**International Journal of Advanced Engineering and Technology**
**ISSN: 2456-7655; Impact Factor: RJIF 5.54**
**Received: 14-09-2019; Accepted: 16-10-2019**
**www.newengineeringjournal.com**
**Volume 3; Issue 4; October 2019; Page No. 54-55**

# Speaker identification using vector quantization and dynamic time warping

**Pardeep Sangwan**
Department of ECE, Maharaja Surajmal Institute of Technology, New Delhi, India

**Abstract**
Speaker Recognition is a system in which an individual is recognized based on voice signals. In this article, we provide a brief overview of the histopry of the methodology used in speech recognition for pattern identification. A method for modelling a speaker recognition system was also addressed and suggested, which involves the pre-processing stage, the extraction phase of features and the classification phase of patterns. LPCC and MFCC are used in this method as text-related speech recognition and the experiment uses vector quantization and dynamic time warping (DTW) to compare a speaking identity recognition rate for LPCC, MFCC or a combination of LPCC and MFCC. This reflects the higher acceptance levels for LPCC and MFCC.

**Keywords:** Speaker recognition, LPCC, MFCC, VQ, DTW

## 1. Introduction
Speaker identification is a method to determine who talks utilizing the data inserted into sound utterances automatically. The recognition of speakers can be further categorised as: the identification of speakers and the verification of speech speakers. Speaker recognition specifies the speaker has a particular utterance from the registered speaker while speaker confirmation is the approval or rejection of the speaker's stated identity. The approach to speech recognition is the processing of a specific audio signal. LPCC represents the differences between biological structure of the human vocal tract and MFCC is dependent on the characteristic of the non-linear frequency of the human ears [1]. The present work proposes a speaker identification system utilizing VQ [2] and DTW which utilizes combined features of LPCC and MFCC as reference point.

## 2. Extracting Features
### 2.1 LPCC
It combines strengths of LPC and cepstral analysis as well as improving the precision of speech recognition applications. LPCC is identical to the smooth envelope of the speech log which allows speaking specific features to be extracted.
LPC is changed into cepstral coefficients by means of the subsequent recursive formula

$$c1 = a1 \quad (1)$$

$$C_n = a_n + \sum_{k=1}^{n-1}(1 - \frac{k}{n})a_k c_{n-k} \quad (2)$$

Where a and C are the *i* th-order cepstrum and linear predictor coefficient, respectively.

### 2.2 MFCC
It focuses on the characteristic of the non-linear frequency of the human ears, and the scale of the Mel frequency corresponds to the relationship of the logarithmic distribution of the real frequency as a whole and to the characteristic of the human ears. Mel frequency's idiographic relationship with real frequency is as follows:

$$M = 2595\, log_{10}(\frac{f}{700} + 1) \quad (3)$$

MFCC begin by breaking the sound signal into smaller frames and windowing each frame to eliminate the effect of discontinuity at frame edge. It converts signal to a frequency domain in the Fast Fourier transform (FFT) phase and then Mel frequency deforms the frames. After the frames are warped by Mel frequency, logarithmic signals are passed to IDFT which converts it again to time domain. Thirteen coefficients called MFCC were obtained for each frame as a result of the final step. The 0th coefficients of each frame are discarded as these merely reflects average energy in a frame and provides no data that can be used. Vectors in 12 dimensions are obtained for each frame as the output of the extraction step of the function. The vectors are used to compare and match the feature sets against the previously stored model in pattern matching / classification technique.

## 3. Speaker recognition techniques
VQ is a technique for processing digitalized signal by mapping vectors to a finite number of regions in that space from a large vector space. Every area is called a cluster, and its centre called a code word will represent it. A codebook is considered the set of all code words. The first step, training, is setting up N codebook of N speaker which don't overlay in the space. It can be assumed that no. of code word of N-codebook is M. DTW utilizes dynamic programming. LPCC and MFCC are stored and processed by frame-based speech processing. Comparison of reference and test templates done with DTW template matching algorithm; DTW measures the distance between the referenced template and test model and the minimum range template is the best matching result [3].

### 3.1 SR using Combined Features
MFC represents the characteristic of non-linear frequency response of human ears, LPC depicts variations in the biological structure of the human vocal tract, and their

differential one-order coefficient each explain dynamic characteristic. When using MFCC parameter to identify, the system tends to judge this speaker as the system's legitimate speaker if the key is mixed with a speaker. So, using LPCC which represents the difference of the human vocal tracts biological structure and its one order which differentiates a function to improve system security. First, the DTW arithmetic template matching is applied in the recognition process to two combined feature vectors. Then set the template matching distance threshold p to reduce the miscarriage of justice. If the model matching range d is greater than p, even if he was a legitimate speaker, the speaker is deemed to be an illegitimate speaker. Compare the minimum range script code I j. If I'm equivalent to j, even if he was an illegitimate speaker, the speaker is treated as a legitimate speaker.

## 4. Experimental setup

We have utilized database including 50 speech utterances of speakers (20 M and 30 F) and recording is done for the training and testing of the developed system in two different sessions with a gap of two weeks. Former speech recordings are used for training and latter recordings are used for test data. Recordings were done with microphones at 8000Hz and 11025Hz. This study used MATLAB 7.0 as the development environment and performed three types of experiments with DTW arithmetic against different function vectors. Accuracy [4] is used in this paper as the criterion for assessing the system's recognition quality.

Experiment one: Train the prototype and test the process with a 12-order LPCC coefficient and its one-order distinction* LPCC to form the entire speaker recognition system.

Experiment two: Train the prototype and test the process with a 16-order MFCC coefficient and its one-order distinction* MFCC to shape the entire speaker recognition system.

Experiment three: Use the combination of experiment one of two forms of feature vector and experiment two and change the DTW arithmetic recognition output.

## 5. Results

Table 1, Table 2 and Table 3 show the results of Experiment 1, Experiment 2 and Experiment 3 respectively. The result of utilizing LPCC and ΔLPCC or MFCC and ΔMFCC alone are different than LPCC, MFCC, ΔLPCC and ΔMFCC combination. In addition, ΔLPCC and ΔMFCC represent complex characteristics of sound and vocal tract, so that combined vectors represents improved individual characteristics of speaker.

**Table 1:** Results of 1st experiment

| Modelling Technique | Feature | Experiment | Sampling Frequency | Accuracy |
|---|---|---|---|---|
| VQ+DTW | LPCC, ΔLPCC | 1st | 8000 Hz | 87.65% |
| VQ+DTW | LPCC, ΔLPCC | 1st | 11025 Hz | 95.52% |

**Table 2:** Results of 2nd experiment

| Modelling Technique | Feature | Experiment | Sampling Frequency | Accuracy |
|---|---|---|---|---|
| VQ+DTW | MFCC, ΔMFCC | 2nd | 8000 Hz | 91.25% |
| VQ+DTW | MFCC, ΔMFCC | 2nd | 11025 Hz | 96.27% |

**Table 3:** Results of 3rd experiment

| Modelling Technique | Feature | Experiment | Sampling Frequency | Accuracy |
|---|---|---|---|---|
| VQ+DTW | Combination | 3rd | 8000 Hz | 95.85% |
| VQ+DTW | Combination | 3rd | 11025 Hz | 98.5% |

## 6. Conclusion

For speaker identification, the paper used various pre-processing stages before extraction of the feature was studied and implemented. The model has been developed to study and test different methods of voice function extraction, such as LPCC and MFCC, for their suitability in speech recognition. The paper used the VQ and DTW approach to determine the personality of a speaker by extracting the combination of LPCC, MFCC, * LPCC and* MFCC and contrasting the strengths and weaknesses of using LPCC, MFCC, * LPCC, * MFCC and their combination as speech characteristics. The experiment showed that the combination of LPCC, MFCC, * LPCC and* MFCC improved the recognition level efficiency.

## 7. References

1. Dey S, Kashyap K. A Dynamic-threshold Approach to Text-dependent Speaker Recognition using Principles of Immune System, IEEE, 2015.
2. Khushboo Desai S, Pujara H. Speaker Recognition from the Mimicked Speech: A Review, IEEE, 2016.
3. Kishori Ghule R, Ratnadeep Deshmukh R. Automatic Speech Recognition System Using MFCC and DTW for Marathi Isolated Words, IJTEEE, 2015.
4. Narang S, Divya Gupta. Speech Feature Extraction Techniques: A Review, IJCSMC. 2015; 4(3):107-114.