ARABIC SPEECH RECOGNITION BASED ON KNN, J48, AND LVQ

Nassren A. Alwahed [1], Talib M. Jawad [2]
[1] Iraqi Commission for Computers and Informatics, Informatics Institute for Postgraduate Studies, Iraq
[2] College of Information Engineering, Al-Nahrain University, Baghdad, Iraq
nassren763@gmail.com [1], talib@coie-nahrain.edu.iq [2]
Received:7/2/2019, Accepted:8/5/2019

***Abstract*- Most systems of speaker recognition work on speech feature primarily classified of being a low level which considerably relies on speaker physical characteristics and, to the lower extent, the acquired speaking habits. In this paper present a system to recognition and identification in Arabic speaker. It includes two phases (training phase and testing phase) each phase includes the using of audio features (Mean, Standard Division, Zero Crossing, Amplitude). after get the feature, the recognition step is using (J48, KNN, LVQ),) where the Nearest Neighbor (KNN) applied o get the similarity of the data training and data testing , LVQ neural network used for Speech Recognition and Arabic language Identification. This sentence contains words especially kidnappings and kidnappers are ten sentences and pronounce these sentences by 10 people, five men and five women of different ages and each of the ten pronunciation of all sentences, so a total of 100 samples and the samples were recorded on audio and wave. The results of the sentences pronounced by women are higher than the results of the same sentences pronounced by men. They achieved better recognition rate 85, 93, 96.4% .**

keywords: **Speaker recognition, K- nearest neighbor(KNN) , Learn vector quantization (LVQ), J48.**

I. INTRODUCTION

Clearly, identical speakers present a difficult task to the technology biometrics that depends on the measuring and distinguishing essential physical properties such as fingerprint, face, iris, voice, of individuals in order to conduct the process of human recognition. This issue served as an important stimulus for performing many multi- disciplinary research efforts, which concentrated considerably on the measurement the similarity extent of biometric characteristics in speakers that is mentioned earlier [1]- [5]. The aim of the research efforts is to identify the influence of identical speakers. speech in the applications of automatic speaker verification. That is to be conducted through making comparison of the speaker error rates by matching experiments between speakers and non-speakers, in addition to the percentage of speaker misidentification in identifying the speaker for each experimental setting. The test basis composes of an i-vector framework on a great database of speakers' audio samples, whether it is read or conversation sample, kept in several recording devices. Practically, verification experiments cover 100 pairs of speakers in sever AL periods and train-test conditions. the sample scenarios almost present an accurate representation of commercial and forensic verification scenarios where duration may vary between microphones and sessions. The results will be compared later with a similar group of unrelated persons.

II. RELATED WORK

There have been numerous applications of automatic speaker identification which can be categorized as commercial applications, like voice- mail, telephone banking, biometrical authentication, and forensic applications. In the application of forensic there are different works of speaker recognition for different languages. [6] propose an FSR system for Hindi Words in which the MFCC technique and vector quantization model are used for extracting the features and pattern matching, respectively. This system achieves about 90% rate of success through the experiment.[7] proposes a simple method to the

recognition of text dependent speaker and are based on the Symlet wavelets for extracting features. Those features are after that categorized with the use of data mining techniques. In this work, J-48, Naive Bayesian and SVMs have been utilized to classify features, and the results have shown that classification precision of up to 86% has been accomplished by the classifiers. The researcher in work [8] used the same concepts which are used in work [6], and they are focus on the problem of finding a trade- off between the time duration of speech sample and the probability of error. [9] presents an FSR scheme which use a combination of MFCC and its delta derivatives DMFCC and DDMFCC. also, the probabilistic neural network (PNN) in the modeling domain is used to achieve lower operational times during the training steps. In this scheme, the feature set attains a recognition precision of 94%. Working in Arabic language, [10] present an Arabic FSR approach by using GMM-UBM, thirty- nine MFCCs for feature extraction, and all experiments are conducted using the KSU Speech Database. This approach has shown that the test sample of a suspect can be recognized, within a noisy environment, with few seconds of speech, and at different times of training and testing. But this approach did not show much improvement; this is mainly due to the big invariance between channels. [11] propose an audio signal classification to determine gender class (male or female). This algorithm is based on 4 extracted features. Mean, Zero Crossing rate, Standard Deviation, and Amplitude. The values of those properties are categorized into two categories using one of the automated algorithm supported by the support.

III. BACKGROUND THEORY

A. *Speaker recognition*

The purpose of automatic speaker recognition is to identify a person from pronounced speech, this work can be further turned to the process of the identification or verification where the first matching is 1: 1 (claiming an identity and matcher task is to accept or reject the individual), or the latter, which is a 1: N match (the identification of speaker depends on making comparison to N registered speakers) [12]. Fig. 1 provides a common guidance and description of a system of voice verification, which is very clear in the diagram below. Speaker recognition technologies can be turned into text- dependent; where the words or sentences that help to be recognized are identified or known in advance or they are text- independent (no previous knowledge available of the uttered sentences). Seemingly, text-independent is seen much more difficult task to achieve in speaker recognition system. Generally, choosing the features of any Speaker Recognition system is regarded a key matter in order to obtain an accurate performance, where this system must have high variability between speakers and low variability within each person,in addition to other standards [12].

There are body- related and cognitive aspects and they play an important role in the determination of the uniqueness existed in aperson's speech. The aspects can be categorized according to a range of features ranging between high to low-level. Relying on the use of this application, it is possible to anyone to choose a specific attribute or a combination of those attributes [13]. Clearly, High- level features have their own features such as the speaker's specific word uses, in addition to speech styles. The traits are described of being environmentally- influenced that anyone can acquire them by education, place of living, and the social or family environment, socio-economic status. Remarkably, these features could

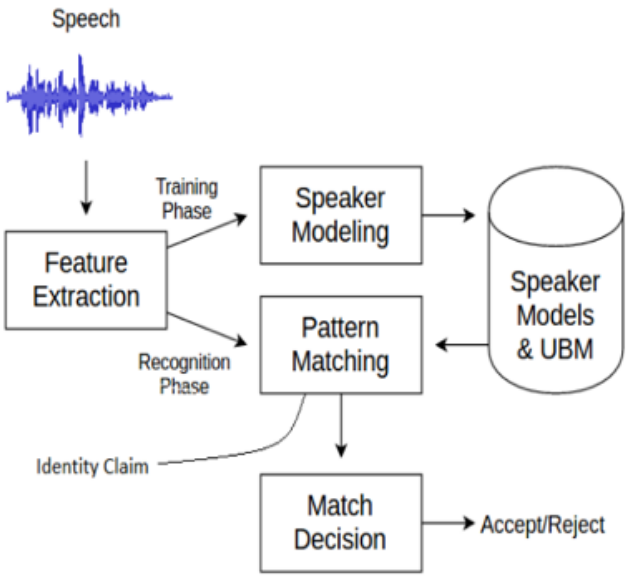

Figure 1: Speaker verification diagram adapted from [12]

have many advantages and the most striking one is robustness to channel effects and noise and. However, they entail advanced extraction techniques[14].

B. k-Nearest Neighbor

The KNN is a method which depends on method called supervised learning, where it used to allow machines to categorize objects, problems or situations relying on related data already fed into the machines. The similarity distance measuring between the data is the depended method of KNN on the categorization of data. The following is the algorithm of KNN [15].

C. J48 Decision Tree

Classification technique is an approach of creating a model of classes from a collection of records containing class label. Decision Tree Algorithm is to know how the attributes- vector behave for an arrays of instances, and it also based on training instances, the classes for the newly created instances are reached. This algorithm basically participates in producing rules to predict the variable of target. However, by obtaining assistance from the tree classification algorithm, the critical distribution of data can be simply understood. J48 is seen as an extension of ID3. Additional features of the J48 are standing for missing values, derivation of rules, continuous attribute value ranges and decision trees pruning etc. In the mining tool for Waikato Environment for Knowledge Analysis data WEKA, J48 is considered as an open source Java application for the C4.5 algorithm. The WEKA tool contributes in providing an array of options which are basically related with tree pruning. If possible over fitting pruning maybe utilized as a device for precising. In other algorithms, the

classification is repeated again and again meaning that the classification of data have to be as perfect as possible. This algorithm participates in producing the rules by which a specific identity is formed for that data. The goal is to bring the decision tree to common use gradually until it gains balance of flexibility and accuracy[15].

D. Learning Vector Quantization

The error back-propagation algorithm is a method widely used in studies regarding artificial neural networks, especially in the case of multilayer perceptron. On the other hand, it has been shown that the easy and rapid training of learning vector quantization which is developed by Kohonen, achieves high discriminatory power such as multilayer perceptron trainers or back-propagation algorithm or Boltzmann apparatus. A number of versions of learning vector quantization can be classified into three versions, such learning vector quantization 1 learning vector quantization 2 and learning vector quantization 3. Obviously ,learning vector quantization can be seen as equal to the algorithm of adaptive learning for a traditional , multi-reference distance classifier used for the purpose of classification to static vector inputs, even if this concept had been developed in the first place in self-organization network framework, here, adaptive learning concept involves a learning scheme that frequently participates in making the small system adjustment incurred for every presentation of a training sample. In all versions mentioned above, which each one has the same dimension as the input, are designated to each class. Typically, the initialization of these reference vector occurred through the use of a traditional method such as Ic-clustering. In the learning vector quantization training phase, an adjustment process is made to the reference vectors therefore all inputs of training have a valid reference vector for the class as its closest reference. It is evident that all versions have their corresponding adjustment rule, so each one [version] can be characterized according to its own corresponding adjustment rule. In regard with decision of classification, when an input vector is unknown, it can be categorized by reaching the reference vector that is the closest to the input vector. Here, we will provide full description of the algorithm of learning vector quantization 2, which is already used in the hybrid algorithm. Originally, learning vector quantization 2(learning vector quantization 2 -E) was defined primarily by use of the formula of squared Euclidean distance, it is well-known distance measure which uses the metric to calculate speaker similarity. Since the measure of Euclidean square distance is not considered as scale-invariant, it makes it difficult to process vector components, where each component has its own diverse value range Apparently, the fact that says squared Euclidean distance is seen as aversion of likelihood-based distance measurement which was simplified to a large degree can be as reminder for us of the method of likelihood-based distance, which excellently employs the available information about the probability distribution model, may play an important part in overcoming this scale problem and assists in the progress of estimating class boundaries accurately [16].

IV. METHODOLOGY

A. Data Set

The samples that were used in the search database is an Arabic- language sentences are configured as private Arabic-language pronunciations phonemes. This sentence contains words especially kidnappings and kidnappers are ten sentences and pronounce these sentences by 10 people, five men and five women of different ages and every one of those ten bands all sentences, so a total of 100 sample and the samples were recorded on audio and wave extension form and address

several samples preprocessing, feature extraction, classification table below contains sentences. Table 1 show sentences in Arabic.

Sentences
تَعَرَّضَ وَلَدُ الْجَارِ لِلْخَطْفِ
خَطَفَ وَلَدًا مِنْ وَسَطِ الْحَيِّ
اذا لم تعطي المبلغ نقتل ولدك
لا تقلق ولدك بامان عندنا
المبلغ المطلوب خمسون الف دولار
ضع المبلغ في المكان المحدد
لا تبلغ الشرطة والا يتم قتله
اجلب المبلغ في الوقت المحدد
تجد ولدك في المكان المتفق عليه
لديك يومان لتسليم المبلغ

V. Feature Extraction

In Different Domains The most significant process in the FSR system is the feature extraction which is used to get best speaker distinguishing and provide good accuracy rates. In this paper, many features have been discovered in frequency and spatial domains by utilizing the MFCC and VQ, and (mean, standard division, Coefficients (MFCCs) are a popular approach for feature extraction which is utilized in speech identification on the basis of frequency domain utilizing the Mel scale that is on the basis of zero crossing, amplitude). Mel Frequency scale of the human ear [5]. Here, the speech signal is initially split to time frames that Cepstral consist of a random number of samples. Whenever frame is then windowed using Hamming window for eliminating edge discontinuities [11]. The filter coefficients w(m) of a Hamming window of length m are calculated based on the equation: Here, M is whole number of samples and m is the current sample. After the windowing, to speed up the processing, Fast Fourier Transformation (FFT) is computed for all frames for extracting elements of frequency of a signal in the time domain. The logarithmic Mel- Scaled filter bank is basically applicable to the frame that is transformed. This scale is roughly linear up to 1 kHz, and logarithmic at higher frequencies. The final step is for converting the log mel spectrum back to time. And this is performed by utilizing the Discrete Cosine Transformation (DCT) of the outputs from the filter bank Vector Quantization (VQ) is used for changing the quantization dimension from one (for scalar) to multi (for vectors). In this paper, the VQ is utilized to improve the MFCC by minimizing the data of the extracted feature.

VI. MATCHING

$$w(n) = \begin{cases} 0.54 - 0.46cos\left(\frac{2m}{M-1}\right), 0 \leq m \leq M-1 \\ 0, otherwise \end{cases} \quad (1)$$

The actual reference method to match voice for speech categorized independent is the KNN, J48 , LQV methods. This approach of matching is working by fitting the input features as shown in description below The Proposed Speaker System system consists of two phases applied to Arabic speech sentences, after applying the cross- validation to the dataset these two phases will be applied. Algorithm 1 and Fig. 2 show the general architecture of the proposed system.

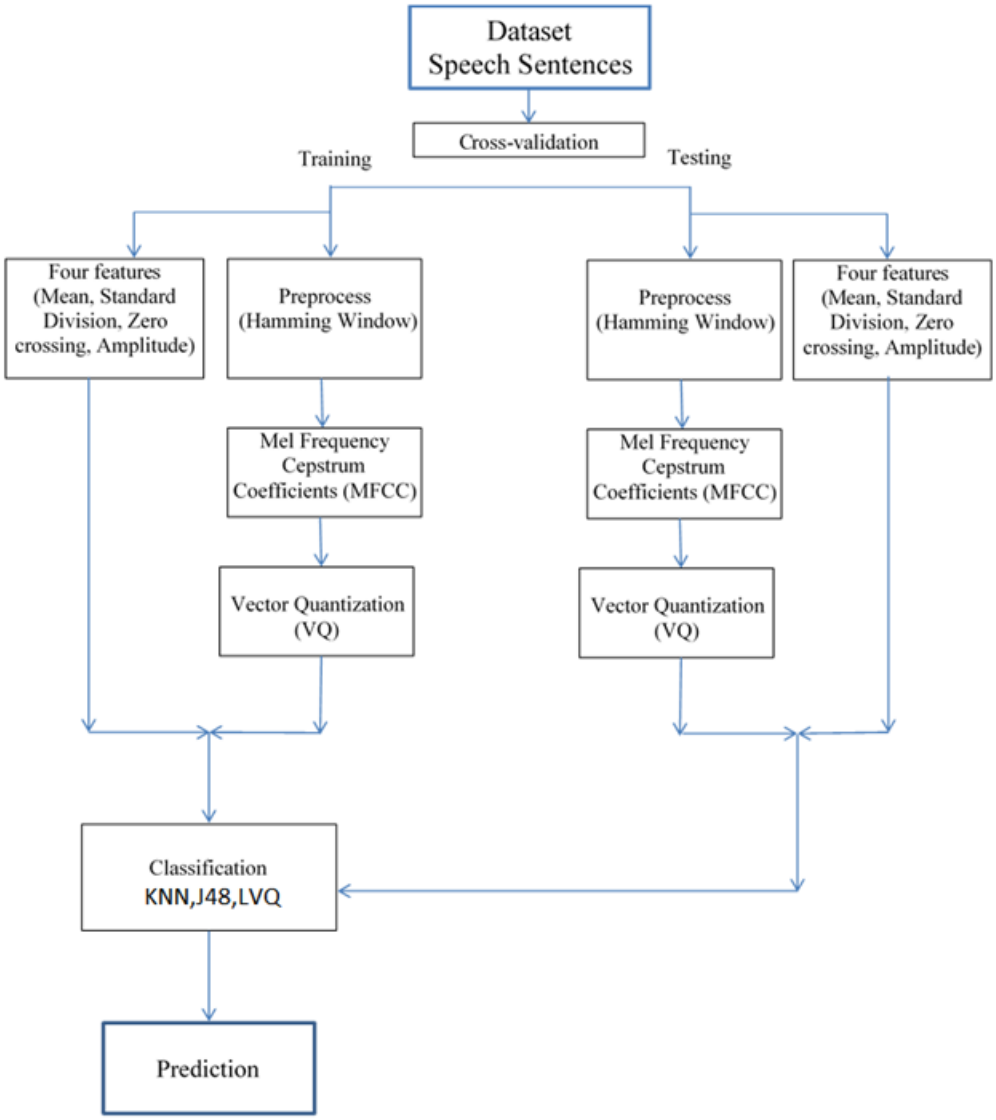

Figure 2: The general architecture of the proposed system

VII. K-FOLD CROSS-VALIDATION

It is a statistical technique that involves splitting data into subsets, training data on a subset and using the other subset to evaluate model performance. To reduce contrast, we conduct multiple rounds of cross-checking with different subsets of the same data.

VIII. THE TRAINING PHASE

The first phase of the proposed system is the training of the dataset which will be produced from the cross validation. In this phase the dataset will be processed using, preprocessing using hamming window, then, the features are extracted and modeled by using MFCC, VQ and these features values are mixed with the features extracted from (Mean, STD, Zc, and Amp) in order to prepare it to the classification algorithms. Algorithm 2 shows the details of this phase. Through the phase of training, the mixed features are stored as reference templates. These templates are then compared with the entered speech signals

IX. THE TESTING PHASE

The second phase of the proposed system is the testing phase which will process the remained speech sentences dataset which result from the cross validation processing. In this phase the dataset will be processed using preprocessing; hamming window, and then, the features are extracted and modeled by using MFCC, VQ and these features values are mixed with the features extracted from (Mean, STD, Zc, and Amp) in order to prepare it to the classification algorithms. Algorithm 3 shows the details of this phase.

X. EXPERIENTIAL RESULTS

Table I and II show the Average results for five speakers, of sex male, and female under three classification techniques KNN, LVQ, and J48. These obtained results are based on the measurements of Mean, STD, Zc, and Amp. Table I Average results for five speakers, Sex: Male

TABLE I
AVERAGE RESULTS FOR FIVE SPEAKERS, SEX: MALE

Feature	Type algorithm		
	KNN	LVQ	J48
MFCC	91.3	97.7	91.5
MFCC- MEAN	90.5	98.8	92
MFCC- STD	94.8	98.8	97.2
MFCC- ZC	91	97	92
MFCC- Amp	90.5	96.9	97.2
MFCC- all	93.6	96.8	99

TABLE II
AVERAGE RESULTS FOR FIVE SPEAKERS, SEX: FEMALE

Feature	Type algorithm		
	KNN	LVQ	J48
MFCC	98.1	98.1	96.1
MFCC- MEAN	98.1	98.1	98.1
MFCC- STD	98.1	98.1	96.5
MFCC- ZC	98.1	98.1	96.3
MFCC- Amp	98.1	98.1	98.1
MFCC- all	98.1	98.1	98.1

The results in the above tables show that the recognizing rates in male are greater than female when using the J48 classification technique and the same samples, where the greater achieved rate in the male is 99%. And we noticed that the using of KNN result in low rates in male, 93.6%.

XI. Conclusions

The proposed system work on the forensic speaker identification for Arabic language. In this system, the VQ works to improve the MFCC technique. The extracted features which results from the mixing of use the MFCC technique and (Mean, STD, Amp, and Zc) has given good results after being applied to a number of classification algorithms like; KNN, LVQ, and J48 with a better recognition rate; in male, 93,6%, 98.8%, and 99%, respectively, and in female, 98,1 %, 98,1%, and 98.1%, respectively. The processing time is reduced due to the using of the cross-validation. In the future, we will work on java languages .Since the proliferation of voice recognition technologies and the high rates of speaker birth, it becomes very important to conduct such efforts to ensure the level of reliability and security of all people's services and resources. Future procedures may include other artificial neural networks in order to participate to a larger extent in improving the accuracy and precision of systems to distinguish identical speaker. Furthermore, high-level speech features may be seen as a greater discriminating power among related persons as they do not depend on physical aspects.

References

[1] Z. Sun, A. Paulino, J. Feng, Z. Chai, T. Tan, A. Jain, "A study of multibiometric traits of identical twins" , Proceedings of Biometric Technology for Human Identification, vol. 7667, 2010, p. 76670T
[2] K. Hollingsworth, K. Bowyer, and P.J. Flynn, Similarity of iris texture between identical twins, Computer Vision and Pattern Recognition Workshops, 2010, p. 22-29
[3] P. J. Phillips, P. J. Flynn, K. W. Bowyer, R. W. V. Bruegge, P. J. Grother , G. W. Quinn, M. Pruitt, "Distinguishing identical twins by face recognition" , IEEE Int. Conf, on Automatic Face and Gesture Recognition and Workshops, 2011, p. 185- 192
[4] A. K. Jain, S. Prabhakar, S. Pankanti, "On the similarity of identical twin fingerprints" , Pattern Recognition 35 (11) (2002) 2653- 2663
[5] J. Hu, J. Lu, Y.-P. Tan, Fine- grained face verification: Dataset and baseline results, Int. Conf. on Biometrics, 2015, p. 79- 84
[6] Nitisha, and Ashu Bansal, "Speaker Recognition Using MFCC Front End Analysis and VQ Modeling Technique for Hindi Words using MATLAB" , International Journal of Computer Applications, Vol. 45, No. 24, pp. 0975- 8887, 2012.
[7] V. Srinivas, Ch. Santhi rani and T. Madhu, "Investigation of Decision Tree Induction, Probabilistic Technique and SVM for Speaker Identification" , International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 6, No. 6, pp. 193- 204, 2013.
[8] Nimesh V Bhimani, " Speaker Recognition System Based On MFCC and VQ Algorithms" , International Journal of Engineering Research and Technology (IJERT), Vol. 3,No. 2, 2014.
[9] K. S. Ahmad, A. S. Thosar ,J. H. Nirmal and V. S. Pande, " A unique approach in text independent speaker recognition using MFCC feature sets and probabilistic neural network" , 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), Kolkata, pp. 1- 6, 2015.
[10] Mohammed Algabri, Hassan Mathkour, Mohamed A. Bencherif, Mansour Alsulaiman, and Mohamed A. Mekhtiche, "Automatic Speaker Recognition for Mobile Forensic Applications" , Mobile Information Systems, Vol .2017
[11] Nidaa F. Hassan, Sarah Qusay Selah, "Gender Classification based on Audio Features", Al - Ma'amoon College Journal, Vol. 31, 2018.
[12] E. San Segundo, H. K nzel." Automatic speaker recognition of spanish siblings:(monozygotic and dizygotic) twins and non- twin brothers" , Loquens 2(2)(2015)021
[13] M. Sullivan, Global Markets and Technologies for Voice Recognition, Information Technology Market Research Reports in BCC Research, January 2017.
[14] Kaghyan, Sahak and Hakob Sarukhanyan, "Activity Recognition Using K-Nearest Neighbor Algorithm On Smartphone With Tri-Axial Accelerometer" ,1(2012).
[15] Nadali, A, Kakhky, E. N. , Nosratabadi, H. E. , "Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system" , Electronics Computer Technology (ICECT), 2011 3rd International Conference on, vol. 6, no. , pp. 161, 165, 8- 10 April 2011.
[16] Higeru Katagiri, "A New Hybrid Algorithm for Speech Recognition Based on HMM Segmentation and Learning Vector Quantization" , IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. I, NO. 4, OCTOBER 1993.