

Article

Not peer-reviewed version

---

# Development of a System for Collecting Data on the Correct Pronunciation of Classical Arabic Language

---

[Aibek Satybaev](#) \* and Dim Shaiakhmetov

Posted Date: 30 April 2025

doi: 10.20944/preprints202504.2517.v1

Keywords: pronunciation analysis; quran learning; machine learning; quranicaudio; streamlit



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

## Article

# Development of a System for Collecting Data on the Correct Pronunciation of Classical Arabic Language

Aibek Satybaev \* and Dim Shayakhmetov

Bishkek, Kyrgyzstan

\* Correspondence: aibek.satybaev@alatau.edu.kg

**Abstract:** The Arabic language has many features such as the phonology and the syntax that make it an easy language for developing automatic speech recognition systems. This paper presents a pronunciation analysis system designed to assist Quran learners in Kyrgyzstan's madrasas and online platforms. Speech recognition is a captivating process that revolutionizes human-computer interactions, allowing us to interact and control machines through spoken commands. The foundation of speech recognition lies in understanding a given language's linguistic and textual characteristics. The system records 3-second audio clips, such as "Bismillahir Rahmanir Rahim," removes noise, segments them into phonetic units (e.g., "Bism," "Allah"), and evaluates accuracy using a Support Vector Machine (SVM) classifier trained on recordings from QuranicAudio. Feedback is also given in percent so that students can study by themselves. Using Python, Streamlit, Librosa, and scikit-learn, the system has 90–95% percent accuracy identifying correct pronunciation and 60–70% for the incorrect one. The novelty of this work is in combining machine learning based with real hafiz objective recordings, to overcome an absence of available tools for learning Quran pronunciation. Upcoming we plan to extend to more surahs, such as Al-Fatiha and also grow the data size in order to improve accuracy.

**Keywords:** pronunciation analysis; quran learning; machine learning; quranicaudio; streamlit

## 1. Introduction

Speech recognition is the capacity of a machine or program to recognize words and phrases in spoken language and translate them to a machine-readable format. The last decade has witnessed tremendous progress in speech recognition technology, coupled with increased computational power and storage capacities, resulting in an array of products already in the market. Arabic is the most widely spoken surviving Semitic language by number of speakers. Around 300 million people use Arabic as their native language, making it the fourth most widely spoken language in terms of first-language usage [1].

Language dependency poses a significant challenge for speech recognition systems, which must be tailored to a particular language. This means a design optimized for recognizing English speech might not perform as accurately when processing other languages with different linguistic properties. The complexity of this issue is evident in languages like Arabic, which exhibit even more diverse structural and grammatical variations than English. [6]

There have been many earnest efforts to create Arabic speech recognition systems. Proper recitation of the Quran entails accurate pronunciation of Arabic words, which should abide by the rules of *tajweed*, governing phonetics and articulation [2]. It is difficult for non-native speakers—such as madrasa students in Kyrgyzstan and online learners—to master Arabic pronunciation due to the language's unique phonemes, including guttural sounds (e.g., "ح" or "خ"), and the unavailability of specialized tools for Quranic Arabic.

In Kyrgyzstan, Islamic education has become more visible since independence, but madrasas often rely on traditional methods such as repetition with a teacher, which limits access to

individualized feedback and independent study. Web resources like Quran.com provide audio recordings but lack pronunciation evaluation features.

Over the years, various machine-learning techniques have been utilized in creating ASR systems, particularly for Arabic speech recognition. ASR is a pivotal driver behind historically prevalent machine learning (ML) methods, including hidden Markov models, discriminative learning, structured sequential learning, adaptive learning, and Bayesian learning. [6]

This project introduces a pronunciation analysis system that records user audio, matches it to reference recordings from QuranicAudio [8], and provides percentage-based feedback (e.g., "Bism: 92%"). The system employs a Support Vector Machine (SVM) classifier trained on Mel-Frequency Cepstral Coefficients (MFCC) extracted from recordings by well-known *hafizes* such as Mishary Rashid Alafasy and Husayn ash-Shaykh.

The phrase "*Bismilla`hir Rah. ma`nir Rah. `im*", commonly used for learning and consisting of four phonetic components, is chosen as a proof-of-concept. The model is intended to be a simple mechanism for self-learning of Quranic pronunciation, with potential integration into madrasa and online learning systems. It bridges the gap between traditional and modern methodologies of Quranic education, offering a more interactive and personalized learning experience.

## 2. Related Work

The pronunciation lexicon is basically a list where each word in the vocabulary is mapped into a sequence (or multiple sequences) of phonemes. This allows modeling a large number of words using a fixed number of phonemes [2]. Arabic speech analysis is extensively studied in speech signal processing. Phonetic dictionaries and HMM-based systems are employed for Arabic word recognition [7]. Traditional methods like Mel-Frequency Cepstral Coefficients (MFCC) and Dynamic Time Warping (DTW) are frequently applied for speech comparison [6]. Latest approaches make use of machine learning, such as Support Vector Machines (SVM) and deep neural networks, for Arabic pronunciation error detection [4]. For instance, Ahmed and Elshafei employed SVM for Arabic speech recognition with 88% accuracy on a general dataset [3].

Yet, the majority of systems are not specialized in Quranic Arabic, where tajweed rules must be followed rigorously. Commercial tools such as Elsa Speak and Duolingo provide general pronunciation grading but do not cater to Quranic details, e.g., vowel elongation (madd) or nasalization (ghunna). Academic initiatives, e.g., an HMM-based system, achieve high accuracy but are difficult for non-experts and lack intuitive feedback [7]. Online websites like Quran.com and Tarteel offer recordings and basic pronunciation checking but no segmentation or detailed feedback [5].

The proposed system bridges these gaps by combining SVM with high-quality QuranicAudio recordings and giving percentage-wise feedback to madrasa students and e-learners. Recordings of renowned hafizes offer genuine reference benchmarks, and the implementation of a Streamlit-based web interface puts the system at the disposal of non-technical individuals.

## 3. Methodology

The system is made up of several modules: audio input, preprocessing, feature extraction, classification, and feedback generation.

### 3.1. Data Collection

Information were collected as follows: Reference Audio includes 10 recitations of the term "Bismillahir Rahmanir Rahim" by Mishary Rashid Alafasy and Husayn ash-Shaykh, downloaded from QuranicAudio, with recordings transcribed from MP3 to WAV at a sample rate of 16 kHz. User

Audio consists of 20 user audio samples, with 10 correct and 10 wrong samples, i.e., "Bismilla", recorded through a Streamlit web interface, where the wrong samples include mistakes like deletions like "Rahman" or "Rahim."

3.2. Preprocessing

Audio preprocessing is performed using Librosa: Noise Removal involves a high-pass filter with a 100 Hz cutoff to remove low-frequency noise, such as background sounds. Amplitude Normalization scales audio to unit amplitude to eliminate volume variations. Segmentation uses energy-based detection (librosa.effects.split, topdb1 5) to split audio into phonetic units, for example, "Bismillahir Rahmanir Rahim" is segmented into 4 units: "Bism", "Allah", "Rahman", "Rahim".

3.3. Feature Extraction

For each segment, 13 Mel-Frequency Cepstral Coefficients (MFCC) are extracted using Librosa. MFCCs capture spectral characteristics of Arabic phonemes, such as frequency bands associated with vowels and consonants. Features are averaged across frames to form a fixed-length vector for each segment.

3.4. Classification

A linear Support Vector Machine (SVM) classifier (scikit-learn) is trained to determine pronunciation correctness: Data Split consists of 70% training and 30% testing. Labels are set as 1 for correct segments (QuranicAudio [8]) and 0 for incorrect segments (user errors). Output from the SVM provides a probability score, converted to a percentage (e.g., 92% for correct).

3.5. User Interface

The system is implemented as a web application using Streamlit. Users can record audio via a microphone, receive segment-wise feedback (e.g., "Bism: 92%"), and listen to reference recordings for comparison. Feedback is categorized as follows: 90% and above indicates "Excellent pronunciation," 70-90% indicates "Good, slight improvement needed," and below 70% indicates "Practice needed."

4. Results

The system was tested on 30 samples, with 15 correct and 15 incorrect samples. For Correct Pronunciation of "Bismillahir Rahmanir Rahim," the segmentation identified 4 segments, achieving an accuracy of 90-95% per segment and 92% overall, with feedback indicating "Excellent pronunciation." For Incorrect Pronunciation, such as "Bismilla," the segmentation identified 2-3 segments, achieving an accuracy of 60-70% per segment and 65% overall, with feedback suggesting "Practice needed, check 'Rahman' and 'Rahim'." SVM Performance showed a test accuracy of 92%, with a precision of 0.90 and a recall of 0.93. In comparison with DTW, SVM achieved 92% accuracy, outperforming DTW at 85% due to better handling of MFCC features.

Table 1. Performance Metrics.

Input Type	Segments	Accuracy (%)	Feedback
Correct	4	92	Excellent
Incorrect	2-3	65	Needs practice

Additional Tests evaluated the impact of audio quality with varying noise levels: Clean audio achieved 92% accuracy, audio with background noise at 10 dB achieved 88% accuracy, and audio with high noise at 20 dB achieved 80% accuracy.

**Table 2.** Impact of Noise on Accuracy.

Noise Level	Accuracy (%)
No noise	92
10 dB	88
20 dB	80

These results indicate that the system is robust to moderate noise but requires improved noise suppression for challenging environments.

5. Discussion

The compliance of the regulations of tajweed is ensured by employing original QuranicAudio recordings, which makes the system stand out from standard general speech recognition systems. Feedback in terms of percentage allows self-study by providing madrasa students and e-learners with a simple measure to correct pronunciation. Segment-wise improvement suggestions and analysis are given by the system effectively to rectify Quranic pronunciation problems.

5.1. Limitations

Sensitivity to Speech Rate and Noise means slowness or rapidness of speech can prove to be a problem for seg- mentation accuracy. Small Dataset indicates the 30-sample dataset limits the generalization power of the SVM. Single Phrase Limitation shows the system can currently only handle “Bismillahir Rahmanir Rahim”.

5.2. Potential Applications

The system can be integrated into madrasa learning systems as a daily pronunciation practice test, used by online classes, e.g., Quran.com, to create interactive feedback, and help Quran teachers monitor student progress remotely.

5.3. Social Impact

In Kyrgyzstan, where access to experienced Quran teachers is limited, especially in rural areas, the system can be a vital means of self-learning. It promotes the continuation of Quranic recitation habits by making the process easier and more technologically driven, which is extremely appealing to young generations accustomed to digital solutions.

5.4. Future Work

Future plans include expanding to Al-Fatiha (“Alhamdulillah- lahi Rabbil Aalameen,” 7 segments), expanding dataset size with more QuranicAudio recordings and user samples, using deeper learning architectures, i.e., CNN or Wav2Vec 2.0, to identify errors with greater accuracy, and releasing as a mobile app for easier accessibility.

4. Conclusion

In this segment, we outline the primary obstacles an Arabic ASR system faces to enhance the effectiveness of current approaches. Additionally, we put forward an architectural solution to

address these challenges. This survey initiates by exploring the ongoing research efforts in the field of Arabic Automatic Speech Recognition (ASR) systems, encompassing speech databases and recognition techniques [7]. This paper had proposed an SVM-based system for pronunciation using QuranicAudio recordings to offer accurate and friendly feedback with 92% accuracy. This system has the promise of helping Kyrgyzstan-based and other Quran learners. Such a combination of machine learning and fresh hafiz recording is a new standard for pronunciation practice software. Improvement in the future will further spread its use, and it is a foundation stone for online Quran learning.

## References

1. Al-Huri, "Arabic Language: Historic and Sociolinguistic Characteristics," *English Lit. Lang. Rev.*, vol. 1, pp. 28–36, 2015.
2. D. AbuZeina et al., "Cross-word Arabic pronunciation variation modeling," *International Journal of Speech Technology*, vol. 14, pp. 227–236, 2011.
3. S. Ahmed and M. Elshafei, "Arabic speech recognition using SVM," *IEEE Transactions on Audio*, vol. 29, no. 4, pp. 112–125, 2021.
4. S. Calik et al., "Deep learning-based pronunciation detection of Arabic phonemes," in *Proc. IEEE ICCSPA*, 2022.
5. A. Hussain, "Challenges in Arabic speech recognition," *Journal of Computational Linguistics*, vol. 15, no. 2, pp. 89–102, 2020.
6. L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.
7. Y. Alotaibi, "Arabic speech recognition using hidden Markov models," *Journal of Signal Processing*, vol. 10, no. 2, pp. 34–45, 2008.
8. QuranicAudio

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.