

A Holy Quran Reader/Reciter Identification System Using Support Vector Machine

Khalid M. O. Nahar, Moyawiah Al-Shannaq, Ahmad Manasrah, Rafat Alshorman, and Iyad Alazzam

Abstract—Holy Quran Reader Identification is the process of identifying the reader or reciter of the Holy Quran based on several features in the corresponding acoustic wave. In this research, we build our own corpus, which contains 15 known readers of the Holy Quran. The Mel-Frequency Cepstrum Coefficients (MFCC) are used for the extraction of these features from the input acoustic signal. These MFCCs are the reader's features matrix, which is used for recognition via Support Vector Machine (SVM) and Artificial Neural Networks (ANN). According to our experimental results, the Holy Quran Reader Identification System identifies the reader with 96.59% accuracy when using SVM, in contrast to accuracy of 86.1% when using ANN.

Index Terms—Reader identification, feature extraction, Mel-Frequency Cepstrum Coefficients (MFCC), support vector machine (SVM), Artificial Neural Networks (ANN).

I. INTRODUCTION

The process of reader identification is a general process that covers a wide area in voice recognition systems. For instance, the reader identification process uses the reader's voice as a biometric signal to recognize an individual's identity from the underlying speech information. The process of reader identification can be used for a broad range of applications such as forensic tests, automatic indexing of speech databases, voice dialling, and in enhancing speech recognition systems, among others. However, although there have been numerous efforts to increase the accuracy of such systems, the identification process still lacks the accuracy required in the recognition of voices, especially in noisy environments.

Generally, speaker recognition applications can be divided into speaker identification and speaker verification categories. Speaker identification determines the identity of an anonymous speaker based on the his/her utterances. On the other hand, speaker verification verifies the speaker's identity from his voice characteristics [1]. Speaker recognition usually requires the speaker to say the same utterance as a text-dependent identification system, or without speech constraints as a text-independent identification system. Over the past few decades, significant research has been conducted to improve text-independent speaker identification performance in terms of accuracy and

systems robustness, special conditions where there, they are noise and vibration. The research in this field focuses on converting a caustic wave form into features that are useful for further processing to identify the owner. The strength of the state-of-the-art algorithms used depends on the features' capability of capturing time frequency and energy into a set of coefficients for cepstrum analysis [2]. Several studies have achieved promising results in identifying the speakers from acoustic waves by using different extraction methods and classification schemes such as those based on vector quantization [3], the Gaussian mixture model [4]-[6] and Gamma Tone Frequency Cepstral Coefficients (GFCC) with I-vectors [7], [8]. It is, however, important to extract features from each frame to capture speaker characteristics. Features such as the Mel-frequency cepstral coefficients (MFCC) [9], [10] and perceptual linear predictive (PLP) coefficients [11], [12] are commonly adopted for speaker identification and recognition algorithms.

The human voice conveys a great deal of information such as the gender, emotions, and identity of the speaker, which is converted into a digital signal to produce digital data representing each level of signal at every discrete time step [13]. The digitized speech samples are converted into the Mel-Frequency Cepstrum Coefficient (MFCC) to produce the features of the wave vectors, which can be then used in learning algorithms such as SVM, and ANN. Even though most of the known acoustic-property-based algorithms have achieved an almost 100% level of accuracy in an uncluttered environment [14], [15], the performance of these algorithms is degraded with channel variations due to the microphones, used or environmental or background distortions [14].

In this paper, a Holy Quran reader/reciter identification model is proposed. Unlike any other voice or speaker identification system, the voices of the readers who recite the Holy Quran and the way they do so is very different from the voice and delivery of ordinary speech. The difference is due to the existence of certain acoustic rules, known as "Ahkam Al Tajweed" "أحكام التجويد", that must be applied when reading the Holy Quran aloud. X). Furthermore, emotional features, which are added by the reader, as well as the transition from one pitch or tone to another when reading the Holy Quran distinguish such readings from normal speech.

The transition from one pitch to another is known as Maqam (مقام). This very important feature must be maintained while reciting the Holy Quran. As a result, the proposed model converts the readers' sound waves into MFCC to extract the feature vector matrix. The proposed model has been practiced and tested using both SVM and ANN. SVM is widely used and has proved to be effective for speaker recognition [16].

This paper also contains a related work in Section II. the Methodological steps and algorithm are presented in Section

Khalid M. O. Nahar, Moyawiah Al-Shannaq, Ahmad Manasrah, Rafat Alshorman are with Department of Computer Sciences, Faculty of IT and Computer Sciences, Yarmouk University, Irbid, 21163, Jordan (Corresponding author: Khalid M. O. Nahar; e-mail: khalids@yu.edu.jo).

Iyad Alazzam is with Department of Computer Information System, Faculty of IT and Computer Sciences, Yarmouk University, Irbid, 21163, Jordan.

III, while Section IV illustrates the results of the experiments. Section V concludes the paper and sheds light on future research potential.

II. RELATED WORK

The nature of the recitation of the verses of the Holy Quran is different from that of any other type of speech or utterance. To date, to our knowledge, no one has addressed the area identifying the speech of Holy Quran readers., which is not true of other types of voice and speaker identification systems. Different approaches exist in the literature for using the MFCC features along with some learning algorithms such as the Hidden Markov Model (HMM), SVM, and ANN among others. One example, [2], is the design of a small-scale system for speaker and speech recognition with specific visual information for the user. Through the HMM, the authors successfully recognized isolated words where the speech recognition was achieved through extracting a set of feature vectors using MFCC. The Vector Quantization (VQ) was used for the practice and classification of the features. The accuracy of discrete speaker and/or speech recognition was 90% over a database of recorded voices. Likewise, [17] a further example presented a real time speaker recognition system using MFCC and VQ. This research was evaluated with 120 speakers with a TIDIGIT database. The proposed real time system uses the LBG algorithm proposed in a further example [18] to group

the set of acoustic vectors into sets of codebook vectors. The experimental results showed 91% accuracy in normal environmental conditions at 20db SNR.

However, the proposed system was tested and assessed using a sample of only seven speakers. This limitation can be addressed through other matching techniques such as HMM and delta spectral Cepstrum coefficient as suggested by the authors. Similarly, [19] a proposed voice recognition approach based on MFCC features and SVM learning algorithm for speaker identification is suggested. The proposed approach was practiced over a database of eight speakers/readers. The feature vectors were extracted using MFCC from the input speech. The feature vectors are used as the speaker features for learning and matching using SVM. The accuracy rate reached was 95%. This result suggests the use of ANN or GMM techniques to improve performance and increase accuracy.

Each language has its own phonetic system. The Arabic sound system, which is different from the English sound system, has a unique phonetic structure as each spoken word has a phonetic combination of certain sound units. Therefore, various scholars have proposed methods for an Arabic -based speaker or voice recognition system. For instance, a new feature called Multi Directional Local Features (MDLF) was suggested and applied on automatic speaker recognition [20]. The MDLF showed significant improvements compared to conventional features. The authors also focused on how to select optimum features from a 48- dimensional MDLF to increase the speaker recognition rate and the possibility of applying the MDLF features to Arabic speech recognition systems. In Arabic, some authors like the one in [21] suggested that knowing the length of a specific phoneme can assist in speeding up the recognition process as well as increasing recognition accuracy. As a result, the authors have presented a statistical study and analysis of the Arabic phonemes. They have also utilized the previous assumption, that knowing the length of a specific phoneme would be helpful, to limit the length of an HMM chain to increase recognition accuracy. They have also improved the manner of searching for phonemes through adopting the median length of each cluster during the recognition phase. However, the proposed statistical analysis requires manual tuning of the recognition and configuration rates for building the best acoustic model.

Since the voice is a signal of infinite information, direct analysis and synthesizing of the complex voice signal is required. To represent the voice signal, various Feature Extraction and Feature Matching algorithms are introduced. Therefore, [13] based on an assumption that the voice signal may have different temporal rates, they propose a non-parametric method for modelling the human auditory perception system. They utilize the MFCC as a feature extraction technique along with the non-linear sequence alignment known as Dynamic Time Warping (DTW) as a feature- matching techniques. The proposed method proved to be an efficient voice recognition algorithm. However, with constant window spacing, the lengths of the input and stored voice signals are unlikely to be the same. Moreover, there will be a variation in the length of individual phonemes that has not been addressed in the proposed method.

Table I summarizes some of the studies which have addressed speaker identification in their research.

TABLE I: LITERATURE SUMMARY

Author/year	Highlights	No. of Speakers	Database	Recognition rate
(Das and Nahar, 2016)	Hidden Markov Model (HMM) to recognize isolated word, Mel Frequency Cepstral Coefficients (MFCC) and (VQ) for speech recognition using MATLAB.	40	Three hours of news recordings	speech recognition has been successfully accomplished with about 90% accuracy
(Bharti, 2015)	Mel Frequency Cepstrum Coefficient (MFCC) and Vector Quantization (VQ) using MATLAB.	120+5 local	TIDIGIT database	The experimental results showed 91% accuracy in normal environmental conditions at 20 db SNR
(Nijhawan and Soni, 2014)	Support Vector Machine (SVM) for speaker identification and feature extraction by using (MFCC).	8	local	the accuracy rate is 95%
(Alsulaiman et al., 2017)	A new feature called Multi Directional Local Features (MDLF) is proposed and applied on automatic speaker recognition.	49	TIMIT database	(MDLF) showed significant improvements compared to MFCC and LF
(Nahar et al., 2012)	A statistical study and analysis of Arabic phonemes designed to improve the performance of current ASR systems.	40	Three hours of news recordings	increase the speed of the clustering of the phonemes and increase the recognition accuracy
(Linde et al., 1980)	The use of MFCC and DTW) to increase the voice recognition performance.	2	Local	the results show that these techniques could be used effectively for voice recognition

In our research we focus on identifying the reader name by exploring his vocal features, which is considered as a challenge in voice recognition systems. Identifying Holy Quran reciter or reader is not like identifying a normal speaker, since the original voice of Holy Quran reciter will not be as is, it is governed by "Ahkam Al Tajweed" "أحكام التجويد". Identifying the person name through his vocal features is one of the main objectives of via voice systems. Sometimes we may find some audio files without any annotation, for example Holy Quran Audio files with unknown reader. It is beneficial to know the reader or reciter name of this files, else we should mark them as unknown reader. In our research we contribute in the direction of annotating unannotated Holy Quran audio files by the corresponding reciter. As mentioned before SVM is widely used and has proved to be effective for speaker recognition [16].

III. THE PROPOSED APPROACH

The speaker recognition's main components are feature extraction, and speaker modelling. Fig. 1 illustrates an overview of the proposed identification system. The proposed system mainly consists of practice and testing phases.

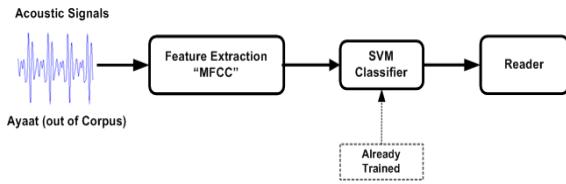


Fig. 1. Proposed approach overview.

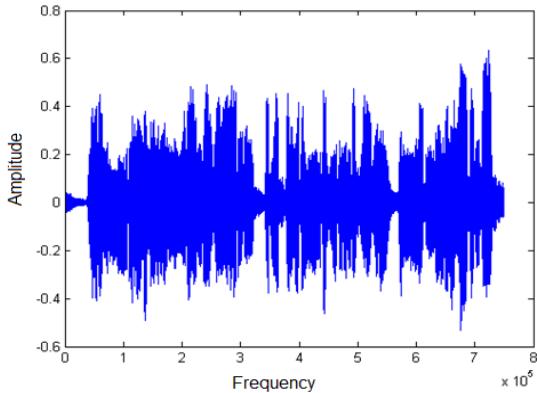


Fig. 2. Sample of acoustic wave.

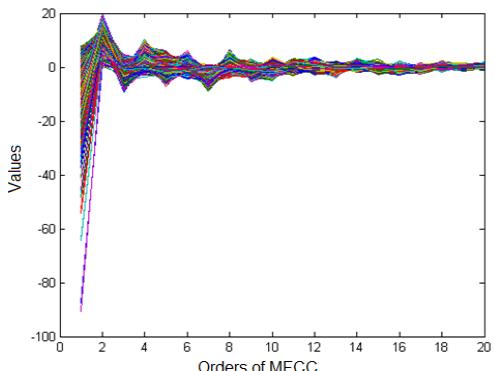


Fig. 3. Sample of MFCC file.

Feature extraction converts the input speech signal into a

suitable feature space to be modelled. The first step of the proposed approach after extracting the features is to build a trained model. Since the proposed approach aims to identify the Holy Quran reader; samples from the Holy Quran recited aloud are required to build the corpus. An example of acoustic waves and their features is illustrated in Fig. 2 and Fig. 3 respectively. Acoustic waves for each reader, based on his readings, were collected. The feature vectors of these acoustic waves were extracted.

A. Data Acquisition and Corpus Construction

The Holy Quran is the Holy Scripture for Muslims as it is a comprehensive guide to a lifestyle based on what is written there. It is the collection of the verbal revelations given to Prophet Muhammad taught to believers over a period of twenty-three years. The Holy Quran was revealed in Arabic with 114 Chapters (Surahs) of varying lengths, and each Chapter consists of several individual verses (Ayat). There are 6,348 different verses in the Holy Quran. To date, there is no corpus for the Ayat of the Holy Quran based on their readers/reciters' performance. In this research we have chosen to segment two different Surahs, the first 120 Ayah from Suraht Al-Baqraa (chapter 1) as outlet data and Suraht Al Kahaf "سورة الكهف" (Chapter 18) into 110 separate Ayah for a specific reader as the corpus. In total, 15 different readers were considered for the corpus construction, as detailed in Table II. Each Holy Quran reader recites the 110 Ayah from Suraht Al Kahaf. A wave file was created for each Ayah. The total number of acoustic waves created is 1650 waves. The recognition system was implemented in MATLAB. The speech signal was divided into segments of 15 ms frames with the use of a Hamming window for further analysis. We experimentally determined that 15 ms frames generated better recognition because the Holy Quran readers generally have a slow rate of recitation. A larger frame size does not enhance the recognition system's ability to learn the characteristics of the signal.

TABLE II: HOLY QURAN READERS CITED IN STUDY

Reader Name	Arabic Name	Samples from Al-Kahaf	Samples from Al-Baqara
Mohammad Rashad.	محمد الراشد	110	120
Maher Almuaiqly	Maher المعيقلي	110	120
Khaled Alqahtani	خالد القحطاني	110	120
Ahmad Alhuthaifi	أحمد الحذيفي	110	120
Ali Alhuthaifi	علي الحذيفي	110	120
Ahmad Alajamy	أحمد العجمي	110	120
Yahya Hawa	يحيى حوى	110	120
Abdullah Basfar	عبدالله بسفل	110	120
Salah Albudair	صالح البدير	110	120
Yasser Aldosari	ياسر الدوسري	110	120
Fares Abbad	فارس عباد	110	120
Abdulrahman Alsudaes	عبد الرحمن السعدي	110	120
Meshary Al-Afasy	مشاري الفقي	110	120
Saud Shuraim	سعود شريم	110	120
Saad Alghamidi	سعد الغامدي	110	120

Since the Ayat are of different lengths, a 20-feature representation of each Ayah was extracted for the 1650 acoustic waves as an MFCC file corresponding to the acoustic waves as discussed in Section III.B. Fig. 4 shows the feature vectors of the whole Ayat.

Each ayah vector is transposed and labelled by the corresponding reader, as illustrated in Fig. 5.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-55.1683	-54.3575	-51.4150	-47.8614	-35.1240	-30.7448	-32.2724	-33.2354	-32.9174	-34.1820	-35.3730	-35.3371	-33.4985
2	4.9671	4.8545	4.5540	4.8081	7.9815	8.8882	9.1816	8.8391	10.0779	9.6682	9.0567	6.6015	8.3921
3	-0.4850	-0.8994	-1.1692	0.9392	1.5180	2.5573	1.9791	0.7630	1.0817	0.9784	-0.1995	1.1566	
4	-0.1677	-0.3894	-0.5723	-0.5238	2.6716	1.5008	-1.2273	-0.0918	0.6134	0.4301	0.4260	-0.2386	0.5766
5	-0.1350	-1.0238	-0.9653	-1.4984	-0.9588	-0.6209	-1.4468	-2.0344	-1.8723	-1.8817	-0.9953	-1.1459	-0.6688
6	0.9713	0.5030	1.6765	0.4643	0.3708	2.1740	0.4943	1.9794	2.9947	2.0271	2.0476	2.3713	
7	0.0716	-1.2867	-1.7126	-1.1019	-1.0785	-0.9136	-1.1033	-1.7855	-0.6965	-1.3884	-1.1551	-1.5551	
8	0.7547	0.3814	-0.0039	-0.0833	1.4208	2.9068	1.8089	1.9690	1.7072	1.4630	1.8351	2.5160	2.1325
9	-0.5826	0.0465	0.5281	-0.3673	-1.3271	-1.3548	-1.1672	-1.6099	-0.7134	-0.9526	-0.7082	-0.1082	-0.8944
10	0.8526	0.5272	1.3723	0.9842	1.5510	1.5118	1.1008	0.9584	0.9117	1.5196	1.4473	0.4103	0.0569
11	0.2054	0.3248	0.1508	-0.6827	0.6550	-1.1771	-2.1258	-2.2005	-0.9746	-2.3188	-2.1581	-2.6374	
12	-0.1627	0.4634	1.2243	-0.7175	0.3881	0.8943	1.3485	1.3052	1.3266	0.9110	0.5611	0.7838	1.1453
13	-1.1869	-0.0774	-0.2165	-1.3071	-1.2668	-1.3869	-0.9340	0.6453	0.4801	-1.0693	-0.2515	-0.0466	0.3921
14	0.2625	0.0757	0.0655	-0.3512	0.1621	0.4348	1.0463	1.4003	1.0524	0.8911	0.9092	1.3110	1.2994
15	-0.0283	-0.2393	-0.4679	-0.0702	0.2925	-0.5260	-0.5659	-0.0861	0.1955	-0.2546	-0.3336	-1.0000	
16	0.3919	0.1949	0.5072	1.2355	0.1358	1.2881	0.8550	0.7835	1.1850	1.0670	0.8514	0.9454	
17	0.5645	-0.0426	-0.5043	-0.0709	0.1306	-0.2846	-0.9712	-0.3230	-0.4291	-1.1742	-0.3751	-0.5887	0.2321
18	-0.5028	-0.5860	-0.6365	0.2119	0.1438	0.2061	-0.9660	-0.6764	-0.5626	-0.7095	-0.3423	-0.3521	-0.6248
19	0.3215	-0.1612	-0.2551	-0.0481	-0.6701	-0.6206	-0.3859	-0.9148	-0.7319	-0.8356	-1.1694	-1.3060	

Fig. 4. Screen shot of the feature vectors matrix of all readers (20*22952).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
7775	Mohammad Rehi	-31.7404	11.0899	4.1470	-0.8517	0.627508	-1.5952	4.040147	-1.2965	-1.9012	1.419544	-1.34848	1.625929	-1.10594	0.654317	0.155391	0.728986	-0.2895	0.9564	0.517	
7775	Mohammad Rehi	-31.8590	11.5574	5.025169	-1.83035	-1.2685	-3.37409	0.593048	-0.73533	1.505801	-1.72687	0.768203	-1.50619	1.745753	-0.88485	-0.47878	-0.563134	0.1196			
7775	Mohammad Rehi	-32.8773	15.3873	5.036459	-1.8304	0.611781	-5.42027	0.67818	-0.8875	1.971273	-1.6472	1.298131	-1.2578	1.590914	-0.7678	0.13861	0.27640	-0.5323	0.2737		
7775	Mohammad Rehi	-30.2561	11.1770	3.547472	1.521391	-0.3804	-0.6594	-5.42027	-0.2653	-0.33122	-0.6123	-1.0422	1.142784	-0.37408	0.044083	-0.04747	0.03831	0.1955	-0.2537	0.422	
7775	Mohammad Rehi	-50.3228	7.686802	0.431504	0.415566	-0.51388	0.587571	-3.3122	-0.6866	1.708666	1.0987	1.708666	-1.593814	0.10938	-0.2640	0.129513	0.129513	0.129513			
7775	Mohammad Rehi	-51.2178	1.495257	1.32134	-1.2316	0.430412	-1.6462	-0.1224	-0.3893	1.72408	-1.3188	1.72408	-1.59384	0.248377	-0.2567	0.137788	0.20957	0.642341	-0.0938	-0.0884	-0.1
7775	Mohammad Rehi	-51.0481	2.848810	0.51185	-1.8047	1.833133	2.74502	-3.5722	-1.3969	-1.3963	1.3967	-1.3963	1.39618	-0.0365	-0.0365	-0.0365	0.03133	0.037553	-0.4697	-0.1	
7775	Mohammad Rehi	-48.1611	1.810944	1.047470	-0.4876	0.0317	0.577508	-3.0868	0.6189	-1.59577	0.323887	-1.16771	0.35434	-1.15554	0.045202	-0.045202	0.045202	0.045202	0.045202		
7775	Mohammad Rehi	-48.1611	2.940308	1.685904	-0.45991	0.45646	-0.38512	-1.12075	-0.4571	1.446257	1.446257	1.446257	1.446257	1.446257	-0.03512	-0.03512	-0.03512	0.03512	0.03512		
7775	Mohammad Rehi	-49.1474	1.805462	1.385938	-0.4049	0.29422	1.274681	-1.55539	1.72343	-0.7765	0.45345	-0.7386	0.36051	0.6958	1.37853	-0.7954	-0.4584	-0.2405	-0.1376	0.525	
7775	Mohammad Rehi	-48.1284	1.766519	0.944175	1.8848	1.30423	2.71772	-2.0689	1.2417	1.05151	0.628086	-0.47476	0.030226	0.853124	0.48807	0.170076	-0.48250	0.20318	0.3092	0.6	
7775	Mohammad Rehi	-50.5740	2.397927	-1.7421	-1.0125	1.087786	0.855374	-1.1295	0.40581	0.7352	-0.6823	-0.04565	0.2997	-0.02126	0.36707	0.274086	0.006621	0.42407	-0.01		
7775	Mohammad Rehi	WEKA	NaN	NaN	NaN	NaN	NaN	NaN	NaN												

Fig. 5. CSV labeled matrix for a reader.

The 15- Feature Matrices were combined into one file for the 15 reader's features. For the purpose of training and testing, a WEKA tool was used with 70% to be in training as illustrated in Fig. 6 and 30% for testing as illustrated in Fig. 7.

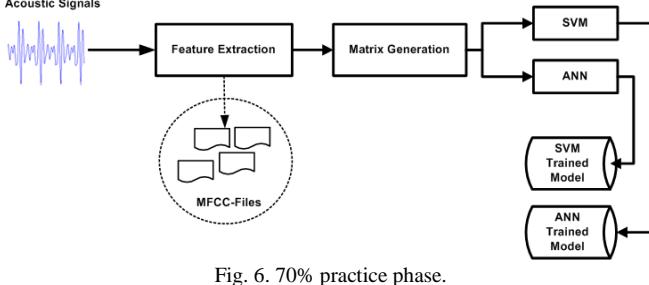


Fig. 6. 70% practice phase.

Moreover, a randomized filter property (from WEKA) was used for rearranging the features with randomly chosen rows. As can be seen from Fig. 6 for each sound wave the features were extracted and all features were combined in one labelled matrix. 70% of this matrix was to be inputted to both SVM and ANN to train the system. The result were two trained models one for SVM and one for ANN.

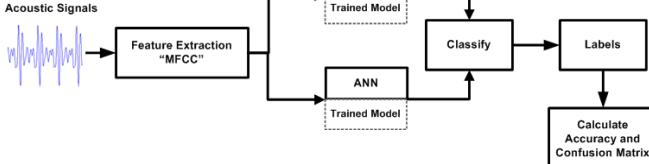


Fig. 7. 30% testing phase.

After obtaining the trained models, the rest of the feature matrix (30%) is used for testing them as seen in Fig. 7. The result of the test gave two confusion matrices one for SVM and one for ANN. The confusion matrix shows both the hit and miss classification of the waves. The correct classification is located on the main diagonal, while others are distributed up or down the main diagonal.

B. Mel-Frequency Cepstrum Coefficients (MFCC)

The Mel-Frequency Cepstrum Coefficients (MFCC) is a

short-period power spectrum of sound wave representation. Mel frequencies are based on known variations in the human ear's critical bandwidth with frequency filters at frequencies below 1 kHz and logarithmically at higher frequencies, to capture the phonetically important characteristics of speech [1]. In the Holy Quran, the correct form of pronunciation and voice depends on the context and is controlled by the reader's voice and his ability move from one pitch and tone to another (Maqam). The stages involved in Mel-frequency Cepstral Coefficient (MFCC) extraction are illustrated in Fig. 8.

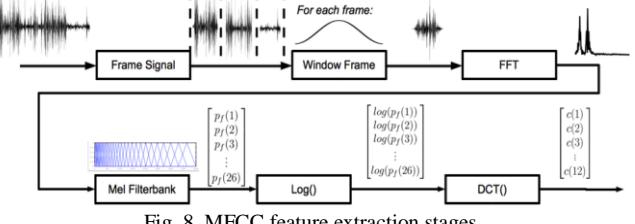


Fig. 8. MFCC feature extraction stages.

The MFCC is calculated using equation 1 and its implementation in Algorithm 1 [22].

$$C_i = \sum_{k=1}^N X_k \cos\left(\frac{[\pi_i(k-0.5)]}{N}\right), \text{ for } i=1,2,\dots,p \quad (1)$$

where C_i is the cepstral coefficients, P is the order, k is the number of discrete Fourier transforms magnitude coefficients, X_k is the k th order log-energy output from the filter bank, N is the number of filters (usually 20). Thus, 19 coefficients and an energy feature were extracted, generating a vector of 20 coefficients per frame. In this research, the first 20-order of the MFCC were extracted.

Algorithm 1: MFCC Method

Inputs: s : the signal, fs : the sampling rate of the signal.

Output: r = transformed signal.

$m = 100$

%Frames blocking phase

%m: the distance between the

begging of %two frames

%n: frame length

%nb Frame: number of frames

for $i = 1$: nbFrame
for $j = 1$: nbFrame
 $M(i, j) = s(((j - 1) * m) + i);$
end
end

$h = \text{hamming}(n);$
 $M2 = \text{diag}(h) * M;$

for $i = 1$: nbFrame
frame(:, i) = $\text{fft}(M2(:, i));$

end

$m = \text{melbankm}(20, n, fs);$

$n2 = 1 + \text{floor}(n / 2);$

$z = m * \text{abs}(\text{frame}(1:n2, :))^{1/2};$

$r = \text{dct}(\log(z));$

%Windowing phase: windowing all frames via %multiply each individual frame by %windowing function.

%FFT Phase: to convert each frame from %time domain into frequency domain to %removes the redundancy of mathematical calculations.

%mel-spaced filter bank

%Length of mel-spaced FFT

%Take log and then the DCT conversion

C. Support Vector Machine

A support vector machines (SVM) is a powerful and popular discriminative classifier. SVMs aim at maximizing the separation margin between two competing classes, where the margin is defined as the distance between the decision

hyper plane and the closest training examples. In other words, finding the line that separates the sample into two classes with the maximum distance between the line and the samples. This line is called the hyper plane: the samples near the hyper plane are called support vectors. The maximum distance between the support vectors and the hyper plane is called the margin [2], [23]. Fig. 9 Is a simple pictorial view of SVM stages.

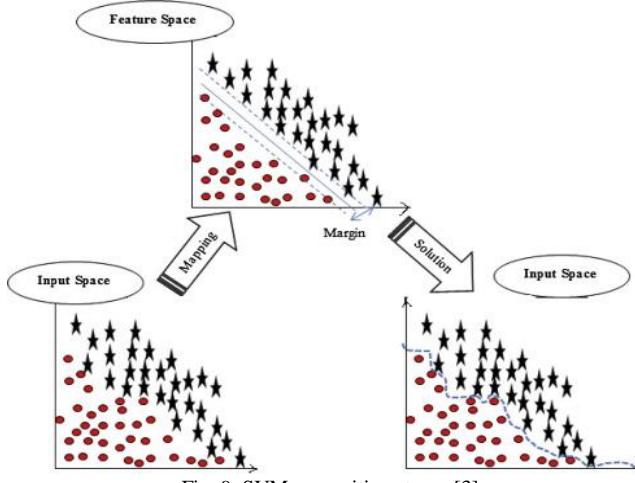


Fig. 9. SVM recognition stages [3].

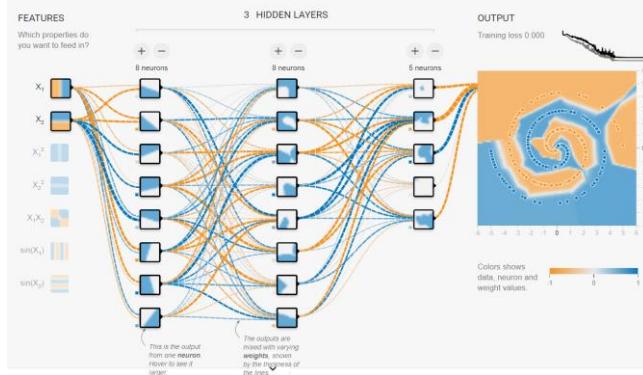


Fig. 10. ANN Paradigm¹.

D. Artificial Neural Networks

Neural networks, a beautiful biologically-inspired programming paradigm which enables a computer to learn from observational data. It's a technique for building a computer program that learns from data. It is based very loosely on how we think the human brain works. First, a collection of software "neurons" is created and connected together, allowing them to send messages to each other. Next, the network is asked to solve a problem, which it attempts to do over and over, each time strengthening the connections that lead to success and diminishing those that lead to failure. Fig. 10 represent a simple ANN with one input layer, three hidden layers, and one output layer.

IV. EXPERIMENTS AND EVALUATION

The Holy Quran Reader Identification system depends mainly on the features extracted from the acoustic waves, as

¹ The picture is taken from: <https://becominghuman.ai/creating-your-own-neural-network-using-tensorflow-fa8ca7cc4d0e>

well as the learning and classification algorithms. The SVM is trained with the 15 readers of the Holy Quran in Table II. For comparison purposes, the same experiment was repeated with the ANN for classification. For training purposes, we have chosen 70% of the features for training in both SVM and ANN. The remaining 30% of the features is for the testing, as shown in Fig. 11.

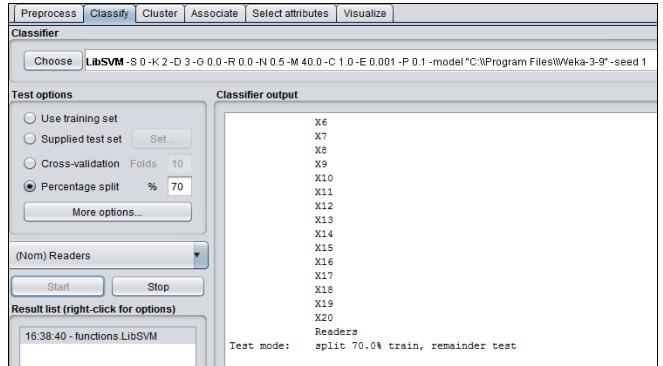


Fig. 11. Model building in WEKA.

As shown in Fig. 11, X1, X2 up to X20 are the feature vector values or the predictors. The "Readers" are the target classes. After training the SVM over 70% of the features, it starts testing the trained model on 30%. The result is a testing confusion matrix of the SVM. The confusion matrix is a data structure used to show the classification results. The rows of this matrix are the desired classification and the columns are the predicted classification.

Table III represents the confusion matrix of the testing phase with SVM. It is clear that some samples are below the hyper plane of the SVM (Main Diagonal) where others are above it. In both cases, it is considered a misclassification of the sample. The sample is correctly classified if it is on the main diagonal. In Table III, most of the samples are correctly classified. To calculate the accuracy, Eq (2) is used.

$$\text{Accuracy} = \frac{\sum \text{Diagonal Sample of confusion matrix}}{\text{Total Sample}} \quad (2)$$

The overall accuracy of reader identification when using SVM is 96.59%. About 4% of the total samples were misclassified. Although we generate samples from only one Surah, it was huge feature file to process. If the samples from Surah Al-Baqarah are entered as an outlet test data, the SVM shows 95.9% accuracy when tested by the outlet data from Al-Baqarah. The Differences between testing on test data from the same corpus (Surah Al-Kahaf) and the outlet data (Surah Al-Baqarah) is about 0.6%, which is highly acceptable.

The misclassified samples were due to the way that the Holy Quran reader recited the different verses. The similarity between the reader's sound and the emotion expressed is critical. Especially since most of the Holy Quran readers follow the same reading rules "Ahkam Altajweed" when they are reciting the various verses of the chapter. Applying "Ahkam Altajweed" will increase the chance of the inter modulation between acoustic waves from different readers. Moreover, every reader has his own standard of delivery "Makam" (مقامات صوتية) which is often like that of other readers.

TABLE III: TESTING CONFUSION MATRIX OF SVM

10	4	11	2	27	3	1	111	22	10	1	5	7	49	11181	<i>o = Fares Abbad</i>
6	22	48	4	42	51	0	148	103	1	42	39	5	11136	60	<i>n = Yasser Aldosari</i>
6	4	0	6	14	29	0	19	8	0	180	38	12740	20	1	<i>m = Mohammad Rashad</i>
0	6	14	4	49	56	1	55	11	0	81	9136	24	51	10	<i>l = Salah Albudair</i>
4	8	42	9	16	25	0	21	9	1	8726	75	193	41	0	<i>k = Saud Shuraim</i>
6	7	6	0	1	4	15	3	3	6080	1	1	0	1	8	<i>j = Saad Alghamidi</i>
30	10	6	8	15	12	1	138	10997	1	11	8	10	82	25	<i>i = Abdullah Basfar</i>
12	19	35	36	28	27	0	12182	134	5	21	64	19	125	108	<i>h = Mishary Alafasi</i>
3	2	1	0	0	1	4648	0	1	23	0	1	9	0	0	<i>g = Maher Almuaiqly</i>
13	47	78	8	71	10195	0	51	15	4	46	112	55	62	15	<i>f = Alsudaes</i>
2	116	77	9	10237	31	2	20	12	0	10	11	2	25	41	<i>e = Khaled Alqahtani</i>
2	1	92	10307	5	12	0	50	3	0	10	5	2	7	0	<i>d = Yahiya Hawwa</i>
17	29	10476	123	73	49	1	48	11	1	17	7	1	45	33	<i>c = Ahmad Alhuthaifi</i>
7	10624	47	4	127	54	2	25	13	7	11	6	9	20	7	<i>b = Ahmad Alajamy</i>
6225	2	21	4	6	10	0	23	41	2	4	2	0	4	20	<i>a = Ali Alhuthaifi</i>
<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	<i>m</i>	<i>n</i>	<i>o</i>	

For comparison purposes, the same experiments were repeated with another learning algorithm, namely the ANN. The same files were used as an input to ANN with the same ratios of 70% for training and 30% for testing. The result shows that ANN converges and stops its progress when the accuracy reaches 86.17%. The ANN confusion matrix is shown in. When feeding Surhat “Al-Baqara” as an outlet data for the ANN classifier, the percentage of accuracy fell to 83%. The SVM shows superiority over ANN in this experiment. Table IV shows the summary of the results of the proposed system using SVM to contrast with ANN

classifiers.

The SVM classifier shows a better result than ANN with our system. It has higher accuracy and lower MSE than ANN.

TABLE IV: SVM AND ANN RESULTS SUMMARY

Accuracy / Learning Algorithm	SVM	ANN
Accuracy-Normal Corpus	96.59%	86.17%
Accuracy-Outlet Data	95.9%	83%
Average Accuracy	96.245	84.585
Average MSE	0.0036	0.0144

TABLE V: THE TEST CONFUSION MATRIX FOR ANN

75	20	77	5	112	4	6	252	81	49	0	25	15	100	10623	<i>o = Fares Abbad</i>
48	275	231	29	229	252	3	373	774	0	171	109	58	8661	494	<i>n = Yasser Aldosari</i>
46	27	9	34	217	158	9	114	102	9	434	302	11514	73	17	<i>m = Mohammad Rashad</i>
28	73	65	42	231	387	15	175	96	5	266	7696	150	151	118	<i>l = Salah Albudair</i>
73	82	150	115	154	181	7	70	89	6	7653	141	274	174	1	<i>k = Saud Shuraim</i>
37	68	6	5	8	9	152	27	47	5654	0	15	10	2	96	<i>j = Saad Alghamidi</i>
264	82	13	12	76	9	4	160	10406	22	7	38	31	185	45	<i>i = Abdullah Basfar</i>
218	127	101	95	144	72	1	10278	598	15	53	279	33	307	494	<i>h = Mishary Alafasi</i>
22	16	0	20	32	2	4242	14	5	287	4	27	17	0	1	<i>g = Maher Almuaiqly</i>
44	262	133	27	256	9102	13	51	152	10	96	241	153	140	93	<i>f = Alsudaes</i>
21	660	231	30	9208	84	5	53	62	1	49	28	9	26	128	<i>e = Khaled Alqahtani</i>
18	3	262	9908	79	9	3	97	27	2	42	9	6	23	8	<i>d = Yahiya Hawwa</i>
129	206	8833	967	244	165	25	44	63	11	13	32	12	73	114	<i>c = Ahmad Alhuthaifi</i>
33	9734	146	11	574	117	18	33	63	55	63	36	38	30	12	<i>b = Ahmad Alajamy</i>
5744	47	92	44	15	5	21	35	223	22	10	8	24	18	56	<i>a = Ali Alhuthaifi</i>
<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	<i>m</i>	<i>n</i>	<i>o</i>	

V. CONCLUSION

In this research, we developed a Holy Quran Reader Identification System from acoustic waves that represent the pronunciation of the “Ayaat”. The MFCC features were extracted from the signals. The SVM and ANN classifiers were used, and their results were compared. The results obtained reveal that the identification accuracy using SVM is much better compared to the ANN. The SVM gives 96.59% accuracy while, 86.17% accuracy was obtained with ANN. The proposed system is the very first system for Holy Quran reader identification. In the future, we suggest extending our work with more than 15 Holy Quran readers and using further extraction techniques and classification algorithms such as the Gaussian Mixture, to increase the identification accuracy.

ACKNOWLEDGEMENT

We extend our sincere thanks to Yarmouk University - Irbid - Jordan, where this scientific research was supported by the Deanship of Scientific Research and Graduate Studies at the University.

REFERENCES

- [1] M. A. Islam, W. A. Jassim, N. S. Cheok, and M. S. A. Zilany, “A robust speaker identification system using the responses from a model of the auditory periphery,” *PLoS One*, vol. 11, no. 7, p. e0158520, Jul. 2016.
- [2] T. K. Das and K. M. O. Nahar, “A voice identification system using hidden markov model,” *Indian J. Sci. Technol.*, vol. 9, January 2016.
- [3] S. B. Dhone, A. Chaudhari, and S. M. Jagade, “Integration of mel-frequency cepstral coefficients with log energy and temporal derivatives for text-independent speaker identification,” pp. 791–797, 2017.
- [4] L. Bao and X. Shen, “Improved Gaussian mixture model and application in speaker recognition,” in *Proc. 2016 2nd International Conference on Control, Automation and Robotics (ICCAR)*, 2016, pp. 387–390.
- [5] F. Meriem, H. Farid, B. Messaoud, and A. Abderrahmene, “New front end based on multitaper and gammatone filters for robust speaker verification,” 2017, pp. 344–354.
- [6] D. A. Reynolds, “Speaker identification and verification using Gaussian mixture speaker models,” *Speech Commun.*, vol. 17, no. 1–2, pp. 91–108, Aug. 1995.
- [7] R. Hasan, M. Jamil, G. Rabbani, and S. Rahman, “Speaker Identification Using Mel Frequency Cepstral Coefficients,” in *Proc. 3rd Int. Conf. Electr. Comput. Eng.*, pp. 28–30, December 2004.
- [8] M. Jeevan, A. Dhingra, M. Hanmandlu, and B. K. Panigrahi, “Robust speaker verification using GFCC based i-vectors,” 2017, pp. 85–91.

- [9] R. M. Fauzi and W. Maharani, "The recognition of hijaiyah letter pronunciation using mel frequency cepstral coefficients and hidden markov model," *Adv. Sci. Lett.*, vol. 22, no. 8, pp. 2043–2046, Aug. 2016.
- [10] S. Srivastava, M. Chandra, and G. Sahoo, *Phase Based Mel Frequency Cepstral Coefficients for Speaker Identification*, pp. 309–316, 2016.
- [11] W. Fei, X. Ye, Z. Y. Sun, Y. J. Huang, X. Zhang, and S. X. Shang, "Research on speech emotion recognition based on deep auto-encoder," in *Proc. 2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2016, pp. 308–312.
- [12] R. Saeidi, P. Alku, and T. Backstrom, "Feature extraction using power-law adjusted linear prediction with application to speaker recognition under severe vocal effort mismatch," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 1, pp. 42–53, Jan. 2016.
- [13] L. Muda, M. Begam, and I. Elamvazuthi, "Voice recognition algorithms using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques," *J. Comput.*, vol. 2, no. 3, pp. 2151–9617, 2010.
- [14] T.-S. Chi, T.-H. Lin, and C.-C. Hsu, "Spectro-temporal modulation energy based mask for robust speaker identification," *J. Acoust. Soc. Am.*, vol. 131, no. 5, p. EL368-EL374, May 2012.
- [15] S. Nakagawa, L. Wang, and S. Ohtsuka, "Speaker identification and verification by combining MFCC and phase information," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 20, no. 4, pp. 1085–1095, May 2012.
- [16] Y. Shao, Z. Jin, D. Wang, and S. Srinivasan, "An auditory-based feature for robust speech recognition," in *Proc. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 4625–4628.
- [17] R. Bharti and P. Bansal, "Real time speaker recognition system using MFCC and vector quantization technique," *Int. J. Comput. Appl.*, vol. 117, no. 1, pp. 25–31, 2015.
- [18] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.
- [19] G. Nijhawan, "Speaker recognition using support vector machine," *Int. J. Comput. Appl.*, vol. 87, no. 2, pp. 7–10, 2014.
- [20] A. Mahmood, M. Alsulaiman, and G. Muhammad, "Automatic speaker recognition using multi-directional local features (MDLF)," *Arab. J. Sci. Eng.*, vol. 39, no. 5, pp. 3799–3811, 2014.
- [21] K. M. O. Nahar, M. Elshafei, W. G. Al-khatib, H. Al-muhtaseb, and M. M. Alghamdi, "Statistical analysis of arabic phonemes for continuous arabic speech recognition," *Int. J. Comput. Inf. Technol.*, vol. 1, no. 2, pp. 49–61, 2012.
- [22] S. A. Majeed, H. Husain, S. A. Samad, and T. F. Idbeaa, "Mel frequency cepstral coefficients (MFCC) feature extraction enhancement in the application of speech recognition: A comparison study," *J. Theor. Appl. Inf. Technol.*, vol. 79, no. 1, pp. 38–56, 2015.
- [23] W. S. Noble, "What is a support vector machine?" *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.



Khalid M. O. Nahar received his BS, and MS degrees in computer sciences from Yarmouk University in 1992 and 2005 respectively. He was awarded a full scholarship to continue his PhD in computer science and engineering from King Fahd University of Petroleum and Minerals (KFUPM), KSA. In 2013 he completed his PhD and started his job as an assistant Prof. at Tabuk University, from 2013 to 2015. In 2015, he back to work at Yarmouk University, Jordan, as an assistant professor in the

Department of Computer Science. For now, he is the dean assistant for quality control. His research interests include continuous speech recognition, Arabic computing, natural language processing, Wireless Sensor Networks (WSNs), multimedia computing, content-based retrieval, artificial intelligence, and software engineering.



Moy'awiah Al-Shannaq is an assistant professor of computer sciences in the Faculty of Information Technology and Computer Science, Yarmouk University. Before joining Yarmouk University, Dr. Al-Shannaq has been working as a lecturer in the Department of Computer Sciences at Kent State University, Ohio, USA for two years. He received his MSc and BSc in computer sciences from Yarmouk University, Jordan. He received his PhD in computer sciences from Kent State University, Ohio, USA. His research interests include natural language processing, algorithmic graph and hypergraph theory, computational geometry, and network algorithms.



Ahmed M. Manasrah is an associate professor at Yarmouk University, Jordan since 2011, prior to joining Yarmouk University, Dr. Manasrah was the deputy director of research and innovation at the National Advanced IPv6 Center, Malaysia for 2 years. Dr. Manasrah has published over 40 international journal and 15 research papers including proceedings and 4 chapters of book, and 3 international patents in the field of network monitoring and security. Dr. Manasrah received his PhD from Universiti Sains Malaysia(USM) in 2009. His research interest includes advanced Internet security and monitoring.



Rafat M. Alshorman is an assistant professor in the Department of Computer Science at Yarmouk University, Jordan. He completed his Ph.D. at Loughborough University, UK in 2009 and his undergraduate studies at Yarmouk University, Jordan. His research interests lie in the area of algorithms and mathematical models, ranging from theory to implementation, with a focus on checking the correctness conditions of concurrent and reactive systems. In recent years, he has focused on theoretical computer science such as graph theory and numerical analysis.



Iyad M. Alazzam is an associate professor in the Department of Computer Information Systems at Yarmouk University in Jordan, he has received his Ph.D. degree in software engineering from NDSU (USA), He got his master degree in electronic Commerce from LMU (UK) and his BSc in computer science and information systems from Jordan University of Science and Technology. His research interests lay in software engineering and software testing and information retrieval.