Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

Survey of Features Extraction and Classification Techniques for Speaker Identification

Sahar Adil Kadhum[1] Ahmed Badri Muslim[2] Ali Yakoob Al-Sultan[3]

1. Computer Dept., College of Science for Women , University of Babylon, Iraq . wsci.sahar.adil@uobabylon.edu.iq
2. Computer Dept., College of Science for Women , University of Babylon, Iraq . wsci.ahmed.badri@uobabylon.edu.iq
3. Computer Dept., College of Science for Women , University of Babylon, Iraq. wsci.ali.yakoob@uobabylon.edu.iq

Article Information
Submission date: 1 / 12 /2019
Acceptance date: 31 /12/ 2020
Publication date: 31 / 12/ 2020

Abstract

Speech processing is more common day by day to provide enormous safety. The speech for the purpose of authentication is commonly used. Recognition of the speaker is the method that can check and recognize the speaker. The scheme of speech recognition is distinct from the scheme of speaker recognition. Recognition of speakers is commonly used in sectors, hospitals, laboratories, etc. Its benefits are safer, easier to implement, more user-friendly. Speaker identification method is one of the most commonly used techniques for the region where safety is very crucial. This article presents an overview of various methods that can be used to recognize speakers' systems, the feature extraction techniques such as Linear Predictive Coding (LPC), Linear Predictive Cepstral Coefficients (LPCC), Unique Mapped Real Transform (UMRT), Real Cepstral Coefficients (RCC), "Mel-frequency Cepstrum" (MFCC), in addition to various classification techniques such as "Gaussian mixture model (GMM)", "Dynamic Time Warping (DTW)", Support Vector Machine (SVM), Neural Network (NN), "Vector Quantization" (VQ). The primary purpose of is to explain the common speaker recognition methods. The obtained results are that, MFCC is chosen for high efficiency and low complexity. and GMM is helpful in classifying less memory and less planning and efficient test results.

Keyword: "Speaker Recognition", Vector quantization, Feature Extraction. MFCC, Classifiers

1.Introduction

One of the most important means of communicating is voice. It gives the audience members various levels of data. It passes on the message; moreover, it gives details on the speaker's sexual identity, emotion and personality. There are several situations where the speaker 's proper recognition is needed. There are various ways to identify the person in biometrics, such as finger print, palm, iris, conflict, recognition of speech. The aim of the speaker statement is to view and isolate the data from speech signal. Autonomous content and subordinate content "speaker recognition" are the kinds of identification of speakers. The human voice's behavioral point of view is used to identify verification by moving from simple to computerized over a spoken word and by extracting vocal features (e.g. pitch), tone, recurrence and rhythm to set up a speaker voice test. The voice affirmation contains enrollment and affirmation techniques. Enrollment prepare portrays there [1]

Speech is given through the vocal folds as a consequence of the time-varying vocal tract structure being excited by the time-varying flag. Accepting the vocal tract setup to be stationary amid the brief interims of time utilized for investigation, the comparing speech flag is the convolution of

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

the excitation flag with the vocal tract motivation reaction within the time space. For speaker distinguishing proof, there are time tried strategies to extricate data from the voice source (e.g. direct expectation remaining) and the vocal tract channel (eg. MFCC) [2].

Speaker distinguishing proof points at distinguishing the speaker based on discriminative highlights extricated from discourse signals. The work in programmed speaker recognizable proof (ASI) has started in 1950. An ASI framework comprises include extraction and classification. Highlight extraction is the method of information lessening with good agent highlights. The foremost widely-used approach for speaker distinguishing proof is based on cepstral examination. Cepstral investigation includes the decomposition of discourse flag outlines into coefficients within the log ghastly space through the modeling of the human ear execution with the Mel channel bank [3].

2 Feature Extraction

Extraction of features is the way to distinguish unique components from the flag of data. This extraction function is completed after the pre-processing. There are several methods for features extraction such as MFCC, pitch, energy, formant and so on. [4].

The notion of "feature extraction" is focused on translating the voice signal into some parametric form kinds for further analysis and evaluation. The acquired spectrum features in automatic systems were known to be more efficient. However, due to their inconsistency as well as elevated dimensionality, it is common acknowledged that direct "spectrum-based" characteristics are inconsistent with the applications of recognition.

The primary concept of extraction of features is to transform the signal space with high-dimensional data to the low-dimensional subspace of a feature while preserving the information of the speaker. The significant of speaker identification characteristics are not influenced by the noise surrounding the place, not influenced by the physical speaker fitness (e.g. disease), further than not influenced by the sound differences, and easy measurement and removal. No specific characteristic, however, has all these features. In automatic identification schemes, therefore, the spectrum-based characteristics are the most efficient [4].

Fig (1): Basic speech Recognition system

2.1 Feature Extraction Techniques

Various methods are used for extraction of features such as "Linear Prediction Coding" (LPC), "Linear Predictive Cepstral Coefficients" (LPCC) and "Mel-Frequency Cepstrum Coefficients" (MFCC) [5].

2.1.1 Linear Predictive Coding (LPC)

An operation for extracting a signal data is a signal processing. LPC is a powerful speech analysis tool that enables the extraction of features that have excellent quality and effective computing results. LPC used a voice synthesis in 1978. LPC performing a formant study decided on a signal formant

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

called reverse filtering, the intensity and frequency were estimated, of the residue voice signal. Because the voice signal varies based on the moment, the estimation cut a signal called the frame [6].

2.1.2 Linear Predictive Cepstral Coefficients (LPCC)

LPCC is a common method to obtain the characteristics from the voice signals. The frequency spectrum and energy of sound signal can be described efficiently by LPC parameters. LPCC provides a smoother spectral envelope and a stable representation in comparison with LPC. This method is based on the linear spacing of the frequency band [5].

2.1.3 Unique Mapped Real Transform (UMRT)

Discrete Furrier Transform DFT is commonly used in signal processing. Fast Furrier Transform FFT is a fast algorithm for implementing DFT, taking advantage properties of the symmetry and periodicity of the DFT exponential kernel in the complex domain, transform the real data into complex. The DFT calculation has been adapted in real addition and also complex multiplication giving the property that 2x2 DFT does not require complex multiplication and mapping to the complex domain by multiplying the N/2 twiddle factor [3] for the NRC data. MRT is derived from the Modified DFT by deleting the complex multiplication N/2. In MRT, information grouping and arrangement based on the exponential kernel's respective stage. The total number of MRT coefficients is N range (N/2) (when data sequence of length N) and it also contains redundancies such as full redundancy and derived redundancy. It abolished the full redundancy and produced a more compact variant of MRT called UMRT. The UMRT for a longitudinal N sequence where N is 2 and $x_1, x_2, .. x_n$ is the input sequence [7].

$$Y_0^{(0)} = \sum_{n=0}^{N-1} x_n \quad \text{..... (1)}$$

UMRT rest coefficients computed using the following equation:

$$Y_0^{(0)} = \sum_{n=0}^{N-1} \left(x_{\frac{jN+q}{k}} - x_{\frac{(2j+1)N+2q}{2k}} \right) \quad \text{.....(2)}$$

Where M=N/2, $k=2^t$, $0<=t<=\log_2 M$, $0<=r<=(M/k)-1$, q: index of phase, q=rk, k: index of frequencies.

UMRT comprises of the terms of both stage and frequency. There are 10 frequency terms in a frame size of 512, while the remainder are stage terms. The duration of stage terms varies with the declining energy order of 2. The framing of the voice information is performed with a frame size of 512 after the silent areas are removed. After that, a power of 2 is applied to the "UMRT" for N, then length N "UMRT" coefficients are obtained. In order to be used as a feature extractor, grouping of UMRT coefficients is now performed. The zero coefficient of UMRT is the value of DC. The summation of the next 256 coefficients, containing different phase terms and the same frequency, will be used as the second feature, then the summation of the next 128 coefficients, containing the same frequency term and different phase terms, will be used as the third coefficient, and then the next 64 coefficients will be summed up and used as the fourth coefficient, and so on, until the phase

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

duration. UMRT-based characteristics are acquired in this manner. 10 characteristics representing 10 frequencies are acquired for a size 512 frame [7].

2.1.4 Real Cepstral Coefficients (RCC):

In RCCs, a Fast Fourier Transform (FFT) is applied to each frame to transform the signal from the time domain to the frequency domain. The result logarithm and the reverse Fast Fourier Transform (IFFT) are then introduced to the signal to get the signal's actual Cepstrum. [8].

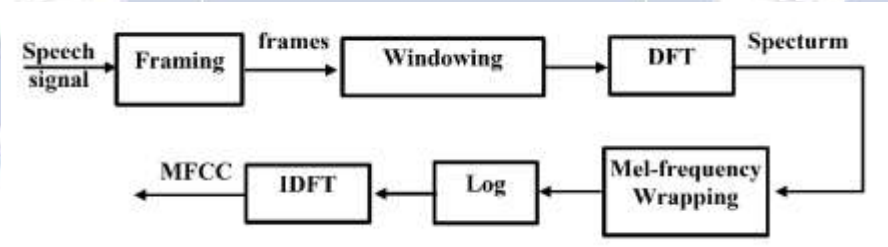
2.1.5 "Mel-frequency Cepstrum" (MFCC)

MFCC are the most common method used for speech recognition. The focus of MFCC calculation is based on the short-term assessment and is equal to the earlier mentioned calculation of "Cepstral Coefficients". The main differences in the using of critical bank filters to realize "Mel-frequency warping". The critical bandwidths of the frequency are based on human ear perception. The block diagram of MFCC seen in fig (2). A Mel is a measuring unit focused on the sound frequency perceived by the human ear and defined according to which, in this manner, 1000 Hz represents 1000 Mels deemed a reference, the physical frequency tailored by hearers until they can hear it, which could be two, ten or one tenth of the reference frequency, and so on.

The measurement of mel is below the first linear frequency spacing of 1000Hz, whereas the logarithmic spacing is above 1000Hz. The frequency-based calculation of Mel could be displayed as the following equation [9]:

$$mel(f) = 2595.\ \mathrm{Log}_{10}(1 + f / 700) \quad \text{..... (3)}$$

T d by *f*, whereas the frequency perceived is referred by Mel(*f*).
tion is comparable to Cepstral Coefficients calculation. The distinction lies in Mel-frequency warping before doing logarithm and reversing DFT. The actual frequency scale that the warping transfers to the scale of perceived human frequency called the Mel-frequency scale. The fresh scale is more than 1kHz logarithmically, with linear spaces less than 1 kHz.



Mel-frequency cepstrum calculation is comparable to Cepstral Coefficients calculation. The distinction lies in mel-frequency warping before doing logarithm and reversing DFT. The actual frequency scale that the warping transfers to the scale of perceived human frequency called the Mel-frequency scale. The current scale is more than 1kHz logarithmically, and linear spaces less than 1 kHz [4][10].

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

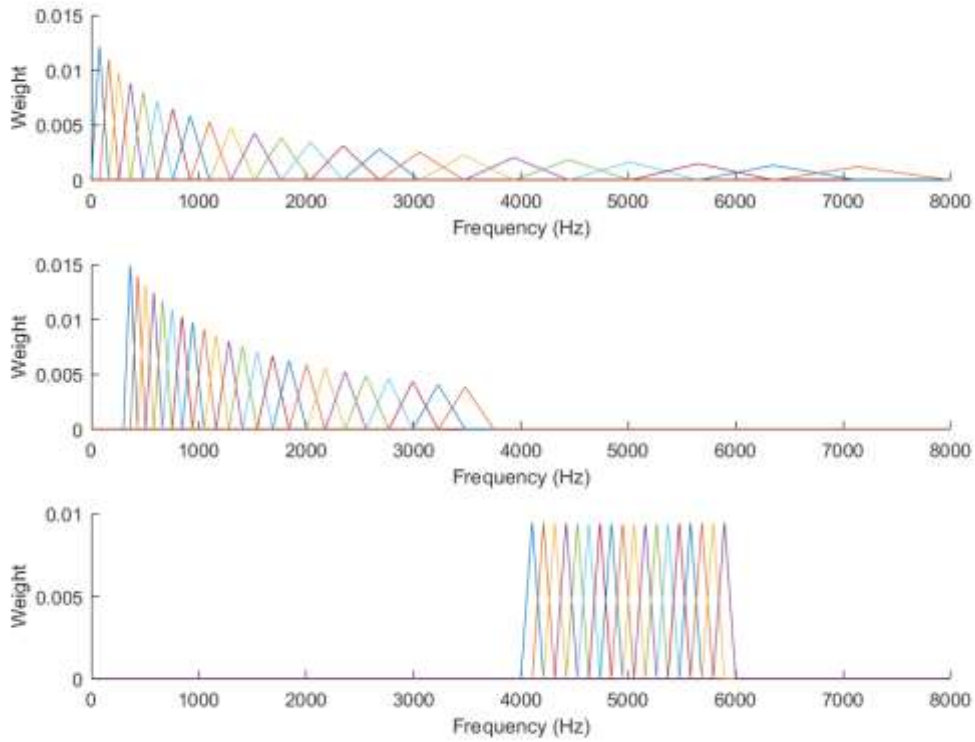


Figure (3) The triangular scaled filter banks of Mel-frequency

MFCC is a common method and is widely used in most voice signal extraction applications [2]. The "MFCC" comprises of five main steps: pre-processing, windowing, FFT, wrapping of "Mel-frequency and cepstrum". The MFCC receives the input signal and we get the required coefficient called MFCC [11] .

The steps of MFCC are given below

- **Pre-processing**: This step involves filtering; filtering converts the specified voice signal to a computer-friendly form. Preprocessing separates the speech component from the unvoiced component.
- **Windowing:** involved to minimize the distortion of the spectrum. To accomplish a stationary conduct, hamming window which used to create the blocking frame at 20-25 ms. Hamming window support the continuity at the start and end of all frames. It is also supporting a best frequency resolution. The windowing results is as the following eq. [11]:

$$z(n) = x(n) \times h(n) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Where, h(n): hamming window, z(n): output signal, x(n): input signal.

- **"Fast Fourier Transform FFT":** It is considered as the most important step in building the fast Fourier Transformation of each frame extracting components from the signals at a rate of 10 ms. Fast Fourier transform converts each N number of samples from time domain to frequency domain. The FFT size is 512, 1024, 2048. It is used to acquire a frequency reaction of magnitude.

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

- **"Mel-Frequency Wrapping"**: Speaker presentation of tone frequency content or expression isn't following a linear scale or not proportional, according to a psychological study. Mel scale is used for measuring distinct pitch. "One Mel is described as 1000 of a 1kHz tone pitch." Mel scale frequency can be approximated by equation [11]:

$$M(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \dots\dots\dots\dots (5)$$

- **Spectrum simulation is performed using filter bank**: The frequency reaction of the triangular band passes is used as a filter bank. Using Mel-scale, the filter bank location is equally spaced.
- **Cepstrum:** In this step, the final stage of MFCC is cepstrum, with the use of DCT (Discrete Cosine Transform) converting Mel spectrum coefficients into time domain. The outcome is achieved as MFCC [11].

3 Speaker Classification Techniques

Classifications are another significant component of the speaker identification scheme. The patterns are categorized into distinct classes at the classification level. Many classifiers like DWT, GMM, SVM, VQ, etc. are used. It is an significant job from this classifier choice. But there are no fixed classifier selection criteria. Many pattern classifiers are studied for the development of language mechanisms such as classification of emotions, speech recognition, speech verification, speech recognition [5].

3.1 "Dynamic Time Warping (DTW)"

Any two-time series that is called wrapping points can vary in time and speed. Thus, the method of information time wrapping is considered as the most commonly used methods for comparing the features (classification). This method is therefore involved to discover the optimum alignment between two time series and to measure the resemblance between those time series. By comparing two time series signals based on the two temporal dimensions linear mapping, the DTW uses linear time wrapping. Thus, by reduce the distance between the two signals, the non-linear alignment of one signal to another is allowed in DTW.

This wrapping can therefore be involved to obtain identification of figures based on resemblance and dissimilarity between these signals. From the point of view of speech signals, the duration of each spoken word or digit may vary, but for the same word or digit, the overall speech waveform is similar. Thus, the respective areas between the two-time series can be readily obtained by implementing the DTW method to be used in matching procedures [12].

3.2 "Gaussian mixture model (GMM)"

GMM is considered as a stochastic model and it is popular method in speaker recognition. GMM is successfully applied in density estimators for "speaker recognition". The GMM may be used as an extension of the VQ model, in which the clusters are overlapping. In GMM each speaker has the independent GMM model. Text independent (TI) recognition can be done using GMM. Therefore, the it crude in which the gross characteristics of the speaker's distribution is modeled [13].

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

3.3 Support Vector Machine (SVM)

For classification, SVM is used. It is possible to classify SVM as binary SVM and multi SVM. We can determine in binary SVM whether or not the individual is acknowledged. Binary SVM compares two speakers' characteristics. But the characteristics of more than two speakers are compared by multi SVM. It comes under supervised classifier. Basic of SVM is to create a hyper plane. This hyper plane differentiates the features. In binary SVM features are classified into two classes, each class for recognized and non-recognized speaker [14].

3.4 Neural Network (NN)

The use of neural networks is another strategy in acoustic modeling. An artificial neural network (ANN) is a mathematical tensile structure capable of defining complicated nonlinear interactions between information sets input and output. ANN models were discovered to be useful and well-organized, primarily in disorders where the process features are difficult to define using physical equations. Information that is more accurate than HMM-based systems through the network, as long as the training data and vocabulary are finite.

Phoneme recognition is a more prevalent method using neural networks. This is an active study area, but the findings are usually better than HMMs [15].

3.5 "Vector Quantization" (VQ)

VQ technique used to generate a small set of feature vectors meaning the distribution centroids and receiving an abundant set of feature vectors of a speaker, to minimize the distance to any points. The justification for using VQ may be due to the impracticability of representing each separate feature vector in the feature space produced from the corresponding speaker's utterance practice. While the VQ requires some time to produce the centroids, it saves a lot of time throughout the test phase as compared to a certain user's overloaded feature space, considering the limited number of feature vectors. As a result, a vector quantizer in the vector space Rk can be used to map k-dimensional vectors into a finite set of vectors Y={yi: i=1,..., N}. The amount of feature coefficients can be referred to as K-dimension in each feature vector. Each vector yi is called a code-word or code vector, while all codewords are set by the codebook. Therefore, code books are developed by using the Vector Quantization technique for each presenter providing a certain number of customers throughout the teaching stage. There is an adjoining neighboring region, known as Voronoi region, for any codeword yi, and is distinguished by [10]:

$$V_i=\{ x \in R^k : \|x-y_i\| \leq \|x-y_j\|, \text{ for all } j \neq i \quad \text{.....(5)}$$

A set of Voronoi regions separate all features space of a given speaker as follows:

$$\bigcup_{i=1}^{N} V_i = R^k \quad \text{..... (6)}$$

$$\bigcap_{i=1}^{N} V_i = \phi \quad \text{for all } i \neq j \quad \text{..... (7)}$$

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

To show what a Voronoi region actually is, considering vectors with only two coefficients of features. Therefore, the two-dimensional (2D) feature space would represent them. Figure (4) shows all the function vectors that a particular user has obtained in a 2D vector space. As highlighted in the figure, by implementing vector quantization method, all feature vectors were correlated with their adjacent neighbor and the corresponding centroids are generated. The centroids are shown by red dot while the various green dots indicate the vectors of the function [10].

Every codeword (centroid) in its own region of Voronoi. These areas are divided by imaginary lines as described in the visualization figure. Euclidian distance is assessed from each codeword for a known input vector, and the one with the least Euclidian distance for that given vector is the right codeword. The cluster region is depicted by the Voronoi region which is linked to a specified centroid for a specific vector. The distance from Euclidean is recognized by [10]: -

$$d(x, y_i) = \sqrt{\sum_{j=1}^{k} (x_j - y_{ij})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

Where x_j: input vector, y_{ij}: the codeword .

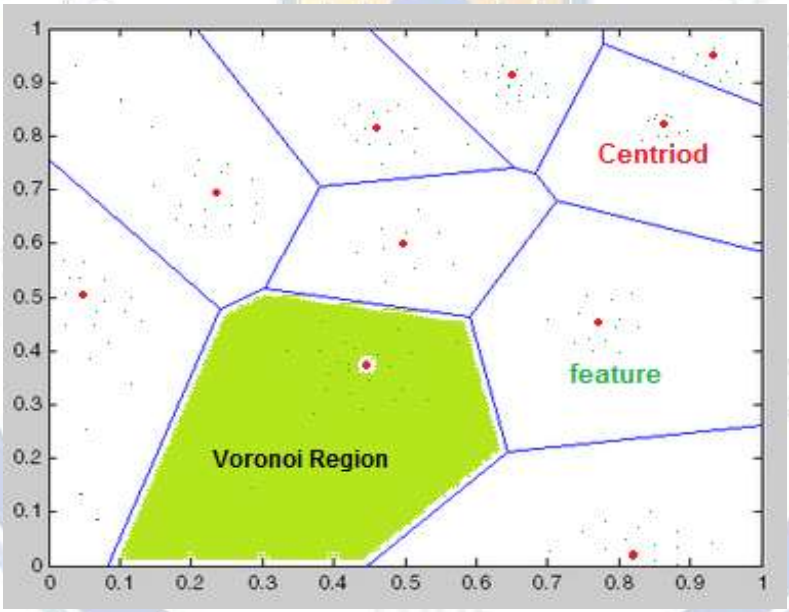


Figure (4) Codewords in 2-dimensional space. x: Input vectors, red circles represent the codewords [10]

3.5.1 VQ properties [10]:

• Lower data storage for spectral assessment. The component of "speech recognition" computation is the identification of spectral resemblance between a couple of vectors. This is usually reduced to a table-based to find the similarities between pairs of codebook vectors based on the depiction of VQ. Also, Reduced calculation to define comparable vectors for spectral analysis.

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

The figure (5) below demonstrates a speaker recognition model using vector quantization:

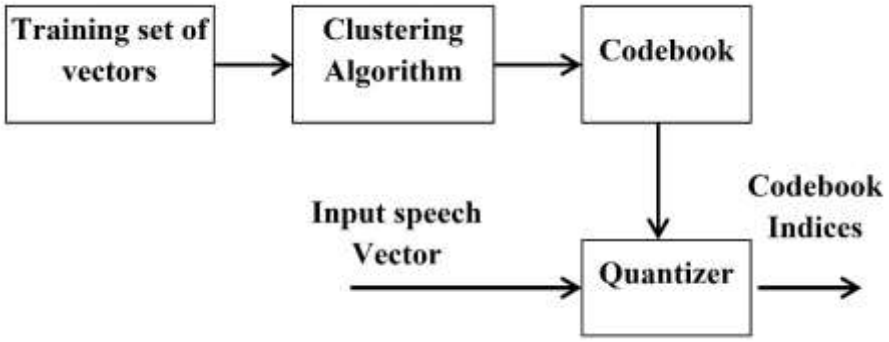


Figure (5) Training and classification structure of VQ Block Diagram

Speaker identification can be accomplished by using the code book already produced by VQ method for each registered user. Figure (5) best describes this process.

There are merely two registered users as shown in figure (5), for whom vectors of features were acquired and quantization of vectors was applied. Only four codewords are developed for any specific user. These codewords commonly indicate each user's codebook. Therefore, a corresponding code book is generated for both users. Mapping unknown user feature vectors into feature space with the presumption that the unknown speaker is one of the qualified speakers throughout the test stage. Thus, by applying Euclidian distance of a given code book for each feature vector, the proximate code word is acquired. The minimum distance observed is marked as distortion of VQ.

VQ distortions are calculated and summarized in a similar way for the remaining feature vectors. For the following speaker, this process is repeated. The wish will be offered by the minimum summation of VQ distortions for each user.

For each voice frame of about 30 ms with overlap, a set of Mel-frequency cepstrum coefficients is calculated using the method shown above. These results from a cosine transformation of the short-term power spectrum logarithm transmitted on a mel-frequency scale. This set of coefficients is represented by an acoustic vector. Consequently, every input utterance is transformed into an acoustic vector order.

3.5.2 Features Matching

The difficulty of recognition of speaker goes into a wider topic in engineering and science, namely the pattern recognition. The objective of pattern recognition is to categorize objects of interest into a number of classes or categories. The speakers are represented as a class. As the categorization system in this case is functional on extracted features, it may also be indicated as feature matching [10].

Vector quantization is a process of vectors mapping from a large vector space to a finite quantity of regions in that space. Each region is termed as a cluster and can be characterized by its center called a codeword. The assembly of all codewords is called a codebook. Figure (6) illustrates a hypothetical diagram to demonstrate this recognition method. Only two dimensions and two speakers of the acoustic space are presented in the figure. The triangles are from the speaker2 whereas the

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

circles denote the acoustic vectors from the speaker1. With the use of the clustering algorithm, a *speaker-specific* VQ codebook is created during the training phase for each acknowledged speaker by clustering their training acoustic vectors. The resulting codewords (centroids) are illustrated in figure (6) by black circles for speaker 1 and black triangles for speaker 2. Then the resultant VQ-distortion represents the distance from a vector to the neighboring codeword of a codebook. The speaker who match the VQ codebook with least total distortion is recognized as the speaker of the input utterance.

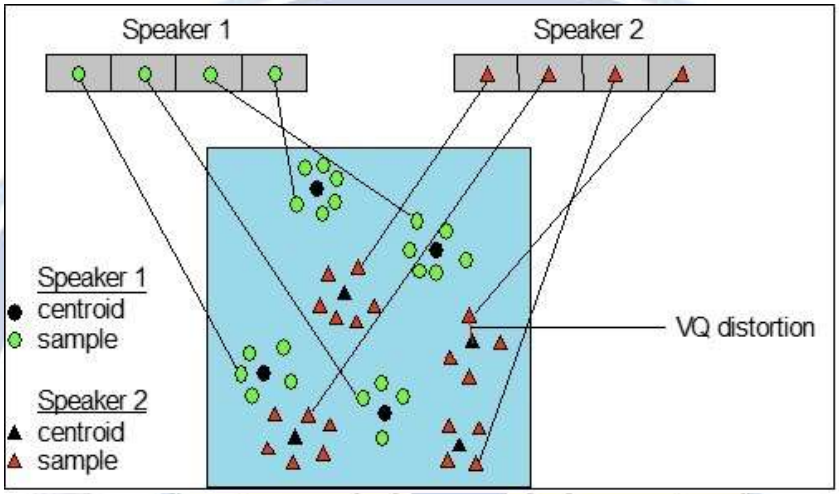


Figure (6) Conceptual diagram illustrating vector quantization codebook formation

4. Factor Affecting Speaker Recognition:

The speaker recognition efficiency is simultaneously influenced by many variables. The interpretation is in details that drive the speaker recognition scheme and influence its performance outlined as follows [10]:

- **Voice Sample Length-Longer samples:** were generally thought to enhance speaker recognition effectiveness but were not verifiable. Cleaning samples does not make any important difference. The brief samples provide better outcomes in background noise, but the distinction lies within the point of confidence.
- **Voice Sample Quality:** The greater quality of microphones leads to better outcomes for clean voice samples. However, normal microphone quality provides nearly the same outcomes as microphones of excellent quality.
- **Modality of Speech**: whether the scheme is text-dependent or text-independent. The function of text is trivial, but it is more complicated to design text-independent systems.
- **Noise**: The background noise is the most important factor for the precision of speaker recognition, which is high for smooth samples but is rapidly deteriorating for noisy samples. There is no important impact on babble noise alone.
- **Microphones**: The best results without mismatch are that the quality of the microphone itself is insignificant and it does not have much more impact on the efficiency of speaker recognition

Journal of University of Babylon for Pure and Applied Sciences, Vol. (28), No. (3): 2020
Online ISSN: 2312-8135, Print ISSN: 1992-0652

systems Disguise-Deliberate cheating is possible, recognition fails in most cases but it does not have much more impact on speaker recognition systems.

- **Language used in practice and testing**: There is no greater effect in native language expression on the effectiveness of speaker recognition. The English linguistic samples provide better outcomes for the noisy samples.
- **Speaker Population:** This has a major effect on the system's effectiveness. As the number of speakers in our database rises, we must compromise with the speaker's right acceptance [16].

Table (1) Summary of advantages and disadvantages of Techniques used by Speaker Identification System (SIS)

Technique	Characteristics	Advantages	Disadvantages
LPC	Formant estimation technique, it is modeled by all pole model	Accurate, Fast, Robust	Cannot detect similar vowel
LPCC	modeled by all pole model	Smoother and stable representation	Compute details for all frequencies
RCC	Approximate the signal through the decompose to sub-bands	Time Frequency Localization	Denoising, Computationally fast
UMRT	it is modeled by all pole model	Fast	Low Accuracy
MFCC	Simulate the human auditory system, Filter bank coefficients	High Accuracy Low Complexity	Affected by the background noise
DTW	Unsupervised	Low storage space	Cross-channel issue
HMM	Unsupervised	Efficient performance	High storage space, High complexity
GMM	Unsupervised	Without training and test data	High complexity
VQ	Unsupervised	Low complexity, easy implementation,	Real time encoding complex
SVM	Supervised	Simple operation	Binary SVM Limited to speaker recognition,

4. Conclusion

In this paper we explore feature extraction techniques and classification techniques for speaker recognition. Table (1) Summarize the advantages and disadvantages of Techniques used by SIS. For high efficiency and low complexity MFCC is chosen. It is useful for LPC encoding with a low bit rate. In DTW, a classification model need not be developed (or trained) in advance. GMM is helpful in classifying less memory and less planning and test.

Conflict of Interests.
There are non-conflicts of interest .

5. References

[1] Sonali T. Saste, P. S. (2017, 3). Comparative Study of Different Techniques in. *International Journal of Advanced Engineering, Management and Science (IJAEMS)*, p. 5.

[2] Amit Meghanani, A. G. (2018, 12 6). Pitch-synchronous DCT features: A pilot study on. arXiv preprint arXiv:1812.02447, p. 5.

[3] El-Samie, M. A.-E.-F.-R. (2018, 6 9). Speaker identification based on normalized pitch frequency and Mel. International Journal of Speech Technology , 21(4), 941-951, p. 12.
[4] Sonali T. Saste, P. S. (2017, 3). Comparative Study of Different Techniques in. International Journal of Advanced Engineering, Management and Science (IJAEMS), p. 5.
[5] Jyoti B. Ramgire, P. S. (2016, 4). A Survey on Speaker Recognition With Various Feature Extraction And Classification Techniques. International Research Journal of Engineering and Technology (IRJET), p. 4.
[6] W.S. Mada Sanjaya, D. A. (2018, 4). Speech Recognition using Linear Predictive Coding (LPC) and Adaptive. Journal of Physics: Conference Series, p. 11.
[7] Antony, A. &. (2018). Speaker identification based on combination of MFCC and UMRT based features. 8th International Conference on Advances in Computing and Communication (ICACC-2018), p. 8.
[8] Parvati J.Chaudhary, K. M. (2015, 2). A Review Article on Speaker Recognition with Feature Extraction. International Journal of Emerging Technology and Advanced Engineering, p. 4.
[9] Attiya, H. L., & Yousif, A. Y. (2015, 1), Mel frequency Cepstrum Coefficients and Enhanced LBG algorithm for Speaker Recognition.
[10] Yousif, A. (2016, 12). Speaker Localization and Identification using Enhanced beamforming Technique. Hilla, Babylon, Iraq.
[11] Ahmed Sajjad, A. S. (2017, 2). Speaker Identification & Verification Using MFCC & SVM. International Research Journal of Engineering and Technology (IRJET), p. 4.
[12] Al-Omari, A. A. (2016, 5). A Comparative Study of Classification Techniques for. p. 115.
[13] Pawar, R. V. (2017, 10 27). Review of various stages in speaker recognition system, performance. Analog Integrated Circuits and Signal Processing, 94(2), 247-257., p. 11.
[14] Swathy M S, M. K. (2017, 4). Review on Feature Extraction and Classification Techniques in Speaker Recognition. International Journal of Engineering Research and General Science Volume 5, Issue 2, March-April, 2017, p. 6.
[15] Ayushi Y. Vadwala, K. A. (2017, 10). Survey paper on Different Speech Recognition. International Journal of Computer Applications (0975 – 8887), p. 7.
[16] Varun Sharma, D. P. (2013, 5). A Review On Speaker Recognition Approaches And Challenges. p. 8.

الخلاصة

تكسب تقنيات معالجة الكلام شيوعًا اكثر يومًا بعد يوم لتوفير قدر هائل من الأمان.كما يشيع استخدام الكلام لغرض التوثيق. التعرف على المتكلم هو الطريقة التي يمكن من خلالها فحص المتكلم والتعرف عليه. يختلف نظام التعرف على الكلام عن طريقة التعرف على المتكلم. يشيع استخدام التعرف على المتكلمين في القطاعات والمستشفيات والمختبرات وما إلى ذلك. فوائده أكثر أمانًا وأسهل في التنفيذ وأكثر سهولة في الاستخدام. تعد طريقة تحديد المتكلم واحدة من أكثر التقنيات شيوعًا في المنطقة حيث تعتبر السلامة أمرًا بالغ الأهمية. تقدم هذه المقالة نظرة عامة على الطرق المختلفة التي يمكن استخدامها للتعرف على المتكلمين مثل الترميز الخطي التنبؤي (LPC) ، معاملات الطيف التنبؤية الخطية (LPCC) ، التحويل الحقيقي الفريد المعين (UMRT) ، معاملات Cepstral الحقيقية (RCC) ، "تردد ميل Cepstrum" (MFCC). بالإضافة إلى مجموعة من المصنفات المختلفة مثل "نموذج الخليط الغاوسي (GMM)"، "تزييف الوقت الديناميكي (DTW)" ، آلة المتجهات الداعمة (SVM) ، الشبكة العصبية (NN) ، "تكميم المتجهات" (VQ). الغرض الأساسي من شرح طرق التعرف على السماعات الشائعة. النتائج التي تم الحصول عليها هي أنه تم اختيار MFCC لكفاءة عالية ومنخفضة التعقيد. و GMM مفيد في تصنيف ذاكرة أقل ونتائج تخطيط واختبار أقل.