ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 6, 173-176 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Quranic Reciter Recognition: A Machine Learning Approach

Rehan Ullah Khan[*,1,3], Ali Mustafa Qamar[2], Mohammed Hadwan[1,3]

[1]*Department of Information Technology, College of Computer, Qassim University, Saudi Arabia*

[2]*Department of Computer Science, College of Computer, Qassim University, Saudi Arabia*

[3]*Intelligent Analytics Group (IAG), College of Computer, Qassim University, Saudi Arabia*

ARTICLE INFO

Article history:
Received: 09 October, 2019
Accepted: 16 November, 2019
Online: 25 November, 2019

Keywords:
Audio analysis
Quran analysis
MFCC
Machine learning

ABSTRACT

Recitation and listening of the Holy Quran with Tajweed is an essential activity as a Muslim and is a part of the faith. In this article, we use a machine learning approach for the Quran Reciter recognition. We use the database of Twelve Qari who recites the last Ten Surah of Quran. The twelve Qari thus represents the 12-class problem. Two approaches are used for audio representation, firstly, the audio is analyzed in the frequency domain, and secondly, the audio is treated as images through Spectrogram. The Mel Frequency Cepstral Coefficients (MFCC) and Pitch are used as the features for model learning in the first case. In the second case of audio as images, Auto-correlograms are used to extract features. In both cases, the features are learned with the classical machine learning which includes the Naïve Bayes, J48, and the Random Forest. These classifiers are selected due to their over-all good performance in the state-of-the-art. It is observed that classifiers can efficiently learn the separation between classes, when the audio is represented by the MFCC, and the Pitch features. In such a case, we get 88% recognition accuracy with the Naïve Bayes and the Random Forest showing that Qari can be effectively recognized from the recitation of the Quranic verses.

1. Introduction

Quranic audio analytics lacks thorough research and understanding from machine learning perspectives. Out of many Qari tilawat recitations available offline and online, an automated system could help in the selection of specific voice of Qari, depending upon the person's mood and choice. This drives the motivation for our work. Altalmas et al. [1] processed the Spectrogram features for the Qalqalah letters, describing the process of Qalqalah correctly. However, in [1], there is no recognition part. In [1], the authors declare classification and recognition as future work. The work of [2] develops an autonomous delimiter that performs the extraction of Quranic verses from the tilawat by using the Sphinx framework. [2] only uses Surah "Al-Ikhlass" for analysis. There are many limitations in [2], not only because of the sample but also due to the usage of the Hidden Markov Model (HMM). The [3] analyzes the recitation of many Surah of the Quran. It is found that there is 21.39% of voiced speech in Quranic recitations, which is 3 times higher than audiobooks. Therefore, a linear predictor component can be used for efficiently representing the Quranic signals. The authors in [4] propose an online speech recognition technique for verification of the Quranic verses. According to Kamarudin et al. [5], the rules of the Quran verse are prone to additive noise. Therefore, they can affect the classification of Quranic results. The authors propose an Affine Projection approach as the optimized solution for echo cancellation. Elobaid et al. [6] develop "Noor Al-Quran" for handheld devices for Non-Arabic speakers for correct recitation learning. Khurram and Alginahi [7] discuss the concerns and the challenges of digitizing and making the Quran available to the masses. The authors in [8] focus on user acceptance of the speech recognition capabilities of mobile devices. Another article [9] represents blind and disabled people to use education-related services for Quran. The [10] demonstrates the Computer-Aided Pronunciation Learning module (CAPL) to detect Quranic recitation errors.

In this article, we analyze the recitation of the Twelve Qari, reciting the last ten Surah of the Quran, thus representing a 12-class problem. For this setup, the audio is analyzed in the frequency domain, and as images through Spectrogram. The Mel

[*]Rehan Ullah Khan, re.khan@qu.edu.sa

https://dx.doi.org/10.25046/aj040621

Frequency Cepstral Coefficients (MFCC) and Pitch are used as the features for model learning. In the scenario of audio as images, the Auto-correlograms are used to extract features. The features are learned with the Naïve Bayes, J48, and the Random Forest, being selected due to their over-all excellent performance in the state-of-the-art. The experimental analysis shows that the classifiers can efficiently detect the reciter of the Quran if the audio is represented by the MFCC and Pitch features. In such a case, we get 88% recognition accuracy with the Naïve Bayes and the Random Forest showing that the Qari class can be effectively recognized from the recitation of the Quranic verses.

2. Recognition Models and Features

In this section, we discuss the classifiers and feature extraction, which are used in experimentation and analysis.

2.1. Naïve Bayes

Naïve Bayes classifiers are a family of probability-based classifiers with the use of strong (naïve) assumptions about the independence in Bayes' theorems [11]. Naïve Bayes assign class labels to classes of a certain problem, where feature label consists of a specific set of class labels. Not only the algorithm for designing such classifiers but the family of algorithms is based on the general principle: all naïve Bayes classifiers assume that the value of a particular attribute does not depend on the value of any other attribute of the data in question.

2.2. J48

J48 is the implementation of the Open Source Quinlan C4.5 decision tree algorithm [12]. Decision tree algorithms start with a series of questions and examples and create tree data structures that can be used to classify new tasks. Each case is described by the attributes (or properties). Each training case has a class label associated with it. Each node within the decision tree is included in the test, which results in which branch to choose from.

2.3. Random Forest

Recently, Decision tree classifiers have gained considerable popularity. This popularity is due to the intuitive nature and overall easy learning paradigm. The classification trees, however, suffer from low classification accuracy and generalization. The accuracy of classification and generalization cannot be increased simultaneously. For this purpose, Breiman [13] introduced Random Forest. It uses a combination of several trees from one data set. A random forest creates a forest of trees, so each tree is generated based on a random grain plus data. For classification stages, the input vector is applied to every tree in the forest. Each tree decides about the class of the vector. These decisions are then summed up for the final classification.

2.4. Mel Frequency Cepstral Coefficients (MFCC)

The first step in any automated speech processing is to extract the functions, that is, the properties of the phonetic features that are effective at identifying words, and all other components containing information such as background sounds, thoughts, etc. [14]. The sound emitted by humans is filtered through the structure of the tongue, teeth, and so on. This structure determines which sounds come out. If we can learn exactly what that looks like, it should give us a true representation of the phoneme produced. The vocal structure is presented in the envelope of the short-time spectrum of power, and the function of the MFCCs is to represent this envelope accurately. MFCCs are widely used features in automatic speech recognition.

2.5. Pitch and Frequency

The pitch is an audio sensation for which the subject assigns music tones to the relative position on the music-based scale on the perception of the vibrational frequency [15]. The pitch is strongly related to the frequency but they are not similar. Frequency is an objective and scientific quality that can be measured. Pitch has a personal perception of the sound wave for each person, which cannot be measured directly. However, it does not mean that most people will not agree on which audio/music notes are lower and higher. Pitch can be quantified as frequencies in Hertz or cycle per second by a comparative analysis of the subjective sound signals with the ones with standard pure tones having aperiodic, sinusoidal wave structure. This approach is mostly used to assign a pitch value to the complex and aperiodic sound signals.

3. Experimental Evaluation

In this section, we discuss the dataset and the experimental evaluation performed for different parameters.

The dataset consists of 120 Quranic recitations performed by the 12 Reciters. The dataset is downloaded and collected from [16]. The 10 Surahs recited by 12 Qaris are as follows:

- Al-Fil
- Quraysh
- Al-Ma`un
- Al-Kawthar
- Al-Kafirun
- An-Nasr
- Al-Masad
- Al-'Ikhlas
- Al-Falaq
- An-Nas

For experimental evaluation, we select Naïve Bayes, J48, and the Random Forest. These are selected based on their good overall performance in the state-of-the-art audio classification based on MFCC and Pitch.

Two types of features are extracted. In the first scenario, the features are extracted based on the MFCC and Pitch of the recitation audio. The second is based on the image domain. The Tilawat (recitation) is first converted to image representation. For the conversion, we use the spectrogram approach.

The spectrogram approach works on the principle of Fourier transform. The key elements are then represented as frequency response coefficients. These are then combined as a time-domain structure. This time-domain structure is represented as an image signal. After an image is obtained, image-based feature extraction can be applied. We adopt the Auto-Correlogram approach [17] for feature extraction. The Auto-Correlogram approach takes into consideration not only the pixel values but also the distance of a particular color from the next similar color. These features have shown excellent performance in state-of-the-art [17].

3.1. Performance analysis based on audio features

Two methods are used to extract features from the audio. In the first case, the features are extracted based on the MFCC and Pitch of the recitation audio. The details of MFCC and Pitch are given in the previous section.

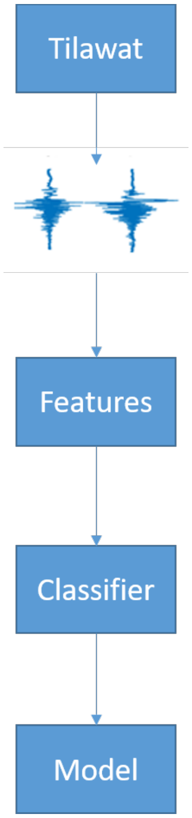

Figure 1. The classification model for Qari recognition

Figure 1 shows the flow of the recognition model. The recitation is converted to the MFCC and Pitch features. The features are then learned by the classifier. The output of classifier learning is generally called the model. The model is then used to test other recitation audios. We use the 10-folds cross-validation scheme. In this scheme, 90% of data is used for training the model, and 10% is used for testing. This procedure is repeated 10 times.

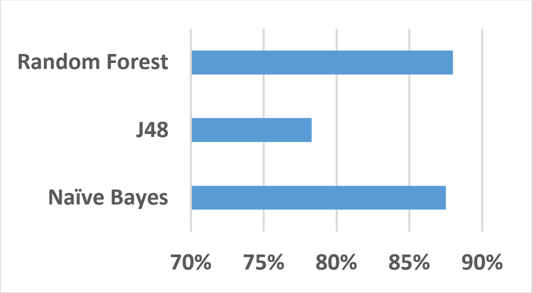

Figure 2. Accuracy of MFCC and Pitch with the three classifiers

Table 1: Performance analysis of the MFCC and Pitch features

Classifier	Accuracy
Naïve Bayes	88%
J48	78%
Random Forest	88%

Figure 2 shows the performance of the classification paradigm for the three classifiers, namely Random forest, J48, and the Naïve Bayes using the features of the MFCC and the Pitch. Table 1 shows similar performance in the form of accuracy values.

Figure 2 shows the accuracy of the models. We select the accuracy parameter because the data is almost balanced with reference to classes. In Figure 2 and Table 1, it can be noted that Random Forest and the Naïve Bayes have the highest accuracy. We represent the accuracy as the percentage. This means that the Random forest accuracy of 88% is linked to the recognition of the Qari. As such, Random Forest can recognize the Qari with 88% accuracy. The total error the Random forest will make in identifying the Qari will be only 12%.

Similarly, the accuracy of Naïve Bayes is also 88%. This is coinciding with the Random Forest. Therefore, the Naïve Bayes classifiers learn the Qari with a good recognition model like that of the Random Forest. In Figure 2, and Table 1, the lower performance is exhibited by the J48. The performance of the J48 is 78%. This means that the J48 classifier model has a 22% chance of not recognizing the Qari correctly. This is higher compared to the Random forest and the Naïve Bayes, which is only 12%. This low performance could be due to the sensitivity of the decision trees (J48) to the noise in the data.

3.2. Performance analysis based on image features

For the conversion of the audio recitation to the image representation, the Fourier transform approach is used. After Fourier transformation, the frequency response coefficients are combined as a time-domain structure. This time-domain structure is represented as an image signal. The image-based feature extraction of the Auto-Correlogram approach [17] for feature extraction is used. These features have shown very good performance in state-of-the-art [17].

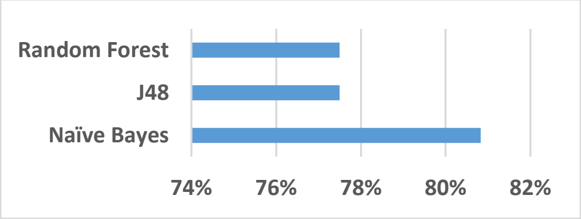

Figure 3. Accuracy of the spectrogram-based recognition models

Table 2: Performance analysis of the Spectrogram features

Classifier	Accuracy
Naïve Bayes	81%
J48	78%
Random Forest	78%

Figure 3 and Table 2 show the performance analysis of the image-based representation of the Quran recitation audio.

In Figure 3 and Table 2, the Naïve Bayesian algorithm has the highest accuracy. The performance parameter is Accuracy, which is used for balanced classes. Similar to Figure 2, the accuracy is represented as the percentage. The Naïve Bayes classifier's accuracy of 81% means that Naïve Bayes is capable of recognizing the Qari with an 81% accurate model. The chance of the error being made by the Naïve Bayes in identifying the Qari is 19%. This error can be explained as such that out of 100 recitations performed by different Qari, only 81 recitations are correctly identified and mapped to the corresponding Qari of the recitation. The accuracy of the Random Forest is 78%. This means that 22 samples out of 100 samples of recitations will be wrongly identified and mapped to a different Qari. Interestingly, the accuracy of J48 is also 78%. The general trend, however, is that Random Forest normally has a higher classification accuracy than J48 in state-of-the-art.

The same detection performance for both the Random Forest and the J48 is an exciting result. Since both of the algorithms work on the same principle of decision trees, similar results in special cases make sense. Moreover, the Naïve Bayes classifier learns the Qari with a good recognition model like that of Figure 2. However, the Naïve Bayes performance in Figure 2 and Figure 3 is not consistent, though higher than other models, especially, in Figure 3. As such, the Naïve Bayes' higher performance in both the cases of audio features and image features is motivating for further analysis and practical applications.

4. Conclusion

By using 120 total recitations, we analyzed the recitations of 12 Qari. We used two approaches to process the audio recitations. The first one being the MFCC and Pitch, and the second one as the Spectrogram-based images. Auto-correlograms are used to extract features in case of image representation. The features are learned with the Naïve Bayes, J48, and the Random Forest, being selected due to their over-all good performance in the state-of-the-art. The experimental analysis shows that the classifiers can efficiently learn the reciter of the Quran if the MFCC and the Pitch features represent the audio. In such a case, we get 88% recognition accuracy with the Naïve Bayes and the Random Forest showing that the Qari class can be effectively recognized from the recitation of the Quranic verses.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The work in this article is funded in its entirety by the Deanship of Scientific Research (SRD), Project number: **3600-coc-2018-1-14-S** at the Qassim University, Kingdom of Saudi Arabia.

We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU which is used for relevant research.

References

[1] T. Altalmas, S. Ahmad, W. Sediono, and S. S. Hassan, "Quranic letter pronunciation analysis based on spectrogram technique: Case study on qalqalah letters," in CEUR Workshop Proceedings, vol. 1539, pp. 14–22, 2015.

[2] H. Tabbal, W. El Falou, and B. Monla, "Analysis and implementation of a Quranic verses delimitation system in audio files using speech recognition techniques," in 2nd International Conference on Information and Communication Technologies: From Theory to Applications, ICTTA, Damascus, Syria, vol. 2, pp. 2979–2984, 2006. https://doi.org/10.1109/ICTTA.2006.1684889

[3] T. S. Gunawan and M. Kartiwi, "On the characteristics of various Quranic recitation for lossless audio coding application," in 6th International Conference on Computer and Communication Engineering: Innovative Technologies to Serve Humanity, ICCCE, Kuala Lumpur, Malaysia, pp. 121–125, 2016. https://doi.org/10.1109/ICCCE.2016.37

[4] A. Mohammed and M. S. Sunar, "Verification of Quranic Verses in Audio Files using Speech Recognition Techniques," in Int. Conf. Recent Trends Inf. Commun. Technol., 2014.

[5] N. Kamarudin, S. A. R. Al-Haddad, M. A. M. Abushariah, S. J. Hashim, and A. R. Bin Hassan, "Acoustic echo cancellation using adaptive filtering algorithms for Quranic accents (Qiraat) identification," Int. J. Speech Technol., 19(2), pp. 393–405, 2016. https://doi.org/10.1007/s10772-015-9319-z

[6] M. Elobaid, K. Hameed, and M. E. Y. Eldow, "Toward designing and modeling of Quran learning applications for android devices," Life Sci. J., 11(1), pp. 160–171, 2014.

[7] M. K. Khan and Y. M. Alginahi, "The holy Quran digitization: challenges and concerns," Life Sci. J., 10(2), pp. 156–164, 2013.

[8] N. Kamarudin, S. A. R. Al-Haddad, A. R. B. Hassan, and M. A. M. Abushariah, "Al-Quran learning using mobile speech recognition: An overview," in International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia 2014. https://doi.org/10.1109/ICCOINS.2014.6868401

[9] S. A. E. Mohamed, A. S. Hassanin, and M. T. B. Othman, "Educational system for the holy Quran and its sciences for blind and handicapped people based on Google speech API," JSEA, 7(3), pp. 150–161, 2014. http://dx.doi.org/10.4236/jsea.2014.73017

[10] S. M. Abdou and M. Rashwan, "A Computer Aided Pronunciation Learning system for teaching the holy quran Recitation rules," in IEEE/ACS International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, pp. 543–550, 2014. http://dx.doi.org/10.1109/AICCSA.2014.7073246

[11] D. Zhang, "Bayesian Classification," 2019, pp. 161–178.

[12] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," Mach. Learn., 16(3), pp. 235–240, Sep. 1994. https://doi.org/10.1007/BF00993309

[13] L. Breiman, "Random Forests," Mach. Learn., 45(1), pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[14] X. Huang, A. Acero, and H.-W. Hon, Spoken Language Processing: A Guide to Theory, Algorithm & System Development, 2001. 2001.

[15] P. A. Cariani and B. Delgutte, "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J. Neurophysiol., 76(3), 1996. https://doi.org/10.1152/jn.1996.76.3.1698

[16] "A2Youth.com - The Youth's Islamic Resource." [Online]. Available: https://www.a2youth.com/. [Accessed: 06-Oct-2019].

[17] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image Indexing using Color Correlograms," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, USA, USA, pp. 762–768, 1994.