VOL. 14, NO. 9, MAY 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences
©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.
www.arpnjournals.com

COMPARATIVE EVALUATION OF DIFFERENT HFCC FILTER-BANK USING VECTOR QUANTIZATION (VQ) APPROACH BASED TEXT DEPENDENT SPEAKER IDENTIFICATION SYSTEM

Mariame Jenhi[1], Ahmed Roukhe[2] and Laamari Hlou[1]
[1]Laboratory of Electrical Engineering and Energy System Faculty of Science, University Ibn Tofail, Kenitra, Morocco
[2]Laboratory of Atomic, Mechanical, Photonics and Energy Faculty of Science,University Moulay Ismail, Meknes, Morocco
E-Mail: mariame.jenhi@uit.ac.ma

ABSTRACT

In the feature parameterization (FP) stages, cepstral coefficients based Short-Term Fourier transform (STFT) have been regarded as one of the most significant features used in speaker identification (SI) system. The widest FP techniques used to extract this feature are the one that attempts to replicate the psycho-acoustic properties of the human auditory system like the currently proposed Human Factor Cepstral Coefficients (HFCCs) based filterbank analysis, which is considered as a modification of the well-known Mel Frequency Cepstral Coefficients (MFCC) approach. Typically, the HFCC process modifies the analysis of the classic Mel scale filterbank in MFCC, using the Equivalent Rectangular Bandwidth (ERB). In this paper, we aim to investigate the HFCC feature extraction with varying the number of the HFCC-filterbank using 10, 20 and 40 filters, to find out how it does affect the identification accuracy of a text-dependent speaker identification system. Furthermore, in this stage, we evaluate the efficiency of the proposed system using the non-parametric Vector Quantization (VQ) speaker modeling approach based on the LBG clustering algorithm by generating a codebooks-size ($\mathcal{COD}_{\text{size}}$) of 1, 2, 4, 8, 16, 32 and 64. The results of the proposed work yielded an identification accuracy rate of 100% for 40 HFCC-filterbank.

Keywords: automatic speaker identification, human factor cepstral coefficients (HFCC), ERB (Equivalent Rectangular Bandwidth), vector quantization (VQ), codebook, Linde-Buzo-Gray (LBG).

1. INTRODUCTION

Over the last few years, development of Speech Processing technology has witnessed considerable interest and tremendous research progress, especially in voice-based authentication process (Speaker Recognition). In fact, speaker recognition technology is interpreted as a special pattern recognition task [1] and can be considered as a promising voice biometric application, which is a process of identifying a person based on his voice [2, 3]. The speaker's voice is able, obviously, to characterize the uniqueness of each individual thanks to various characteristics contained in the speech signal such as language, accent, physiological state, and identity [4]. Although, this technology can be considered as a multidisciplinary field of research that brings together, computer scientists, linguists, logicians and psychologists. The widespread use of speaker recognition has boosted many relevant applications, that makes benefit of the speaker's voice in order to gain easy access to different services, for instance, secure access voice control applications including voice calls dialing, telephone-banking, telephone-shopping, database voice access services, remote access to personal computers and voicemail [5].

Most Speaker Recognition processes were typically divided into speaker identification and verification [6]. As discussed in [6], the list of speakers to identify in the speaker identification process is actually known by the system. In fact, the system must be capable to decide, from a voice sample, which known identity of the system corresponds to the sample. On the contrary, in speaker verification stage [7], the identity claimed of a speaker is either accepted or instantly rejected. Moreover, the input of such processes can be divided into two subcategories: text-dependent and text-independent, where the so-called text-independent system ignores the linguistic content of the speech signal [8]. In contrast, a text-dependent system requires the speaker to use all or part of the speech content [8]. In this paper, we typically focus on the text-dependent speaker identification task.

The speaker identification analysis is mainly comprised of three distinct steps namely pre-processing, feature extraction (parameterization) and feature classification. The pre-processing step is performed before feature extraction step to ensure that the speech utterance contains accurate information that conveys the identity of the speaker. Since the selection of an efficient set of the acoustic feature is considered as a kernel step in the identification process, many researchers have been motivated to develop robust and reliable feature extraction process in order to produce a system that capable of identifying people exactly as the human auditory system does. Most state-of-the-art voice feature extraction techniques are categorized into two groups: (i) features based on linear prediction modeling and (ii) features based filterbank analysis [9]. Under the first category, Linear Predictive Coefficients (LPC) proposed in [10], Linear Predictive Cepstral Coefficients (LPCC) [11] and Predictive Linear Prediction (PLP) [12], while in the second category, we found the Mel-Frequency Cepstral Coefficients (MFCC) introduced by Davis and Mermelstein [13] which describes the spectral envelope [14], and Gammatone Feature Cepstral Coefficients (GFCC) [15, 16].

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

Following feature extraction, the speaker classification phase transforms the voice feature vectors to an identical representation. The objective of this stage is to generate speaker models using speaker-specific feature vectors. A number of different classifiers have been proposed in the literature, remarkably Support Vector Machines (SVM) [17], Artificial Neural Networks (ANN) [18], Gaussian Mixture Models (GMM) [19], Hidden Markov Models (HMM) [20] and Vector Quantization (VQ) [21, 22].

Although being motivated by the great performance of MFCC technique, that has provided better identification accuracy over the said feature extraction methods, it turns out that the relationship between the center frequency and the bandwidth of the critical bands is ignored in the design of the MFCC based filterbank analyses. Hence, in this line of work and due to the need to find parameters to improve MFCC and allow overcoming the current recognition challenges, Skowronski and Harris [23], propose a new form of parameterization called Human Factor Cepstral Coefficients (HFCC). The HFCC is considered as a modification of the well-known MFCC at filterbank design level [24], where human auditory system information is added and the filter bandwidth used in the filterbank is a parameter of free design, independent of the separation between filters and linked to the known relationship between the center frequency and the critical bandwidth of the human auditory system [24].

In essence, we intend in this paper to explore the influence of increasing the number of filters in the HFCC filterbank using 10, 20 and 40 filters for a text-dependent speaker identification system. To compare the performance of our presented system, the well-known Vector Quantization (VQ) classifier based Linde-Buzo-Gray (LBG) clustering algorithm [25] is used in this study due to his fast and not too complicated way of modeling speakers, using 1, 2, 4, 8, 16, 32 and 64 codebook-size.

The remainder of this paper is structured as follows: Section 2 gives an overview of the HFCC feature extraction model used to represent the speaker feature. In Section 3, the operation of VQ based Linde-Buzo-Gray (LBG) clustering algorithm is explained. Section 4 presents the discussed results. Finally, general conclusions and proposals for future work are presented in Section 5.

2. HUMAN FACTOR CEPSTRAL COEFFICIENTS (HFCC) FEATURE EXTRACTION METHOD

The selection of a good set of feature vectors is considered as the paradigm in the design of any speaker identification system. Although, the voice signal is processed to produce a new representation of the voice in the form of a sequence of vectors which, must represent the information contained in the envelope of the spectrum. This section describes the fundamentals of the new Human Factor Cepstral Coefficients (HFCC) feature extraction technique that uses as a base the parameterization process carried out in MFCC [22], modifying the analysis of the Mel filterbank, adding, for its design, information about the human auditory system [23]. The computation steps of the HFCC feature extraction method used for Speaker Identification system is illustrated in Figure-1.

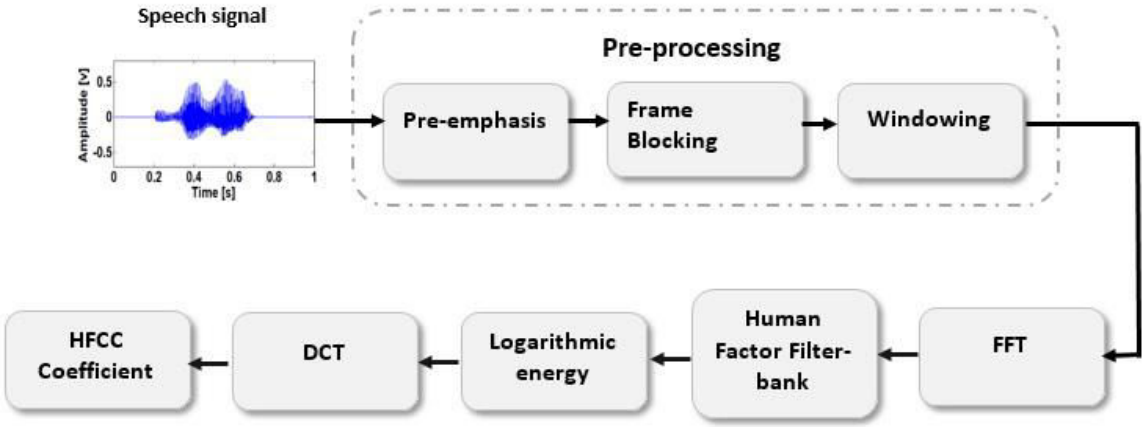

Figure-1. Generation process of Human Factor Cepstral Coefficients (HFCC) feature extraction.

2.1 Pre-processing

The pre-processing of an acoustic signal is considered as the basis step of the feature extraction system, including signal pre-emphasis, frame blocking and windowing. Figure-2 illustrates the speech waveform and the spectrogram of two-dimensional representation that displays time in the horizontal axis and frequency on the vertical axis of the first and second speakers speaking the same utterance. It is observed from the spectrograms in Figure-2, that the frequency range of the first speaker is clearly distinguishable from the second speaker. The differences are especially evident at low frequencies (<1500 Hz).

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

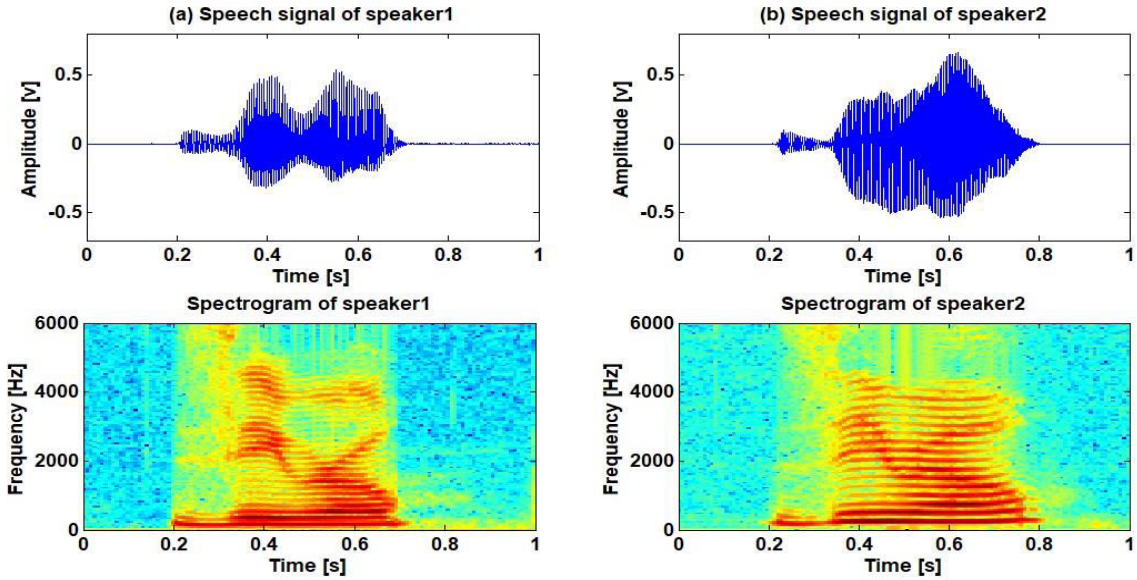

Figure-2. Speech signal and spectrogram showing characteristics of (a) first speaker, (b) second speaker.

2.1.1 Pre-emphasis

Pre-emphasis step has shown an improved performance during spectral analysis. It's purpose to raising up the energy in the higher frequency regions in the speech signal [26]. Here, we apply a pre-emphasis filter to the speech waveform $x(n)$ where the pre-emphasized signal $x_{pre}(n)$is related to the signal $x(n)$ by the following formula:

$$x_{pre}(n) = x(n) - \sigma_{pre}x(n-1),\ 0 \le n \le \mathcal{N}-1 \quad (1)$$

With σ_{pre} = 0.93 [27] themust commonly used value.

2.1.2 Frame blocking

We split the pre-emphasized signal $x_{pre}(n)$ into an overlapping series of quasi-stationary discrete segments. In fact, the speech signal changes continuously and needs to be broken into a sequence of segments according to the assumption that signal can be considered as stationary over short (approximately of 15-30 ms) segments. Let us denote $x_{pre}(n)$ be a pre-emphasized speech signal with a sampling frequency of f_s, and split into N_f frames each of length $\mathcal{F}$ samples with an overlap of $O_v = \mathcal{F}/2$ samples such that

$$x_{pre}(n) = \left\{x_{pre_1}(n), x_{pre_2}(n) \dots x_{pre_{n_f}}(n) \dots x_{pre_{N_f}}(n)\right\}$$

where$x_{pre_{n_f}}(n)$ denotes the ${n_f}^{th}$frame of $x_{pre}(n)$ and is

$$x_{pre_{n_f}}(n) = \left\{x[n_f * (O_v - 1) + i]\right\}_{i=0}^{\mathcal{F}-1} \quad (2)$$

2.1.3 Windowing

In this stage, a tapered window (Hamming or Hanning type) is needed to smooth the transitions at the beginning and the end of the $x_{pre_{n_f}}(n)$ frame and thus to limit the amplitude of the side lobes in the estimation of the spectrum. In this paper, we define $W_{ham}[n]$as a Hamming window given by Eq. (3).

$$W_{ham}(n) = \begin{cases} 0.54 - 0.46\,cos\left(\frac{2\pi n}{N}\right), & 0 \le n \le \mathcal{N}-1 \\ 0 \quad , & otherwise \end{cases} \quad (3)$$

Where $\mathcal{N}$ is the number of samples in each frame; Then the windowing result is the signal$\mathfrak{Q}_{n_f}$, given by Eq. (4)

$$\mathfrak{Q}_{n_f}(n) = x_{pre_{n_f}}(n).W_{ham}(n), 0 \le n \le \mathcal{N}-1 \quad (4)$$

2.2 FFT (Fast Fourier Transform)

During this step, the Fast Fourier transform (FFT) is applied to windowed signal $\mathfrak{Q}_{n_f}(n)$ for converting the ${n_f}^{th}$ frame of the pre-emphasized speech signal $x_{pre}(n)$ of $\mathcal{N}$ samples from the time domain into the frequency domain. Eq.5:

$$\mathfrak{X}_{n_f}(k) = \sum_{i=0}^{N-1} \mathfrak{Q}_{n_f}(n) * e^{\frac{-2j\pi ki}{\mathcal{N}}}, \quad (5)$$

Where $k = 0,1,2,\dots,\mathcal{N}-1$ is the index of the Fourier coefficients $\mathfrak{X}_{n_f}(k)$. In general, the values $\mathfrak{X}_{n_f}(k)$ are complex numbers and we consider only their absolute values (energy of the frequency).

2.3 Human factor cepstral coefficient filter-bank

The main characteristics of HFCC algorithm are that the filter bandwidth is linked to the known relationship between the center frequency and the critical bandwidth of the auditory system [23] and the center frequencies are equally spaced in Mel-frequency scale [7], where the relation between the Hertz scale frequency and its correspondence in Mel is the following:

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

$$\mathcal{M}_{el} = 2595 * \log_{10}(1 + f_{or}/700) \quad (6)$$

The filter bandwidth used in the HFCC filter bank is given by the equivalent rectangular bandwidth (ERB) introduced by Moore and Glasberg [28]. The ERB vs. the center frequency f_c expressed in Hz resulting in the following equation:

$$\text{ERB} = 6.23 \cdot 10^{-6} \cdot f_c^2 + 93.39 \cdot 10^{-3} \cdot f_c + 28.52 \quad (7)$$

Where f_c is the center frequency of each filter in Hz.

Figure-3 depicts a set of (a) 10 HFCC-filterbank, (b) 20 HFCC-filterbank and (c) 40 HFCC-filterbank respectively. Each filter's magnitude frequency response is of equal height at the center frequency, triangular in shape and covering the frequency range between [300, 8000 Hz]. Notice that the filters can overlap not only with their closest neighbors but also with the other distant filters.

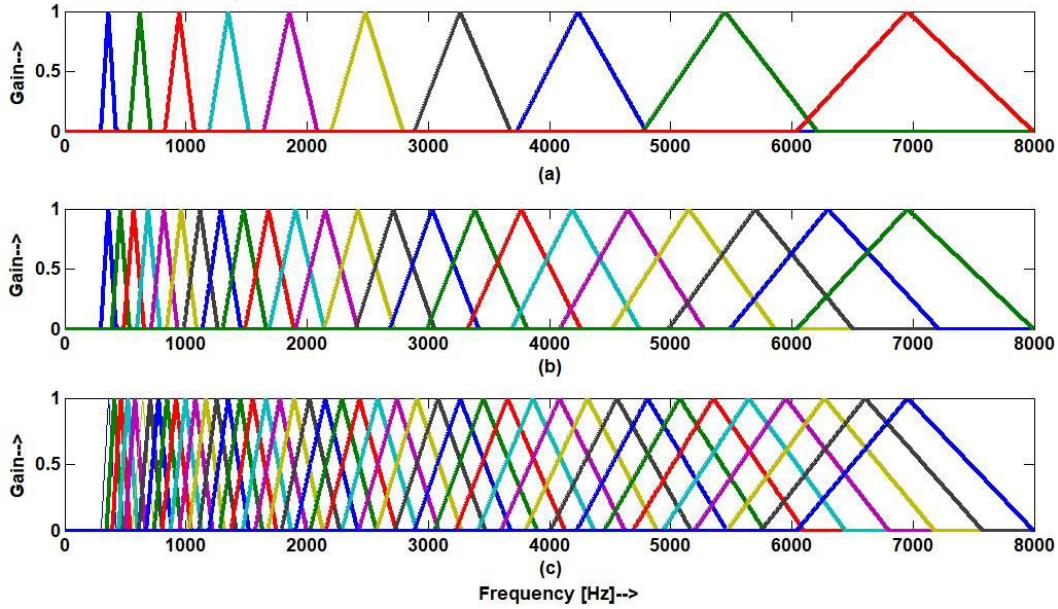

Figure-3. Illustration of Human Factor Cepstral Coefficients (HFCC) filterbank frequency responses for (a) 10- filterbank, (b) 20-filterbank, (c) 40- filterbank.

2.4 Discrete cosine transform (DCT)

Since, the high energy in the given filterbank corresponds to a high energy in the surrounding filters; here the Discrete Cosine Transform (DCT) tends to decorrelate the log-energies and generates K real coefficients:

$$\wp_{n_f}(n) = \frac{1}{K}\sum_{m=0}^{M_f-1} \mathcal{G}_{n_f} * \cos\left(\frac{\pi r(m+\frac{1}{2})}{\mathcal{F}}\right) \quad (8)$$

For $0 \leq r \leq \mathcal{F}$, where $\mathcal{G}_{n_f}$ represents the log-energy output of each of the HFCC frequencies and $\wp_{n_f}(n)$ represents the r^{th} HFCC of the ${n_f}^{th}$ frame of the pre-emphasized signal.

3. SPEAKER MODELING

In speaker modeling phase the objective is to build a model for each speaker using speaker-specific feature vectors. This section describes the fundamentals of the VQ speaker modeling method used in this paper.

3.1 Application of the Vector Quantization (VQ) as feature classifier

In terms of it simple to implement as feature classifier technique, the Vector Quantization (VQ) approach [20, 21], have been extensively validated in a variety of applications including speaker recognition and identification process [29, 30]. The VQ technique consists in summarizing the distribution of a given set of an acoustic speech entrance and partitioning it into subspaces, each represented by a reference vector (codevectors or centroids), where each set of codevectors is called the quantization dictionary (codebook) [31]. In fact, the popularity of VQ comes from the fact that is a probabilistic problem of finding the centroids, which approximate with the smallest possible distortion the input data [20]. Figure-4 shows the process of VQ feature classification technique using HFCC feature extraction method.

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

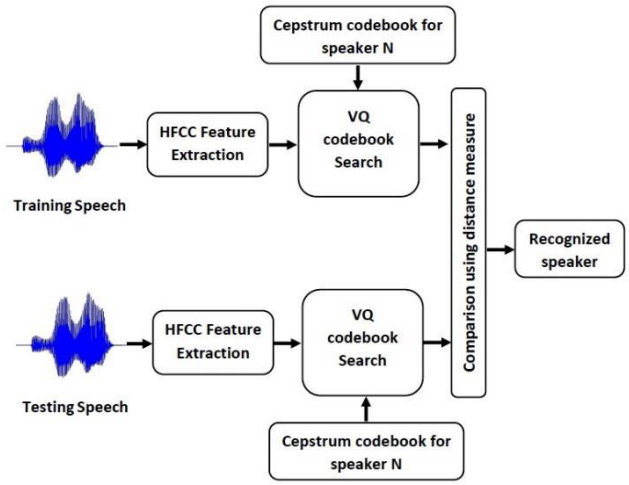

Figure-4. Block diagram of Vector Quantization (VQ) feature classification process using Human Factor Cepstral Coefficients (HFCC) feature extraction technique.

The codebook synthesis is the most important task that decides the performance of VQ process. Assuming that a training set of $\mathbb{T}$-dimensional acoustic input feature vectors represented as $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \dots, \boldsymbol{x}_N)^t$, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \dots, \dots, x_{i\mathbb{T}})^t \in \Re^{\mathbb{T}}$ is mapped to another $\mathbb{T}$-dimensional codevectors $\hat{G}$ where $\hat{G}_j = (\hat{G}_{j1}, \hat{G}_{j2}, \dots, \dots, \hat{G}_{j\mathbb{T}})^t \in \Re^{\mathbb{T}}$. The reconstructed set of vectors (called codevectors $\hat{G}$) belongs to a finite set of representatives called codebook, which is denoted here as $Y = (\hat{G}_1, \dots, \hat{G}_K)$, contains K codevectors where K is called the codebook size $K \ll \mathbb{T}$. It is then estimated a matching score, which is the distance between the input acoustic feature vectors and their reproduction, vectors (codevectors). The minimal distance of the closest codevector from the codebook is calculated by the Euclidean distance shown in Eq. (9)

$$z = \sum_{i=1}^{\mathbb{T}} \min_{1 \le k \le K} (d(\boldsymbol{x}_i, \hat{G}_k)). \quad (9)$$

3.2 Linde-Buzo-Gray (LBG) codebook design algorithm

One of the best-known adopted learning algorithms that generate a local optimal codebook is the Linde-Buzo-Gray (LBG) approach introduced by Linde *et al.* [32]. Generally, the LBG algorithm implements in a first step the growing evolution of the codebook not only with respect to the partitioning of the codevectors but also with respect of the number of codevectors in the codebook, which begins to reach the desired value K using a splitting strategy [32]. In a second step, the LBG algorithm executes according to whether the distortion in step (i-1) and the distortion in step (i) are more or less different. When this difference falls below a threshold that we have set, the algorithm ends and we get the optimal dictionary. Figure-5 shows the flowchart of the LBG algorithm.

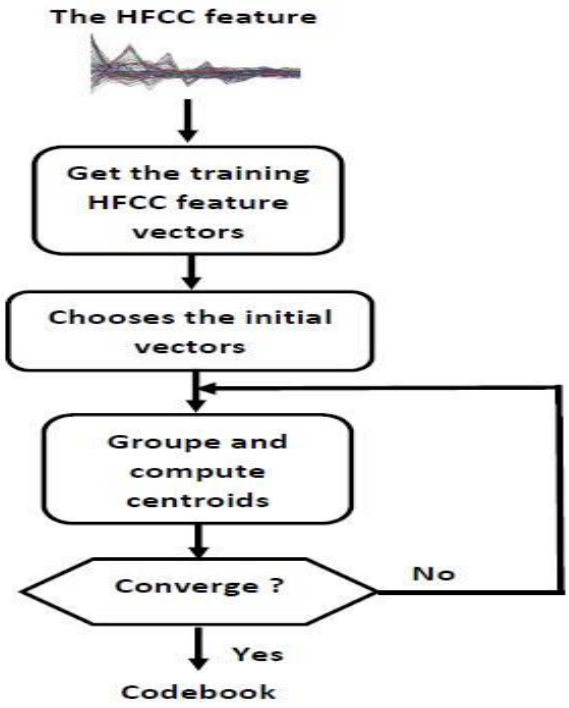

Figure-5. Flowchart of the Linde-Buzo-Gray (LBG) clustering algorithm.

4. RESULTS AND EVALUATION

The evaluation of the proposed system has been carried out on a database of 10 speakers used in training and testing stage. All speakers (five Male and five Female) were invited to pronounce the same utterance "zero" at a comfortable level. All algorithms were executed on a desktop computer using Intel ® Core ™ i3 CPU, a processing speed of 2.27 GHz and Windows 10 operating system. In the feature extraction stage, a frame length of 20 ms with a 50% overlap for extracting short-term features, Hamming window and a pre-emphasis filter $H_{pre}(z) = 1 - 0.97z^{-1}$ were used. The human-frequency cepstral coefficients (HFCCs) dimension were fixed to 12 and the magnitude spectrum is filtered with a bank of 10, 20 then 40 triangular filters respectively. To evaluate the performance of the presented HFCC feature using 10, 20 and 40 HFCC filterbank, the vector quantization (VQ) classification technique based LBG clustering algorithm is used. We use seven codebook of size ($\mathcal{COD}_{size}$) 1, 2, 4, 8, 16, 32 and 64. As a measure of performance, the identification rate is computed for the correctly classified speakers out of the total speakers used for testing. In our work, the performance of the identification system is given in terms of identification rate I_R by the following mathematical expression in Eq. (10).

$$\%\ \ I_R = \frac{N_c}{N_t} * 100 \quad (10)$$

Where N_c denotes the number of correctly identified speakers and N_t represents the total number of speakers.

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

Figures 6 and 7 shows the plot of a 12 HFCC coefficients set extracted from the first speaker for (a) 10HFCC filterbank, (b) 20 HFCC filterbank and (c) 40 HFCC filterbank before and after applying the VQ classifier approach and using a codebook of size 8. We can observe from Figure-7 compared to figure-6 that after applying the VQ technique the number of extracted features vectors are reduced to the eight most significant frames.

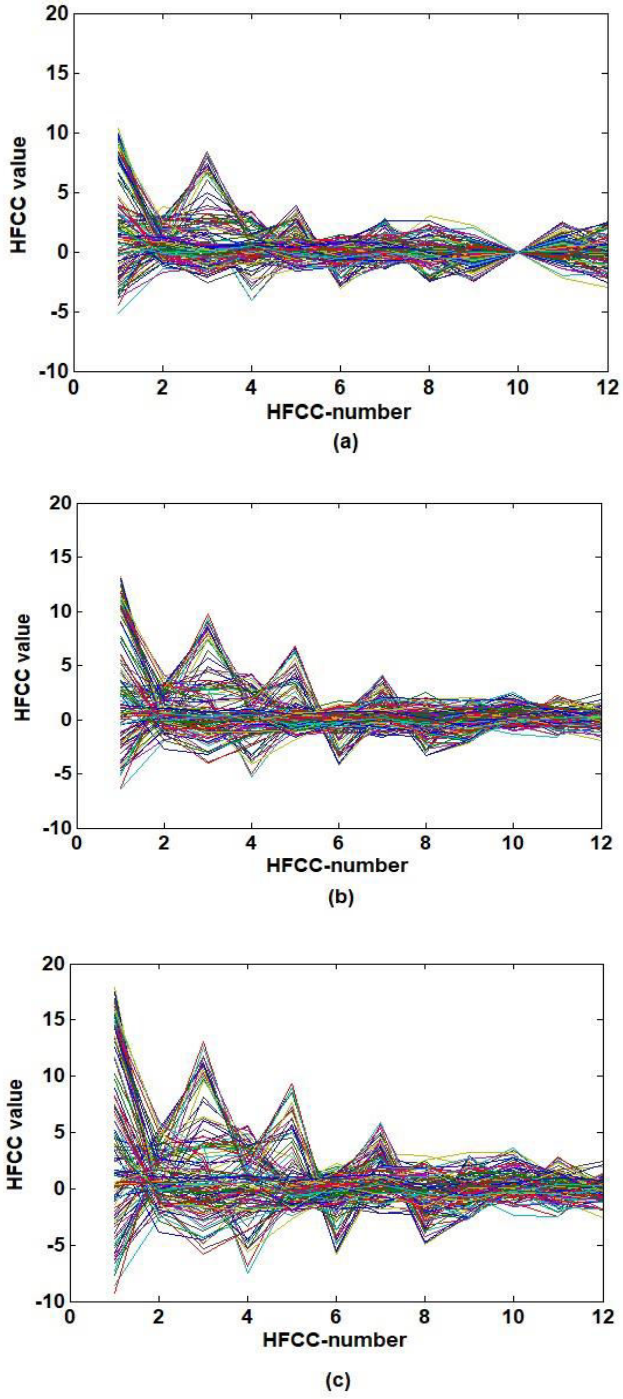

Figure-6. Extracted features vectors plots for speaker1 before applying VQ technique for (a) 10- HFCC filterbank, (b) 20- HFCC filterbank, (c) 40- HFCC filterbank.

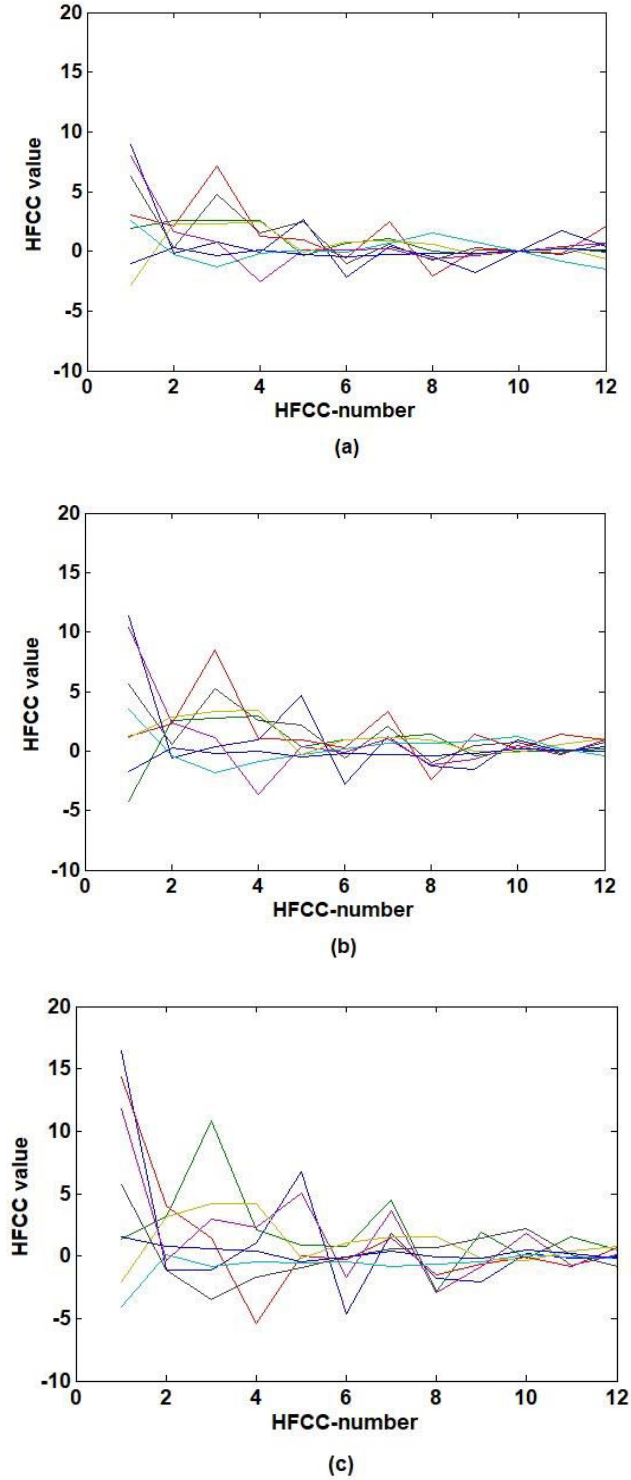

Figure-7. Extracted features vectors plots for speaker1 after applying VQ technique for (a) 10- HFCC filterbank, (b) 20- HFCC filterbank, (c) 40- HFCC filterbank.

Figures 8, 9 and 10 present the two-dimensional partition plot of the acoustic feature vectors of the first and second speaker using HFCC algorithm. The VQ algorithm generates a separate codebook for both speaker1 and speaker2 using the codebook of size 4, 16 and 64respectively for 10, 20 and 40 HFCC filterbank. We

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

notice that with the application of VQ algorithm the speaker reference model is built and then clustered into a number of classes, here the Red Cross signs refer to the acoustic vectors from speaker 1 and the blue one is referred to speaker 2. The red and blue triangle refers to the centroid of speaker 1 and 2 respectively. Using the LBG-clustering algorithm the VQ codebook is generated for both speakers by clustering his/her acoustic vectors. We can observe from Figures 8, 7, and 9 with the increasing number of codebook-size (we use here just 4, 16 and 64 codebook size plot) the vectors are well clustered and the codeword are therefore capable of modeling a particular speaker accurately.

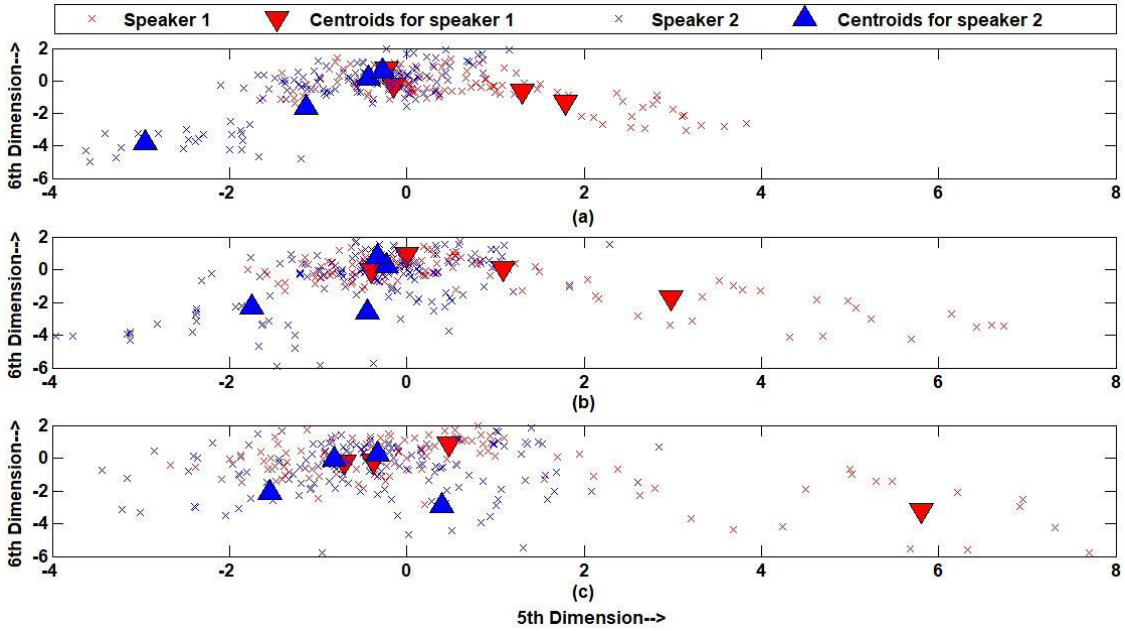

Figure-8. 2D acoustic space plot of codebook-size 4 for both speaker 1 and 2 for (a) 10 HFCC filterbank, (b) 20 HFCC filterbank and (c) 40 HFCC filterbank.

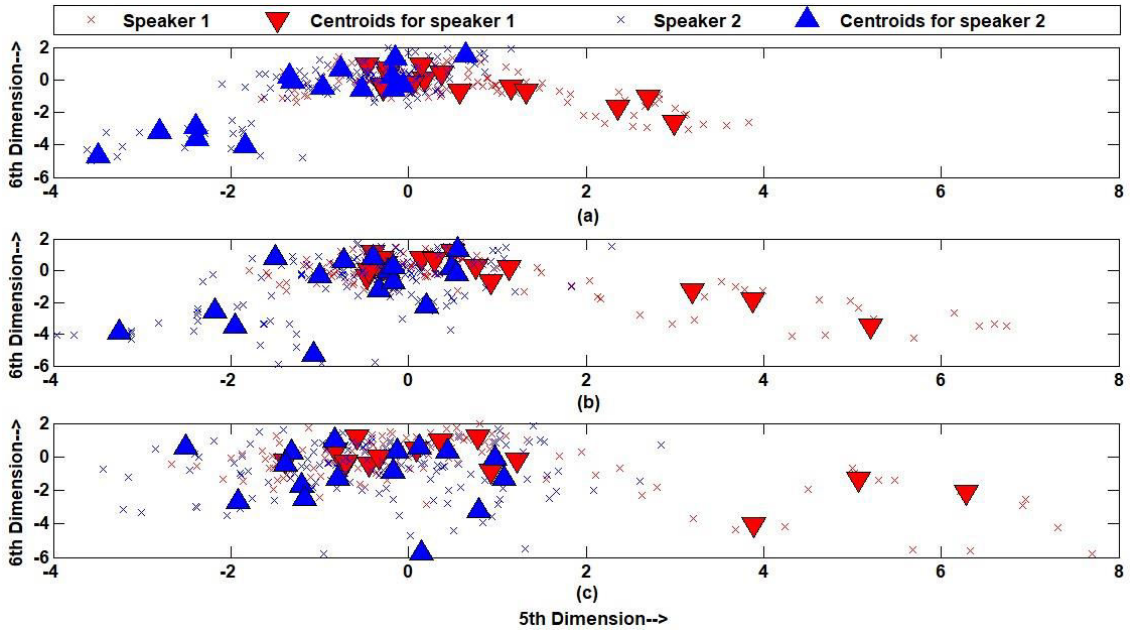

Figure-9. 2D acoustic space plot of codebook-size 16 for both speaker 1 and 2 for (a) 10 HFCC filterbank, (b) 20 HFCC filterbank and (c) 40 HFCC filterbank.

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

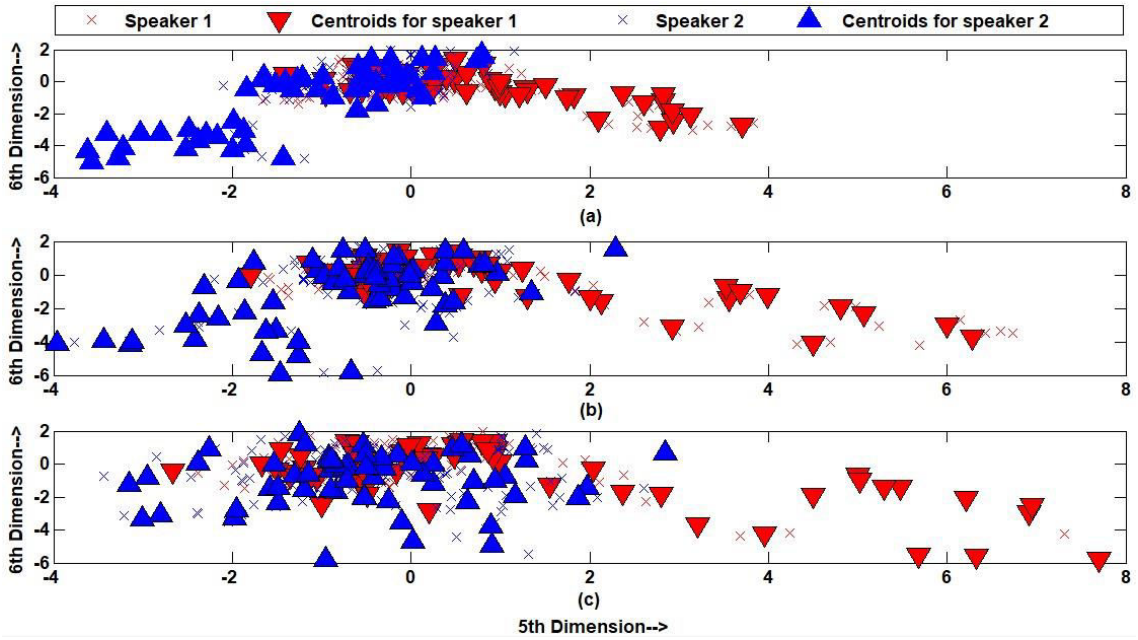

Figure-10. 2D acoustic space plot of codebook-size 64 for both speaker 1 and 2 for (a) 10 HFCC filterbank, (b) 20 HFCC filterbank and (c) 40 HFCC filterbank.

The number of filterbanks determines the frequency resolution of the HFCC analysis. As can be seen from the Figure-11 (a) red dots represent the center frequency for 10 filterbanks, in (b) the blue plus represents the center frequency for 20 filterbanks and finally in (c) green cross represents the center frequency for 40 filterbank. Thus, it can be noticed as the number of filterbank increases the central frequencies get closer to each other.

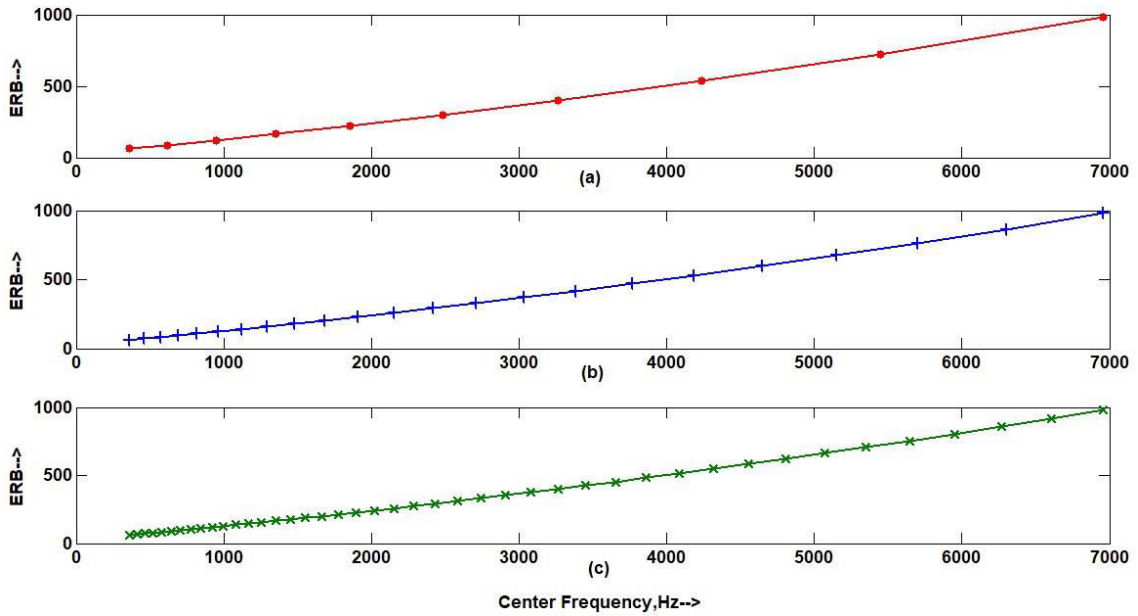

Figure-11. The plot of center frequency vs. the ERB scale for (a) 10 filterbank, (b) 20 filterbank and (c) 40 filterbank.

Up to this point, we have shown the plots of the HFCC feature extraction technique and the feature-clustering step. Next, we analyze the results obtained when varying the number of HFCC filterbank against different codebook size. The Table-1 show the results of the speaker identification task using different $\mathcal{COD}_{size}$ vs. HFCC filterbank.

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

Table-1. Speaker identification results using different codebook-size vs. 10, 20 and 40 HFCC filterbank.

Codebook size	10 filter-bank	20 filter-bank	40 filter-bank
1	15%	25%	50%
2	25%	50%	75%
4	75%	87.5%	87,5%
8	75%	87.5%	100%
16	87.5%	100%	100%
32	87.5%	100%	100%
64	100%	100%	100%

From Table-1, it is clearly shown that either $\mathcal{COD}_{size}$ or number of HFCC filterbank can be both considered as a factor in the accuracy of the system. Thus, increasing the number of $\mathcal{COD}_{size}$and the HFCC filterbank leads to a noticeable increase in the identification rate.

The results show an identification accuracy that range from 15 % to 100% and peaks at 100% by using 64 $\mathcal{COD}_{size}$, for "10 HFCC filterbank", whereas for the "20 HFCC filterbank" the findings show that the accuracy of the identification rate increases up from 25% and stays stagnant at 100% for 16, 32 and 64 $\mathcal{COD}_{size}$. Finally, we can observe for "40 HFCC filterbank" that the identification rate peaks at 100% using 8, 16, 32 and 64$\mathcal{COD}_{size}$. We noticed that, when using large codebook-size ($\mathcal{COD}_{size} = 64$), the identification rate for the three HFCC filterbank model are very close to each other.

The overall identification results are display in Figure-12. From this comparison, we see that the highest identification accuracy of 100% is achieved using 40 HFCC filterbanks with less codebook size 8-$\mathcal{COD}_{size}$followed by 20 HFCC filterbank with 16-$\mathcal{COD}_{size}$ and finally 10 HFCC filterbank with 64-$\mathcal{COD}_{size}$. However, for 40 HFCC filterbank, we have to take into consideration that increasing the $\mathcal{COD}_{size}$ beyond ($\mathcal{COD}_{size}$> 8) will not have any effect for the identification rate of the system; we will only get more computational cost.

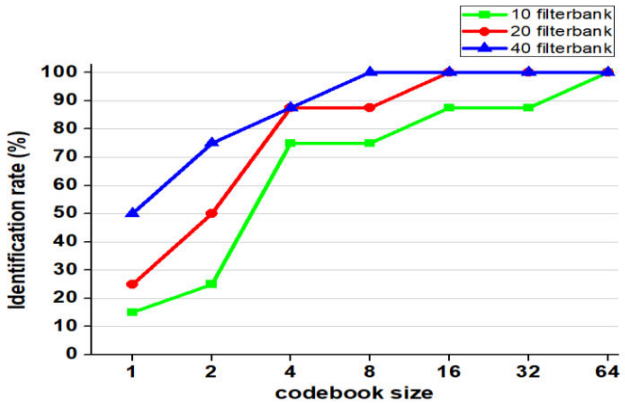

Figure-12. Comparison of the identification accuracy for 10, 20 and 40 HFCC filterbanks using different codebook size.

5. CONCLUSIONS AND FUTURE WORK

Human Factor Cepstral Coefficients (HFCCs) is considerate as the new feature extraction technique that attempts to approach the behavior of the human auditory system in order to achieve the most reliable speaker identification results. In this study, the authors propose to investigate the Human Factor Cepstral Coefficients (HFCC) filterbank influence on a text-dependent speaker identification process. Thus, for this purpose, we attempt to vary the number of HFCC filterbank using: N_{fb}=10, 20 and 40 filters. Moreover, the performance of our proposed process is evaluated using the Vector Quantization (VQ) classifier based LBG algorithm. We vary the codebook-size ($\mathcal{COD}_{size}$) despite the number of HFCC filterbank using 1, 2,4,8,16,32 and 64 $\mathcal{COD}_{size}$.

We found that there is an interesting tradeoff between the number of filterbanks and the$\mathcal{COD}_{size}$, where the choice of filterbanks number plays a part in the final identification results. In fact, the HFCC filterbank is directly related to the system performance since they determine the frequency resolution of the HFCC analysis. We can conclude that more the number of filterbank increase more the center frequency get closer, that means taking more data from the input speech and that influence the system for a better identification rate.

As a continuation of this work, we plan to use other feature extraction algorithms and classification techniques in a noisy acoustic environment.

REFERENCES

[1] Rajesh R., Ganesh K., Koh S. C. L., Singh N., Khan R. A. & Shree R. 2012. International conference on modelling optimization and computing applications of speaker recognition. Procedia Engineering. 38: 3122-3126.

[2] M. He, S.J. Horng, P. Fan, R.S. Run, R.J. Chen, J.L. Lai, M.K. Khan, K.O. Sentosa. 2010. Performance Evaluation of Score Level Fusion in Multimodal Biometric Systems Pattern Recognition. 43(5): 1789-1800.

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

[3] S.J. Horng, D. Mulyono. 2008. A study of finger vein biometric for personal identification. Proceedings of the IEEE International Symposium on Biometrics and Security Technologies, Islamabad. 22-23.

[4] Reynolds, D. 2002. An overview of automatic speaker recognition technology. In: Proc. of IEEE international conference on acoustics, speech and signal processing. ICASSP'02. (4): 4072-4075.

[5] N. Singh, R.A. Khan, R. Shree. 2012. Applications of Speaker Recognition. Procedia Eng. 38: 3122-3126.

[6] Campbell, J. P., Jr. 1997. Speaker Recognition: A Tutorial. Proceedings of the IEEE. 85(9): 1437-1462.

[7] Reynolds D. A. 2002. An Overview of Automatic Speaker Recognition Technology. Proceedings of IEEE International Conference on (ICASSP '02).

[8] Hébert M. 2008. Text-Dependent Speaker Recognition. In: Benesty, J., Sondhi, M., Huang, Y. (Eds.), Springer Handbook of Speech Processing. Springer-Verlag, Heidelberg. 743-762.

[9] L. Rabiner, B.-H. Juang. 1993. Fundamentals of Speech Recognition, Prentice Hall PTR.

[10] B.S. Atal, S.L. Hanauer. 1971. Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. In journal of the acoustical society of America. 50(2): 637- 655.

[11] D.A. Reynolds. 1994. Experimental Evaluation of Features for Robust Speaker Identification. IEEE Trans. Speech Audio Process. 2(4): 639-643.

[12] H. Hermansky. 1990. Perceptual Linear Predictive (PLP) Analysis of Speech. J. Acoust. Soc. Am. 87(4): 1738-1752.

[13] S. B. Davis and P. Mermelstein. 1980. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE Trans. Acoust., Speech, Signal Processing. 28(4): 357-366.

[14] B. Sudhakar, R. Bensraj 2015. An Expressive HMM-Based Text-To-Speech Synthesis System Utilizing Glottal Inverse Filtering For Tamil Language. ARPN Journal of Engineering and Applied Sciences. 10(6).

[15] A. Biswas, P.K. Sahu, A. Bhowmick, M. Chandra. 2014. Hindi vowel classification using GFCC and formant analysis in sensor mismatch condition, WSEAS Trans. Syst. 13.

[16] F. Francis, V. Rajan. 2015. A Novel Noise Robust Speaker Identification System. ARPN Journal of Engineering and Applied Sciences. 10(17): 7641-7646.

[17] W.M. Campbell, J.P. Campbell, D.A. Reynolds, E. Singer and P.A. Torres- Carrasquillo. 2006. Support Vector Machines for Speaker and Language Recognition. Computer Speech and Language. 20: 210-229.

[18] C. Wang, D. Xu, C.P. Jose. 1997. Speaker Verification And Identification Using Gamma Neural Networks. In international conference on neural networks.

[19] D.A. Reynolds. 1995. Speaker Identification And Verification Using Gaussian Mixture Speaker Models. Speech Communication. 17: 91-108.

[20] Yuk, C.C.Q.L.D.-S. 1996. An HMM Approach to Text Independent Speaker Verification. In IEEE International conference on Acoustics, Speech and Signal Processing.

[21] R. Gray. 1984. Vector Quantization. IEEE Magazine on Acoustics Speech and Signal Processing. 1: 4-29.

[22] A. Gersho. 1982. On The Structure Of Vector Quantizers. IEEE Transactions on Information Theory. 28(2): 157-166.

[23] Skowronski Mark D, Harris John G. 2004. Exploiting independent filter band-width of human factor cepstral coefficients in automatic speech recogni-tion. J Acoust. Soc Am. 116(3): 1774-80.

[24] Wielgat Robert, *et al.* 2007. HFCC based recognition of bird species. In: Signal processing algorithms, architectures, arrangements and applications. IEEE.

[25] A. Jose Albin and N. M. Nandhitha. 2015. Text Independent Human Voice Ranking System for Audio Search Engines Using Wavelet Features. ARPN Journal of Engineering and Applied Sciences. 10(4).

[26] H.P. Combrinck, E.C. Botha. 1996. On The Mel-Scaled Cepstrum. Department of Electrical and Electronic Engineering, University of Pretoria.

©2006-2019 Asian Research Publishing Network (ARPN). All rights reserved.

[27] L. Rabiner, B.H. Juang, Fundamentals of Speech Recognition, Prentice Hall, Englewood Cliffs, NJ, 1993.

[28] B. C. J. Moore and B. R. Glasberg. 1983. Suggested Formula For Calculating Auditory-Filter Bandwidth And Excitation Patterns. J. Acoust. Soc. Am. (74): 750-753.

[29] Burton D. 1987. Text-Dependent Speaker Verification Using Vector Quantization Source Coding. IEEE Trans. Acoust. Speech Signal Process. 35(2): 133-143.

[30] F.K. Soong, A.E. Rosenberg, B.-H. Juang, Rabiner, L.R. 1987. A Vector Quantization Approach to Speaker Recognition. AT & T Tech. J. 66: 14-26.

[31] Soong F., Rosenberg A., Rabiner L. & Juang B. 1985. A Vector Quantization Approach to Speaker Recognition. IEEE International Conference on ICASSP '85.

[32] Y. Linde, A. Buzo, R.M. Gray. 1980. An algorithm for vector quantizer design. IEEE Transactions on Communications. 28(1): 702-710.