

Mathematics

Experimental Nuclear Physics

English Language

Myanmar Language in Literature

Author Index

Text Independent Speaker Identification System By Perceptual Linear Prediction (PLP)

Aye Thida Win[1] and Khin Hninn Phyu[2]
Faculty of Information Science Department, university of computer studies, kyaing tong
ayethitarwinn07@gamil.com
Faculty of Computer Science Department, university of computer studies, kyaing tong
khinhninnphyu32@gmail.com

Abstract

Speaker recognition is the identification of a speaker from characteristics of voices. Automatic speaker recognition system generally consists of two main phrases; feature extraction and recognition. In speaker recognition, feature extraction phrase is the most imperative phases. The concise and effectiveness of extracted features effects the recognition rate. This paper presents the speaker recognition system based on PLP (Perceptual Linear Predictive) for feature extraction phrase and VQ/GMM (Vector Quantization/Gaussian Mixture Model) techniques for recognition. To evaluate the proposed method, a popular feature extraction method LPC is also evaluated for comparison. The system tested on 8 speakers (both English and Myanmar speech samples and evaluated for both text-dependent and text-independent speaker identification. Comparison of the proposed approach shows that the proposed method got better recognition rate and obtained promising results.

Keywords: PLP, LPC, analysis, feature, extraction, speaker, identification

1. Introduction

Speaker recognition is the identification of a person from characteristics of voices. Speaker recognition technologies have wide application areas such as the detection of Earthquakes, security, forensics and authentication etc. Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech wave. Speaker recognition generally consists of training phase and a testing phase. In training phrase a user enrolls by providing voice samples to the system. The voice characteristics of every speaker are different. The system extracts speaker-specific information from the voice samples and builds the voice model database. In the testing phrase, the system compares the inputted voice sample with the previously enrolled speakers' features model and subsequently, makes a decision.

Speaker recognition could be categorized as speaker verification and speaker identification. Speaker verification is the process of accepting or rejecting the identity claim of a speaker. In speaker verification, the obtained features are only compared with the stored features of the speaker he/she claimed to be. Speaker identification is the process of determining which registered speaker provides a given utterance. In speaker identification, the obtained features are compared with all the speakers' features in voice model database.

Speaker identification system can also be classified into two types.

(1) Text-dependent system: The recognition of the speaker's identity is based on his or her speaking one or more specific phases or word. Text-dependent that is when the same text is spoken on both training and test phases.

(2) Text-independent system: Speaker models capture characteristics of somebody's speech which show up irrespective of what one is saying. Text-independent that is there is no restriction of voice sample and it could be differ in training and test phases. The general phrase of speaker identification system is shown in Figure 1.

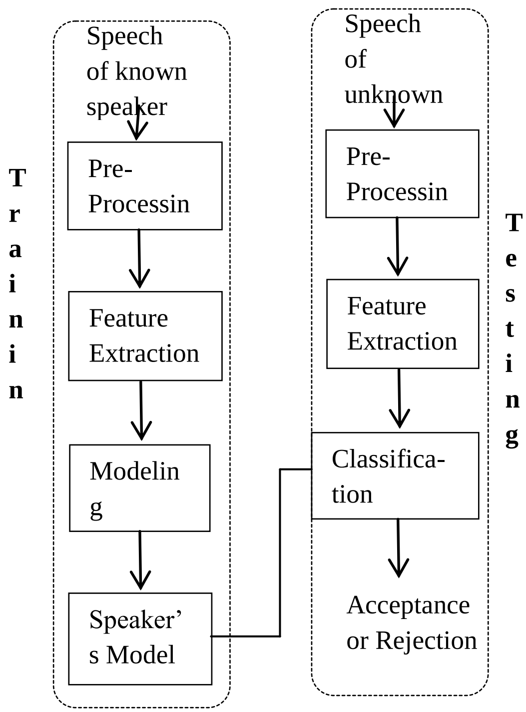

Figure 1. Speaker identification Phases

(1) Training or Enrolment phase

(2) Testing or Identification phase

In the training phase, each registered speaker has to provide samples of their speech so that the system can build or train a reference model for that speaker. During the testing phase, the input speech is matched with stored reference model(s) and recognition decision is made.

This paper presents speaker recognition both in text-dependent and text-independent. The main focus of this paper is to analyze the better feature extraction method for English and Myanmar speech sample. And compare the proposed approach with conventional common used feature extraction method LPC. The PLP(Perceptual Linear Predictive) method is used to extract the voice characteristic features in the first phrase. And then hybrid VQ/GMM (Vector Quantization/Gaussian Mixture Model) method is used to identify the speakers based on their speech samples into respective classes and recognition. The experimental result shows that the accuracy is better and the execution time of the system is less by using PLP rather than other feature extraction methods. The rest of this paper is organized as follows: Section 2 briefly describes the feature extraction method used in our system. Section 3 explains the feature matching, Section 4 presents the decision making, Section 5 presents the experimental results and finally section 6 presents the conclusion.

2. Feature extraction

The objective of automatic speaker recognition is to identify the speaker by extracting features, characterizing and recognizing the information contained in the speech signal. Feature extraction is the process of extracting information from speech in a compact form to model the speaker. The purpose of this scheme is to convert the speech waveform into a set of features or rather feature vectors for further analysis. Different techniques are used for feature extraction such as Linear Predictive Coding (LPC), LPCC, MFCC, PLP. LPC is device used generally in audio signal processing and speech in compressed representing the spectral envelope of speech in compressed form by using the information of a linear predictive model. Feature extraction in speech recognition is the way towards holding valuable data of the signal while disposing of repetitive and undesirable information. The main point of feature extraction is obtaining the set of features with low rate of change in order to keep the computational feasible. In the proposed system, Perceptual Linear Prediction (PLP) is used to extract features from the input speech while the training phase. To compare the effectiveness of the method, we evaluate the LPC method that is commonly used and powerful speech analysis technique.

2.1. Linear Predictive Coefficient (LPC)

Linear Predictive Coding (LPC): LPC is the good analysis technique for extracting features, determining the basic parameter and computational model of speech. In LPC, speech samples can be approximated as a linear combination of past speech samples. The number of samples is referred to as the order of LPC. LPC has the capability for compression, synthesis, identification of the speech signal. The LPC calculates a power spectrum of the signal. It is used for formant analysis. LPC is one of the most powerful speech analysis techniques and it has gained popularity as a formant estimation technique. While we pass the speech signal from speech analysis filter to remove the redundancy in signal, residual error is generated as an output. It can be quantized by smaller number of bits compare to original signal. So now, instead of transferring entire signal we can transfer this residual error and speech parameters to generate the original signal. A parametric model is computed based on least mean squared error theory, this technique being known as linear prediction (LP). By this method, the speech signal is approximated as a linear combination of its p previous samples. In this technique, the obtained LPC coefficients describe the formants. The frequencies at which the resonant peaks occur are called the formant frequencies. Thus, with this method, the locations of the formants in a speech signal are estimated by computing the linear predictive coefficients over a sliding window and finding the peaks in the spectrum of the resulting LP filter. We have excluded nth coefficient and used next ten LPC Coefficients as shown in Figure [2].

2.2. Perceptual Linear Prediction (PLP)

PLP provides a representation corresponding to a smoothed short-term spectrum that has been compressed and equalized much as done in human hearing. It can be assumed similar to mel-cepstrum based features. In PLP technique, several well-known properties of hearing are simulated by practical engineering approximations, and the resulting auditory like spectrum of speech is approximated by an autoregressive all-pole model. PLP provides reduced resolution at high frequencies that indicates auditory filter bank based methods, yet provides the orthogonal outputs that typify cepstral analysis. PLP uses linear predictions for spectral smoothing; hence the name is perceptual linear prediction. The different steps of PLP analysis are as follows.

1. Power spectral estimate for the windowed speech signal is computed. This is done by windowing the analysis region with Hamming window, calculating the FFT and computing its squared magnitude.
2. The power spectrum within overlapping critical band filter responses is integrated. For PLP, trapezoidal shaped filters are applied at 1-bark intervals, where the bark axis [12] is derived from the frequency axis by using a warping function from Schroeder, given in

Equation 1. The Bark scale is linear at low frequencies and logarithmic at high frequencies.

$$\Omega(\omega) = 6\ln\{\frac{\omega}{1200\pi} + [(\frac{\omega}{1200\pi})^2 + 1]^{0.5}\} \qquad (1)$$

Where ω is angular frequency in radians/seconds. This effectively compresses the higher frequencies into a narrow band. The critical band masking, symmetric frequency domain convolution on the Bark warped-frequency scale then allows low frequencies to mask the high frequencies while at the same time smoothing the spectrum an effect consistent with the psycho-acoustic results.

3. The spectrum is pre-emphasized to approximate the unequal sensitivity of human hearing at different frequencies. This step is implemented as an explicit weighting of the elements of critical band spectrum.

4. Spectral amplitude is compressed. The effect of this step is to reduce amplitude variations for the spectral resonances.

5. An inverse DFT is performed. As a result of this step autocorrelation coefficients are obtained, but these coefficients are from a compressed spectrum. Since the power spectrum values are real and even, only the cosine components of inverse DFT are to be calculated.

6. Spectral smoothing is performed. This step is done by solving the autoregressive equations constructed from the autocorrelations of the precious step.

7. The autoregressive coefficients are converted to cepstral variables as shown in Figure 2 [14].

3. Feature matching

Feature matching involves identification of the unknown speaker by comparing the extracted features from his/her voice input with the ones from a set of known speakers. The feature matching techniques used in speaker identification includes DTW, HMM, GMM, VQ and so on. The proposed system uses hybrid VQ/GMM model because execution time is less in VQ/GMM than baseline GMM. Moreover, the speed of operation is significantly increased because the number of features is reduced which consequently decrease the complexity of identification system as shown in Figure 3.

8.

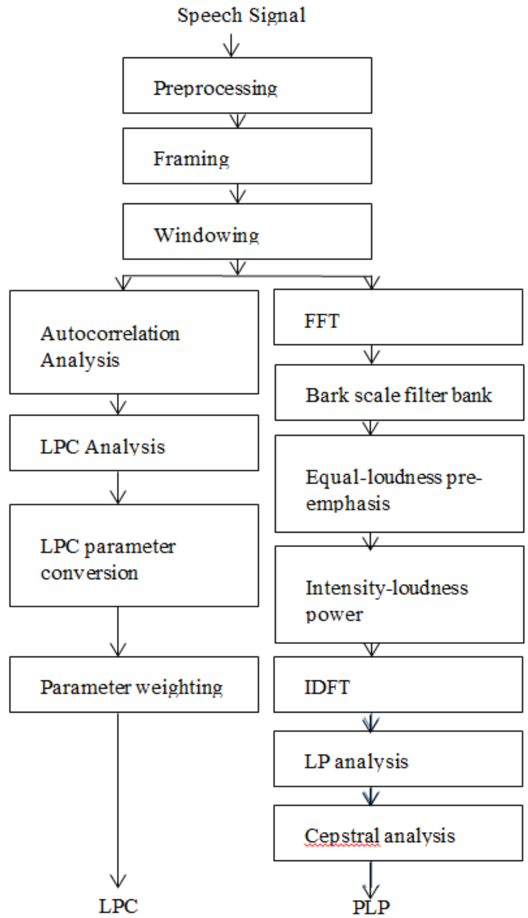

Figure 2. Block Diagram of PLP and LPC

1. Vector Quantization

Vector Quantization (VQ) is a classical technique for compressing signal processing capabilities. It can model probability density by distributing prototype vectors. It was used to compress data [1]. It works by dividing a large set of points (vectors) into groups for having the same number of points closest to them. The central point represents each group as in the k-means and other pooling algorithms.

The d-dimensional vectors are mapped in the vector space R_k into a finite set of vectors $Q=\{q_i: i=1,2,\dots, n\}$ by vector quantizer. VQ works well for lossy data compression. The main elements in the mapping are code word and code book. The each vector q_i associated with quantizer is called a code word or a code vector and the set of all the code words are called a codebook. The region that belongs to nearest neighbour q_i is Voronoi region [1].

2. Gaussian Mixture Model

A Gaussian mixing model is a probabilistic model that assumes that all data points are produced from a mixture of a finite number of Gaussian distributions with unknown parameters. Blending models are the generalization of k-means clusters to incorporate information about the covariance structure of the data as well as the centers of latent Gaussians. The importance aspect of the any accent modelling is to collect and find the weight of the mean vector of each accent and mixture from the training speed utterance. The parameters of Gaussian Mixture Model are estimated with the help of maximum likelihood Gaussian Mixture Model [1]. The probability density function of arbitrary shape is approximated [1].

4. Decision making

There are two decisions for speaker identification: acceptance or rejection. If the input speech in the testing phase is same with the speech in the speaker database while training, the system accepts and identifies the speaker. Otherwise, the speaker will be rejected. A voice match will produce logic 1 while a mismatch, logic 0.

5. Experimental Results

To evaluate the performance of the proposed method, we collect a large number of speech signals, total 250 English and Myanmar speech from 8 different speakers. Each speaker speaks 15 short speeches in different emotions.

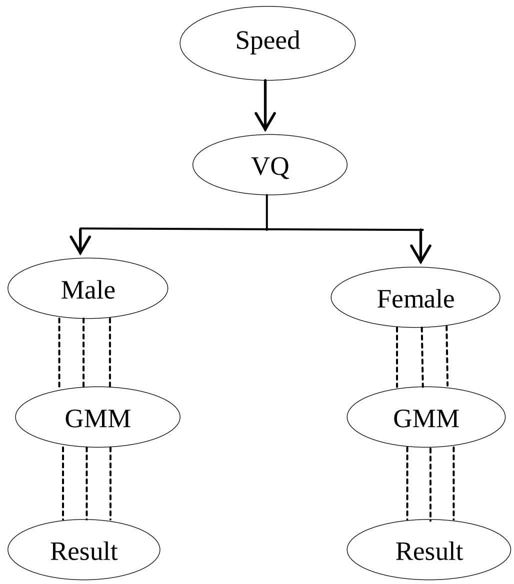

Figure 3. Classification Process using VQ and GMM

All speech samples are recorded using a microphone in normal environment (not too noisy) at 20 KHz sampling frequencies and are stored in wav format files. A set of 180 utterances are used for training the recognition model and the 70 utterances are used for testing. The performance of identification system are given in terms of Correct Identification Rate (CIR) which is :

$$\text{CIR} = \frac{\text{no. of tested correctly recognized}}{\text{total no. of trials}} \text{X } 100 \quad (1)$$

The CIR accuracy is calculated by dividing number of correct recognition to total no. of trials by speaker. Correct recognition means the number of times the speaker is identified by system. The total number of trials for each speaker can be taken anything but for more accuracy we have taken it 15. Example: speaker 1, the total number of trials is 15. If the total number of correct recognition is 12, then the CIR will be 80%.

We build the voice feature model using the PLP and LPC and then hybrid VQ/GMM methods is used to identify the speaker. We use the 30ms Hamming window advanced by 10ms for each frame.

Table1. CIR result for text independent

Speakers	Text independent	
	PLP_CIR(%)	LPC_CIR(%)
P1	80	66.11
P2	87	73.65
P3	73	59.28
P4	93	81.98
P5	87	73.24
P6	80	66.69
P7	73	59.18
P8	87	73.24
P9	80	66.69
P10	73	60.21

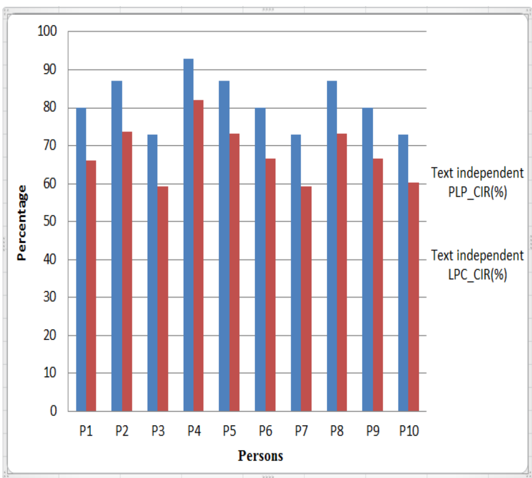

Figure 4. Shows the CIR(%) result of each speaker.

Table 2. CIR result for text dependent

Speakers	Text dependent	
	PLP_CIR(%)	LPC_CIR(%)
P1	89	77
P2	95	83
P3	71	59
P4	85	73
P5	82	70
P6	94	80
P7	67	55
P8	83	71
P9	77	65
P10	84	72

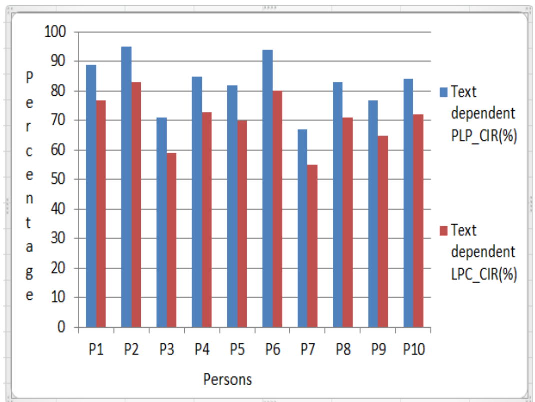

Figure 5. Shows the CIR(%) result of each speaker.

Table 1 and table 2 shows result for text independent and text dependent speaker recognition respectively. The graph for text independent speaker recognition is shown in Figure 4. Figure 5 shows the graph for text dependent recognition. By examining the recognition results, PLP method got better accuracy than LPC method. Especially , PLC offered high accuracy in processing Myanmar speech where LPC cannot. By considering the processing time factor, it is noticed that the LPC method was found to be time consuming when compared with PLP.

6. Conclusion

In this paper, we have investigated the use of the PLP method for feature extraction and VC/GMM method for speaker recognition using a text-independent and text-dependent system. We found that the use of the PLP would definitely better the performance of a speaker recognition system both for English and Myanmar speech. Formulation of the automatic speech recognition system is a challenging due to various effecting factors of human's speech like emotions, diseases, noises, session variably, etc. We plan to extend the proposed system to handle speeches in noisy environment and also in various emotions.

ACKNOWLEDGEMENTS

I would like to express my heartful thanks to Dr. Tin Myat Htwe, Rector of Computer University of Kyaing Tong for her great advices towards the success of this paper.

I also thanks to Daw NingLunMang, Lecturer, Head of Department of Mathematics, University of Computer Studies, Kyaing Tong who helps me with invaluable supporting. Finally, I would like to acknowledge and extend my thanks to several people who are great interested in my paper and read it in their precious time.

References

[1] A. Mauria, D. Kumar and R.K. Agarwal, "Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach", Peer-review under responsibility of the scientific committee of the 6th international Conference on Smart Computing and Communication, Kurukshetra, India,December 2017, pp 7-8,

[2] Athiramenon.G and Anjusha.V.K, "Analysis of Feature Extraction Method for Speech Recognition", IJISET- International Journal of Innovative Science, Engineering & Technology, April 2017, ISSN(Online) pp 2348-7968,Vol.4 Issue 4

[3] Brookes, M., Voicebox: Speech Processing Toolbox for Matlab [on line], Imperial College, London, available on the WorldWideWeb:http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[4] Daniel Jurafsky, and James H. Martin, "Speech and language Processing", PearsonEducation, 2000

[5] Douglas O'Shaughnessy, "Interacting with Computer by Voice Automatic Speech Recognition and Synthesis", Sept 2003Y, pp.1272-1305, Proceeding of the IEEE, Vol.91, No.9

[6] FatmazohraChelali, Amar.Djeradi and Rachida.Djeradi, " Speaker Identification System based on PLP Coefficients and Artificial Neural Network", London, U.K, July 2011, pp 6-8,Proceedings of the world congress on engineering 2011 vol II

[7] H.Fletcher, " Auditory Pattern Review of Modern Physics" , Jan 1940

[8] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech", Journal Acoust. Soc. Amer,199, pp. 1738–1752, vol. 87, no. 4

[9] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy", Magnetism G.T. Rado and H. Suhl, Eds, New York: Academic,1963, pp. 271-350, vol. III

[10] J. Goldstein, L. D. Harmon, and A. B. Lesk, "Identification of Human Faces", May 1971, pp 748-760,No. 5, Vol. 59, Proc. IEEE

[11] Ms. ManpreetKaur and Prof. Puneet Mittal, "Speaker Recognition Based on Feature Extraction in Clean and Noisy Environment", IJRASET(UGC Approved Journal), April 2018,Vol.6 Issue IV, ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887

[12] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[13] Namrata Dave, “Feature Extraction Methods LPC, PLP and MFCC In Speed Recognition”, www.ijaret.org, July 2013,Vol.1 Issue VI.

[14] P.Kumar and M.Chandra, “Speaker Identification Using Gaussian Mixture Model”, MIT International Journal of Electronics and Communication Engineering, Jan 2011, pp 27-30,Vol. 1, ISSN 2230-7672

[15] Poddar, Arnab; Sahidullah, Md; Saha, Goutam (March 2018). "Speaker Verification with Short Utterances: A Review of Challenges, Trends and Opportunities". *IET Biometrics*. **7** (2): 91–101. doi:10.1049/iet-bmt.2017.0065.

University of Computer Studies (Pakokku)
Department of Higher Education
Ministry of Education
Myanmar