African Journal of Advanced Pure and Applied Sciences (AJAPAS)
*Online ISSN: 2957-644X*
Volume 2, Issue 4, October-December 2023, Page No: 369-377
Website: *https://aaasjournals.com/index.php/ajapas/index*
معامل التأثير العربي 2023: (1.55) SJIFactor 2023: 5.689 ISI 2022-2023: 0.557

# Smart Spoken and the Diacritics Arabic Words Recognition System based on Qaidah Noraniah model

Omar Balola Ali [1*], Abdulhafid Bughari [2], Hamed F. Hamed [3], Abdelrahman Ali Salih [4]
[1,2] Computer Engineering, Faculty of Science Engineering, Bright Star University, El-Brega, Libya
[3] Computer Department, Faculty of Science, Ajdabia University, Ajdabia, Libya
[4] Sudan University for Sciences and Technology, Sudan

[*]Corresponding author: omar.balula@bsu.edu.ly

| Received: October 24, 2023 | Accepted: December 16, 2023 | Published: December 24, 2023 |
|---|---|---|

**Abstract:**
This paper presents the Recognition System of spoken Arabic words and the diacritics based on Qaidah Noraniah concept. The Arabic language is one of the most popular languages in the world. There are not many speech recognition systems for the Arabic language. Because of its extensive vocabulary and morphology, it is more difficult to develop systems to recognize it. For some of these reasons, it uses agglutinative letters and diacritics (الحركات). The Qaidah Noraniah language model was designed to solve some of these problems. The model was trained using the CMU Sphinx program, and tested by dependent and independent speakers. The model achieved good results in recognizing single letters, recognizing letters with diacritics and the ability to differentiate between different diacritics of a single letter, as well as recognizing words not used in training where they consisted of letters used in training. The recognition rate achieved by the model for letters is 67.17% and 41.75% for words.

**Keywords**: Qaidah Noraniah, diacritics, Recognition.

**Cite this article as:** O. B. Ali, A. Bughari, H. F. Hamed, A. A. Salih, "Smart Spoken and the Diacritics Arabic Words Recognition System based on Qaidah Noraniah model," *African Journal of Advanced Pure and Applied Sciences (AJAPAS)*, vol. 2, no. 4, pp. 369–377, October-December 2023.



# نظام ذكي للتعرف على الكلمات العربية المنطوقة وعلامات التشكيل بناء على نموذج القاعدة النورانية

عمر بلولة علي[1*]، عبدالحفيظ بخاري[2]، حمد فوزي حمد[3]، عبد الرحمن علي صالح[4]
[2،1] قسم هندسة الحاسوب، كلية العلوم الهندسية، جامعة النجم الساطع، البريقة، ليبيا
[3] قسم الحاسوب، كلية العلوم، جامعة اجدابيا، اجدابيا، ليبيا
[4] كلية الدراسات العليا، جامعة السودان للعلوم والتكنولوجيا، السودان

**الملخص**
يقدم هذا البحث نظام يتعرف على الكلمات العربية المنطوقة والتشكيلات الصوتية استنادًا إلى مفهوم القاعدة النورانية. اللغة العربية هي واحدة من أكثر اللغات شيوعًا في العالم. لا يوجد العديد من أنظمة التعرف على الكلام للغة العربية. بسبب المفردات الواسعة والصرف، فإنه من الأصعب تطوير أنظمة للاعتراف بها. لبعض هذه الأسباب، يستخدم في اللغة نظام حروف متلاصقة وتشكيلات صوتية (الحركات). تم تصميم نموذج لحل هذه المشاكل تطبيقاً على القاعدة النورانية. تم

تدريب النموذج باستخدام برنامج CMU Sphinx واختباره بواسطة متحدثين تابعين ومستقلين. حقق النموذج نتائج جيدة في التعرف على الحروف المنفردة، والتعرف على الحروف مع التشكيلات الصوتية، والقدرة على التمييز بين التشكيلات الصوتية المختلفة لحرف واحد، فضلاً عن التعرف على الكلمات التي لم يتم استخدامها في التدريب والتي تتألف من الحروف المستخدمة في التدريب. حقق النموذج معدل دقة للحروف بنسبة 67.17٪ وبنسبة 41.75٪ بالنسبة للكلمات.

**الكلمات المفتاحية:** القاعدة النورانية، علامات التشكيل، التعرف.

## Introduction

Automatic Speech Recognition (ASR) is a technology that has been developed to allow the computer to understand spoken speech and convert it to text[1], in order to make the computer behave like a human, which provides an environment for the user to use the computer in a simpler and easier way. Many ASR integrated systems have been developed for some languages, such as English, French, and Chinese, but some languages, despite their popularity and spread, have no integrated systems or are still in early stages[2]. This is due to several reasons, including language morphology and rich vocabulary. If we take the Arabic language as an example, it is one of the most widely spoken languages in the world. It represents the mother tongue for all Arabs and also for its religious status as the language of the Holy Quran, which makes non-Arab Muslims have an interest in learning the language. The agglutinative letters, letters with diacritics (الحركات), rich vocabulary and multitude of dialects are some of the reasons that make it difficult to create an Arabic ASR integrated system[3]–[5]**.**

The Qaidah Noraniah method is one of the greatest sciences related to the Holy Quran. This method is used to teach Arabic by teaching the pronunciation of letters, then how to connect letters with each other, and then learning to link diacritics (الحركات) with letters. It is also used for teaching other things like the prolongation (المد) and Shadda (الشدة) using a scientific method of gradual and by the teaching of voice[6]. The name of this rule comes from the person who created it, Noor Mohammed Haquani, who wrote a book called Al-Qaidah Al-Noriah (القاعدة النورية).

This paper focused on the problem of creating a dataset that contains all Arabic words due to the lack of audio files for the Arabic language and because most Arabs speak a dialect. And there's the issue of distinguishing between the diacritics of words with similar sets of letters.

In [7], a character-based language model was created for continuous speech recognition, with the goal of predicting words that were not seen in training data and producing a large vocabulary of words. The corpus of the language model is split into characters rather than words to allow reconstructing word boundaries from character data. In [2] three approaches were used to enhance the Arabic ASR system: the first approach to deal with the problem of Tashkeel of words by using a Decision Tree to generate pronunciation variants; the second approach by using hybrid acoustic models from two models; and the third approach by using processed text to improve the language model because of the lack of Arabic resources. The results of these improvements reduced the Word Error Rate (WER). In [8] a system was made to classify the correct pronunciation of Quran for both males and females. The system extracts formants from speech signals using the spectrogram feature and also PSD extracted using MATLAP software. The two features have been classified separately and then combined together, but no effect has been gained from this combination.

## Dataset

### Preparation of data

This step contains two important steps: preparing text data and collecting and preprocessing audio data.

### Text data

The words of text data are divided into letters. The first stage contains only the first letter. Then at each stage, the next character is added with the previous separately and then integrated into the group of two letters. Because the pronunciation of the word is determined by pronouncing a single letter or two letters together. **Figure 1** depicts word dividing stages for the Arabic word Marhaba (**مرحبا**). The first stage begins with the first letter, then adding the second letter to the first letter. At stage 2, the letter can be separated or compounded with the previous letter. Some letters must be compounded to appear on pronunciation, such as the letter which has Sokoon (**سكون ـْ**) must be compounded with the previous letter. Stage 5 represents all the letters of the word but with different forms, and one of these forms can be used to correctly pronounce the word.

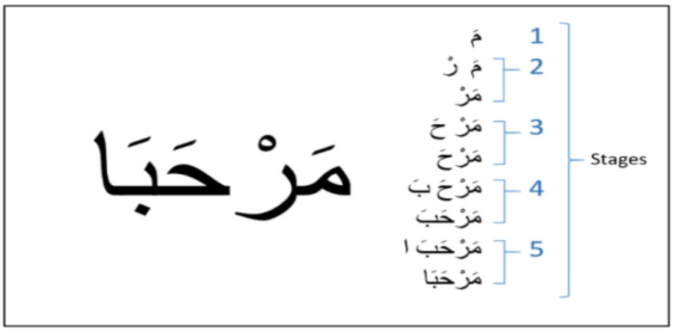


**Figure 1:** Word divide stages.

In this study, the letters of the Arabic sentence (**بُنِيَ الْإِسْلَامُ عَلَى خَمْسٍ** ), which contains 9 Arabic from 28 Arabic letters are used for training the model. The Arabic sentence is sliced into 17 parts. Every part contains a single letter or composed letters. These represent the pronunciation of a word or part of the word from the sentence.

Some Arabic letters have multiple forms; for example, the phoneme Alif (ا) in some words has multiple forms depending on the word (ى ,ا). One form will be used for butter recognizing. **Table 1** shows the words and the letters divided from them.

**Table 1** Letters of Arabic sentence (بُنِيَ الْإِسْلَامُ عَلَى خَمْسٍ)

| Word | Letter | | | | | | |
|---|---|---|---|---|---|---|---|
| بُنِيَ | بُ | بُنِ | نِ | نِيَ | يَ | | |
| الْإِسْلَامُ | ا | الْ | إِ | إِسْ | لَ | لَا | مُ |
| عَلَى | عَ | عَلَ | لَ | لَا (لَى) | ا (ى) | | |
| خَمْسٍ | خَ | خَمْ | سٍ | | | | |

## Audio data

The audio data will be recorded for text data for both genders' male and female, at different ages. Each person will be asked to record 10 audio samples. Then preprocessing will be done for these samples, such as removing noise, removing silence, and enhancing audio quality, and then the audio will be divided into letters and sub-words

The audio samples will be converted to wav files with a 16 KHz sample rate and a mono channel in order to be used for training. The CMU Sphinx toolkit [9] will be used for training and testing the model. To get the best results for this experiment, the sound samples will be evaluated by experts to determine the mispronounced letters and words. These steps are shown in **Figure 2**.

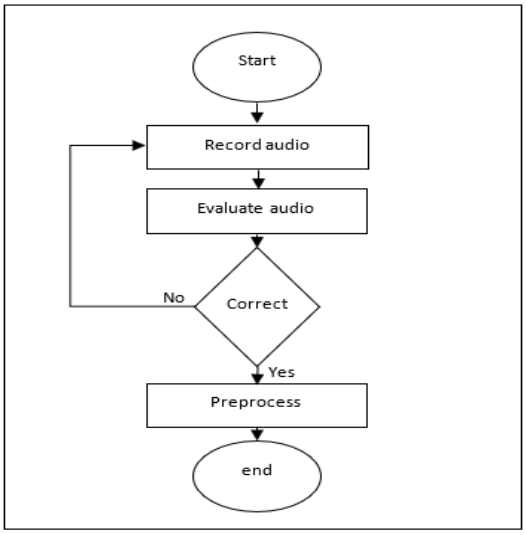


**Figure 2:** Audio sample evaluation process.

## Methodology

A total of 30 people, 21 male and 9 female, recorded audio data for text data. There ages ranged from 5 to 60 years old. Each person has 10 audio samples. Then preprocessing was done for the samples. The audio preprocessing was done by open source software called Audacity [10]. The audio samples are converted to wav files with a 16 KHz sample rate and mono channel in order to be used for training. The total amount of data is more than half an hour.

## Feature Extraction

For training, the feature will be extracted using MFCC, MFCC is the most used among other techniques for its high performance in extracting features [11], [12]. MFCC will be used to extract features from 4,250 audio samples that have been recorded by 25 people. For both male and female, each person has 10 recorded audio files. Every file contains 17 phonemes.

## Classification

The feature will be passed to the classifier to match the feature with the data from the model. HMM is used for classification. It is based on probabilities; it is used to find the next sequence based on probability[13]. The steps for training the model are explained in **Figure 3**.

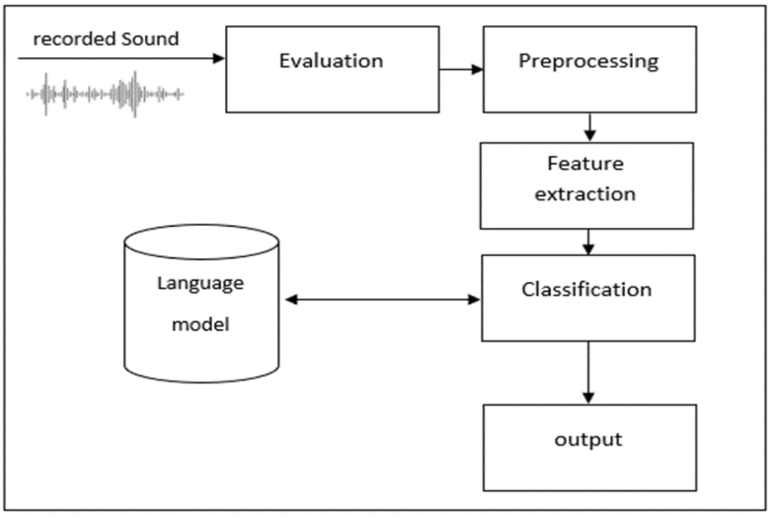


**Figure 3:** Training process.

## Testing the model

The model is tested by 10 speakers, 5 dependent speakers and 5 independent speakers (not used to train the model) from both genders. The test is done by recording audio samples for the words, sup-words, and letters in **Table 1** and **Table 2**. Each person is asked to record 5 audio samples. The total of audio samples is 1250.

**Table 2:** Words constructed from letters of Arabic sentence (بُنِيَ الْإِسْلَامُ عَلَى خَمْسٍ)

| Word |
|---|
| عَمُلَ |
| إِسْمُ |
| عَلَمُ |
| خَلَعَ |

## Results and discussion

The test results of the model for dependent speakers and independent speakers are described in **Table 3**. Letters and words are tested separately. The results are measured by WER in the following equation.

**Equation 1:** Word error rate

$$WER = \frac{Number\ of\ errors}{Total\ number\ of\ data}$$

**Table 3** Test results of dependent and independent speaker.

| Type | Amount | Word Error Rate (WER) |
|---|---|---|
| **Letter** | 17 | 32.9% |
| **Word** | 8 | 68.75% |

The results also showed letter recognition. Some letters are hard to recognize because the pronunciation of the letter makes it sound like another letter. In this test, the letter Alif (ا) is recognized in some incorrect letter recognition as the letter Aa (ع). **Table 4** show the error rate percentages of letters pronunciation. Some of the error rates for dependent speakers for letters (عَلَ, يَ) have been observed to be high, owing to mispronunciations and unclear pronunciation.

**Table 4:** Error rate of letters.

| NO | Letter | Error rate | |
|---|---|---|---|
| | | Dependent Speakers | Independent Speakers |
| 1 | ا | 12% | 26% |
| 2 | الْ | 8% | 16% |
| 3 | إِ | 32% | 46% |
| 4 | إِسْ | 6% | 20% |
| 5 | بُ | 8% | 8% |
| 6 | بُنِ | 2% | 2% |
| 7 | خَ | 12% | 18% |
| 8 | خَمْ | 10% | 20% |
| 9 | سِ | 4% | 12% |
| 10 | عَ | 6% | 32% |
| 11 | عَلَ | 16% | 10% |
| 12 | لَ | 12% | 14% |
| 13 | لَا | 30% | 38% |
| 14 | مُ | 12% | 28% |
| 15 | نِ | 34% | 40% |
| 16 | نِيَ | 14% | 0% |
| 17 | يَ | 8% | 2% |

On **Table 4** the 1st letter (ا) Alif with diacritic Fatha (فتحة ـَ) and 3rd letter (إِ) Alif with diacritic kasra (كسرة ـِ) are recognized as different letters.

The results for training the words and constructed words are shown in **Table 5** and **Table 6**. The error rate percentage is presented for words from both dependent and independent speakers. The error rate of words (بُنِيَ) is higher for dependent speakers due to mispronunciations of some speakers.

**Table 5:** Error rate of words of the Arabic sentence.

| NO | Word | Error rate | |
|---|---|---|---|
| | | Dependent Speakers | Independent Speakers |
| 1 | بُنِيَ | 6% | 4% |
| 2 | الْإِسْلَامُ | 42% | 50% |
| 3 | عَلَى | 2% | 16% |
| 4 | خَمْسٍ | 4% | 10% |

**Table 6:** Error rate of constructed words.

| NO | Word | Error rate | |
|---|---|---|---|
| | | Dependent Speakers | Independent Speakers |
| I. | عَمَلَ | 40% | 50% |
| II. | إِسْمُ | 38% | 50% |
| III. | عَلَمُ | 28% | 46% |
| IV. | خَلَعَ | 30% | 50% |

In **Table 6**, the error rate of words is not high for words not in training data. The error rate of the 4th word is high because the system recognizes the first and second letters as (عَلَ) and not (خَلَ). This is because the pronunciation sounds similar.

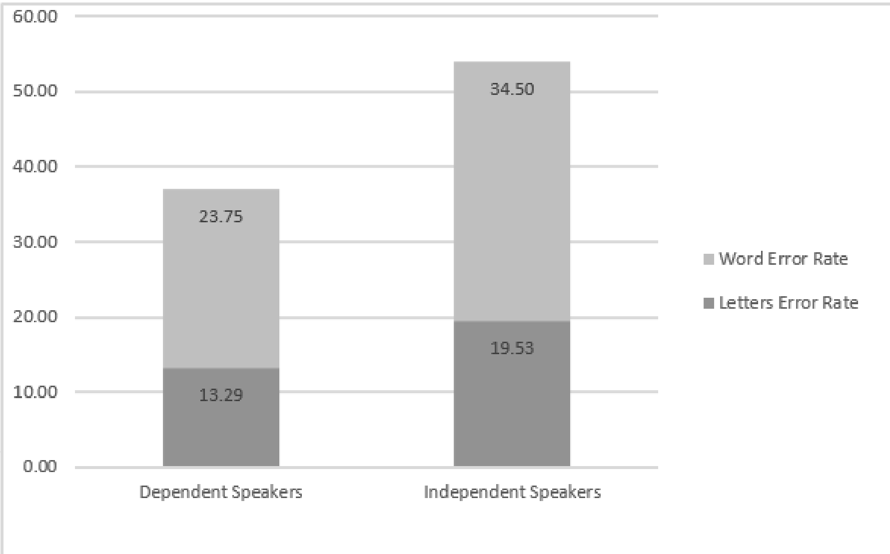


**Figure 4:** Total error rate for dependent and independent speakers.

**Table 7** shows the total number of letters and words used for testing and also shows the error rate for both dependent and independent speakers.

**Table 7:** Total data amount and error rate.

| Name | Description |
|---|---|
| Total letters | 850 |
| Total words | 400 |
| **Total data amount** | **1250** |

| | |
|---|---|
| Dependent error rate of letters | 13.29% |
| Dependent error rate of words | 23.75% |
| **Dependent error rate** | **16.64%** |
| Independent error rate of letters | 19.52% |
| Independent error rate of words | 34.5% |
| **Independent error rate** | **24.32%** |
| Total error rate for letters | **32.82%** |
| Total error rate for words | **58.25%** |
| **Total error rate** | **40.96%** |

The total error rate is 40.96%. This rate is achieved by re-evaluating the test results. A letter or word can have more than one form. See **Figure 1**. The error rate is reduced by giving all types of forms of composed letters or words. This helps in making the recognition process fixable, which reduces the error rate. **Table 8** shows example of possible forms.

**Table 8:** Example of possible forms of word.

| Word | Possible forms |
|---|---|
| بُنِيَ | بُ نِ يَ |
| | بُ نِيَ |
| | بُنِ يَ |

There are some obstacles to creating the system. These are: some of the audio samples were unclear and the pronunciation was not correct or clear; some people have a problem when they pronounce single or composed letters but they don’t have a problem when they pronounce the word. These reasons have affected the recognition rate for the system.

The total recognition rate for the proposed method is 59.04%, but the recognition rate that is achieved for letter recognition is 86.7% for dependent speakers and 80.47% for independent speakers. The total recognition rate for both is 67.17%. In comparison to a system made for recognizing single Arabic letters [14], the system used MFCC and LPC for extracting features for comparison study. The recognition rate achieved by MFCC is 59,87% and by LPC is 78,92%. Although the system is intended to recognize Arabic single letters with diacritics, it has difficulty distinguishing between diacritics for single letters. Another research[8] this research is classify the correct pronunciation for Arabic letter with Sukoon (ْ) for both male and female. The system extracted the formant of audio signal and used PSD as a classifier. The total recognition rate is 62%. these comparisons are presented in **Table 9**.

**Table 9:** A comparison of the proposed method with another research.

| **Reference** | [14] | | [8] | **This research** |
|---|---|---|---|---|
| **Feature extraction** | **MFCC, LPC** | | **formant** | **MFCC** |
| **Classification** | **KNN** | | **PSD** | **HMM** |
| **Amount of data used for training** | **4032 samples from 6 people** | | **22 samples from 22 people** | **4250 samples from 35 people** |
| **Recognize word from letters** | **no** | | **no** | **yes** |
| **Recognition rate** | **59,87% by MFCC,** | **78,92% by LPC** | **62%** | **67.17%** |

This proposed method has the ability to recognize words that are constructed from the letters used for the training, the recognition rate is 76.25% for dependent speakers and 65.5% for independent speakers, the total recognition rate for both is 41.75%, the system achieved this rate although it was not trained by these words. The system also has the ability to distinguish between the pronunciation of diacritics of single letter.

**Conclusion**

Arabic speech recognition systems are facing many problems. There are a lot of factors that affect recognition due to the Arabic language morphology. Some of these factors were explained in the first chapter, and one of them is letter diacritics (الحركات).

By developing the Qaidah Noraniah model, which focuses on the pronunciation of letters with diacritics to help in differentiating between words with the same letters and different diacritics, and by offering a solution by utilizing letter recognition to recognize words, the proposed method seeks to solve some of these issues.

Using the Qaidah Noraniah model for speech recognition proves the ability to recognize words constructed from letters used in training and recognize the difference between diacritics of one letter. However, the total recognition rate of both dependent and independent speakers for letters was 67.17%, which is not great for some reasons; first the amount of audio data used for training, second the number of letters, and finally the speaker's pronunciation of letters. However, the system is able to recognize words constructed from letters used for training. The recognition rate for words was 41.75%.

The working method of the Qaidah Noraniah model for dividing words into letters and using many forms for words helps reduce error rates and gives the ability to recognize single letters, composed letters, words, and words generated from sets of letters.

A lot of work can be done to reduce the error rate by improving the model, such as increasing the amount of text and audio data and using different techniques for feature extraction or classification.

**References**


[1] P. Saini and P. Kaur, "Automatic Speech Recognition : A Review," *Int. J. Eng. Trends Technol.*, vol. 4, no. iii, pp. 132–136, 2013.

[2] B. H. A. Ahmed and A. S. Ghabayen, "Arabic Automatic Speech Recognition Enhancement," *Proc. - 2017 Palest. Int. Conf. Inf. Commun. Technol. PICICT 2017*, no. May, pp. 98–102, 2017.

[3] T. R. Saeed, J. Salman, and A. H. Ali, "Classification improvement of spoken arabic language based on radial basis function," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 1, pp. 402–408, 2019, doi: 10.11591/ijece.v9i1.pp402-408.

[4] A. H. A. Absa, M. Deriche, M. Elshafei-Ahmed, Y. M. Elhadj, and B. H. Juang, "A hybrid unsupervised segmentation algorithm for arabic speech using feature fusion and a genetic algorithm (July 2018)," *IEEE Access*, vol. 6, no. July, pp. 43157–43169, 2018, doi: 10.1109/ACCESS.2018.2859631.

[5] A. Al-Sabri, A. Adam, and F. Rosdi, "Automatic detection of Shadda in modern standard Arabic continuous speech," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4–2, pp. 1810–1819, 2018, doi: 10.18517/ijaseit.8.4-2.6813.

[6] A. Mohammed, "Noorani Qaida," 2018. https://mawdoo3.com/ما_هي_القاعدة_النورانية.

[7] P. Smit, S. R. Gangireddy, S. Enarvi, S. Virpioja, and M. Kurimo, "Character-based units for unlimited vocabulary continuous speech recognition," *2017 IEEE Autom. Speech Recognit. Underst. Work. ASRU 2017 - Proc.*, vol. 2018-Janua, pp. 149–156, 2018.

[8] S. Khairuddin *et al.*, "Classification of the Correct Quranic Letters Pronunciation of Male and Female Reciters," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 260, no. 1, pp. 0–12, 2017.

[9] U. Carnegie Mellon, "CMUSphinx Open Source Speech Recognition." 2015, Accessed: Jun. 23, 2022.

[10] T. Audacity, "Audacity," *Name Audacity Is a Regist. Trademark Dominic Maz. Retrieved from http//audacity. sourceforge. net*, 2017, Accessed: Jun. 23, 2022.

[11] Z. Sakka, E. Techini, and M. S. Bouhlel, "Using geometric spectral subtraction approach for feature extraction for DSR front-end Arabic system," *Int. J. Speech Technol.*, vol. 20, no. 3, pp. 645–650, 2017.

[12] W. A. Pradana, Adiwijaya, and U. N. Wisesty, "Implementation of support vector machine for classification of speech marked hijaiyah letters based on Mel frequency cepstrum coefficient feature extraction," *J. Phys. Conf. Ser.*, vol. 971, no. 1, 2018.

[13] S. R. Eddy, “What is a hidden Markov model?,” *Nat. Biotechnol. 2004 2210*, vol. 22, no. 10, pp. 1315–1316, Oct. 2004.

[14] Adiwijaya, M. N. Aulia, M. S. Mubarok, W. Untari Novia, and F. Nhita, “A comparative study of MFCC-KNN and LPC-KNN for hijaiyyah letters pronounciation classification system,” *2017 5th Int. Conf. Inf. Commun. Technol. ICoIC7 2017*, vol. 0, no. c, pp. 2–6, 2017.