Speech Recognition Mobile Application for Learning Iqra'Using PocketSphinx

by Arbi Haza Nasution

Submission date: 18-Mar-2023 10:52AM (UTC+0700)
Submission ID: 2039826380
File name: P9_dragged.pdf (770.17K)
Word count: 2708
Character count: 14104

Speech Recognition Mobile Application for Learning Iqra' Using PocketSphinx

Arbi Haza Nasution[2(✉)], Winda Monika[1], and Tengku Waldi Firmansyah Masnur[2]

[1] Library Science Department, Faculty of Cultural Studies, Universitas Lancang Kuning, Pekanbaru, Indonesia
windamonika@unilak.ac.id

[2] Informatics Engineering Department, Faculty of Engineering, Universitas Islam Riau, Pekanbaru, Indonesia
arbi@eng.uir.ac.id, tengkuwaldi95@student.uir.ac.id

1 Introduction

Quran is a holy book of Moslem that conceives the teaching of Islam and is used as a guidance to live in this world and the hereafter. Every Moslem is obligated to recite and learn the Quran in order to fully understand and follow the teaching. By learning the Quran properly and thoroughly, the misinterpretation of the Quran could be avoided to have a peaceful life.

Reciting the Quran is an important activity for Moslems because it is believed as a way to receive the reward from God abundantly. Learning the Quran is quite hard for self-learning. Most of the time, an expert that teaches Quran is available at mosque and special schools with a fixed schedule. Sometimes, it is difficult for adults to adjust their time to join the programs. Due to the complexity conceived in learning Quran and the time concern for those who wish to learn from the expert, technologies that enable people to learn fastly, easily, and independently are highly needed.

Several methods had been invented to help learning and reciting the Quran. One of the old and famous methods to learn the Quran in Indonesia is by using a book named Iqra. This book was invented in the 1990s and is still being used now. Iqra helps people to read and recite Arabic letters and words that are used in the Quran. Studies on development of language resources are conducted for low-resource languages like Arabic language [1] and Indonesian ethnic languages [2–7]. Learning the Quran is not merely about knowing the letters or words, but there are rules that should be followed accordingly. In Iqra, those rules such as makhraj and tajwid are taught in the simplest way. Makhraj is how the Arabic letters should be read/pronounced and Tajwid is the rules that should be followed so that the meaning of the words is conveyed correctly.

Several studies showed good use of information technology in helping people learning Quran [8–12]. Some studies had done on developing multimedia tools and applications to learn Iqra [13, 14]. However, the multimedia tools are limited to convert the printed Iqra to a digital one and direct how the words are pronounced.

Speech recognition is a current technology that is massively being developed. This technology enables a computer to process speech into written text and vice versa. A few

P. K. Pattnaik et al. (eds.), *Proceedings of 2nd International Conference on Smart Computing and Cyber Security*, Lecture Notes in Networks and Systems 395,
https://doi.org/10.1007/978-981-16-9480-6_23

studies had been conducted implementing speech recognition in learning Quran [15] that uses machine learning application as a classifier by comparing the accuracy from three different classifiers (i.e., random forest, J48, and Naïve Bayes) with the highest accuracy of 0.8 for random forest and Naïve Bayes. Another study conducted [16] uses the Mel Frequency Cepstrum Coefficient (MFCC) method as a voice feature extension. This study carried out a high percentage of the accuracy of the speech recognition, however, the training data set that is being used was level one (1) of iqra which provides basic hijaiyah or single word only. Pocketsphinx has been used in the development of Android-based speech recognition applications in several domains [17–20].

This study aims to apply speech recognition using Mel Frequency Cepstrum Coefficients and Hidden Markov Model using PocketSphinx [21]. We try to figure out the accuracy of the speech recognition applied and factors that might be able to work on the accuracy. Furthermore, mobile phone application is used to help Moslems in learning Quran easily while correcting the mistakes when pronouncing the Arabic letters and words fastly and independently.

2 Research Method

In this study, there are 2 types of data collected for data training: vocab data and speech data. The vocab data is the data that contains a collection of hijaiyah/Arabic words. There are 70 words from the Iqra' book collected. The speech data is the data that contains a collection of voice recording data. We record the voices of 6 trainers with various range of age and different gender by using an audio recorder. Each trainer is asked to record the voice 4 times for each word in a quiet room. There are 1,168 voices recorded as data that consist of 292 male voices data, 292 female voices data, 292 boy voices data, and 292 girl voices data.

2.1 Speech Recognition

The first process in voice processing (speech recognition) is feature extraction. Feature extraction is a process to characterize the changing of sound data into image data in the form of a wave spectrum. The vocal structure is presented in the envelope of the short-time spectrum of power, and the function of the MFCC (Mel Frequency Cepstral Coefficients) is to represent this envelope accurately. Hidden Markov Model (HMM) is a statistical model of a system that assumes a Markov process with unknown parameters. In HMM, the state cannot be observed directly, but what can be observed are the variables that are affected by the state. Each state has a probability distribution of possible output tokens. The token sequence generated by the HMM, therefore, provides some information about the sequence of states. The results of the feature extraction process in the form of mel cepstrum will be compared with the phonetic file in the training data, after obtaining the appropriate pattern, the system will provide text output to the user.

The training process is a process where vocab data and speech data are processed to produce output in the form of training data that will be used as a dataset on the system. The training process is carried out as shown in Fig. 1. The vocab data was inputted

into the system, then the data is compiled using the Sphinx Knowledge Base Tools[1] to generate dictionary files and language model files. The speech data was inputted and converted using an online Audio Converter into an audio file with WAV (Waveform Audio Format) format, monotype, 16-bit resolution, and 16,000 Hz sampling rate. All these files will be entered into the pocket sphinx.

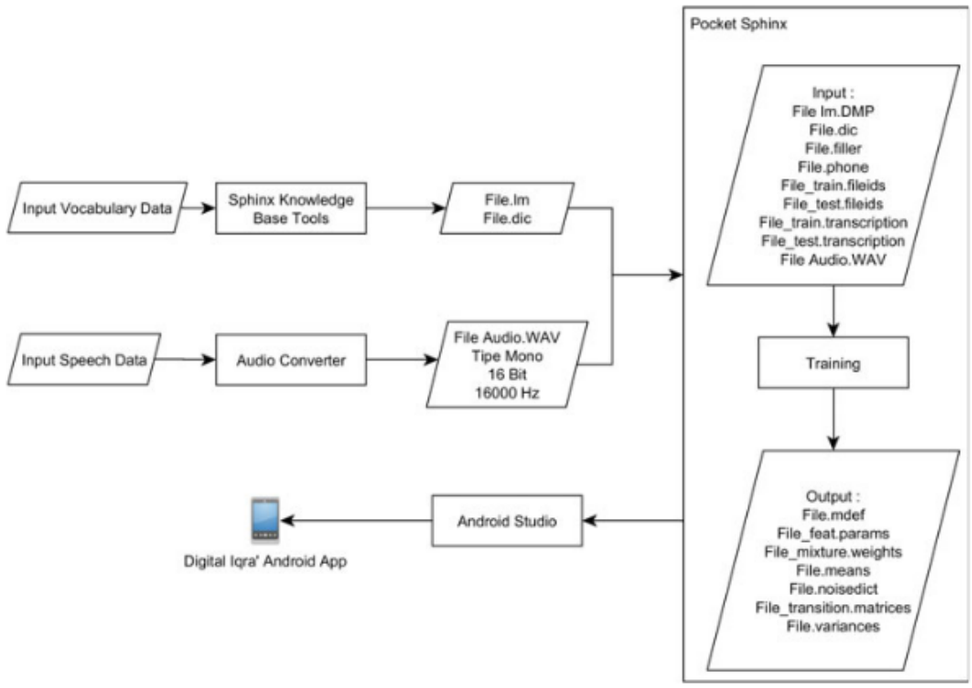

Fig. 1. System architecture

At the training stage, a number of data are needed such as dictionary files, language model files, filler files, phone files, transcription files, file ids, and WAV files. This training stage will produce output in the form of training data consisting of mdef files, feat.params files, mixture_weights files, means files, noisedict files, transition_matrices files, and variances files that will be used in android studio. At this stage, the model is deployed into Android apps. The steps are as follows:

[1] http://www.speech.cs.cmu.edu/tools/lmtool-new.html.

Data Training Processing

```
private void setupRecognizer(File assetsDir) throws
IOException {
    recognizer = SpeechRecognizerSetup.defaultSetup()
      .setAcousticModel(new File(assetsDir, "folder_akustik"))
      .setDictionary(new File(assetsDir, "file.dic"))
      .setRawLogDir(assetsDir)
      .getRecognizer();
    recognizer.addListener(this);
    File hijaiyahGrammar = new File(assetsDir, "file.gram");
    recognizer.addGrammarSearch(DIGITS_SEARCH,
            hijaiyahGrammar);
}
```

The program logic design is a flowchart that will clarify the application of this digital iqra' as shown in Fig. 2. The user can start speaking and giving a voice input by pressing the mic button, the system will perform the feature extraction process and compare the voice data with the training data. Then the system will give a result in the form of the word hijaiyah output. If the output word hijaiyah is the same as the target word hijaiyah listed on the form, a correct notification will appear, however, if the output word hijaiyah is not the same as the image, an incorrect notification will appear and the system will play the pronunciation audio of the correct hijaiyah words.

3 Result and Discussion

System testing is carried out to determine the validity of the final result or output of the system in the form of text and sound. This test uses training data as many as 1,168 speech sound data files that represent 73 hijaiyah words in the dictionary. Figure 3 shows the Mobile Application Digital Iqra' interface. Black box testing is carried out to evaluate whether the application can recognize the exact hijaiyah words or not.

To determine whether the distance from the sound source to the microphone can affect the accuracy of the system, 2 experiments were carried out by trained speakers where the speaker's voice was recognized by the system with different distances, which are 5 cm for the first experiment and 10 cm for the second experiment.

The result of the first experiment shows that from the results of testing the program from a total of 292 trials, the total average accuracy is 72.94%. While for the second test, the total average accuracy is 68.49%. Based on the two tests above, it can be concluded that the safe distance from the sound source should be about 5 cm to get the best result.

As shown in Table 1, the experiment was carried out on 73 hijaiyah words. The experiment was carried out by 6 speakers, which are 4 trainers (i.e., a male, a female, a boy, and a girl) and 2 testers (i.e., a male and a female). Each speakers needs to record 4 voices for every hijaiyah word. The accuracy of the Mobile Application Digital Iqra' is shown in Fig. 4.

The result shows that the application can detect testers' voices with just 10.27% and 8.94% decrease in accuracy compared to the trainers' voice for male and female, respectively. The accuracy for the boy and girl trainers are lower than the male and female trainers due to the low pronunciation quality and consistency of the young trainers during the training phase.

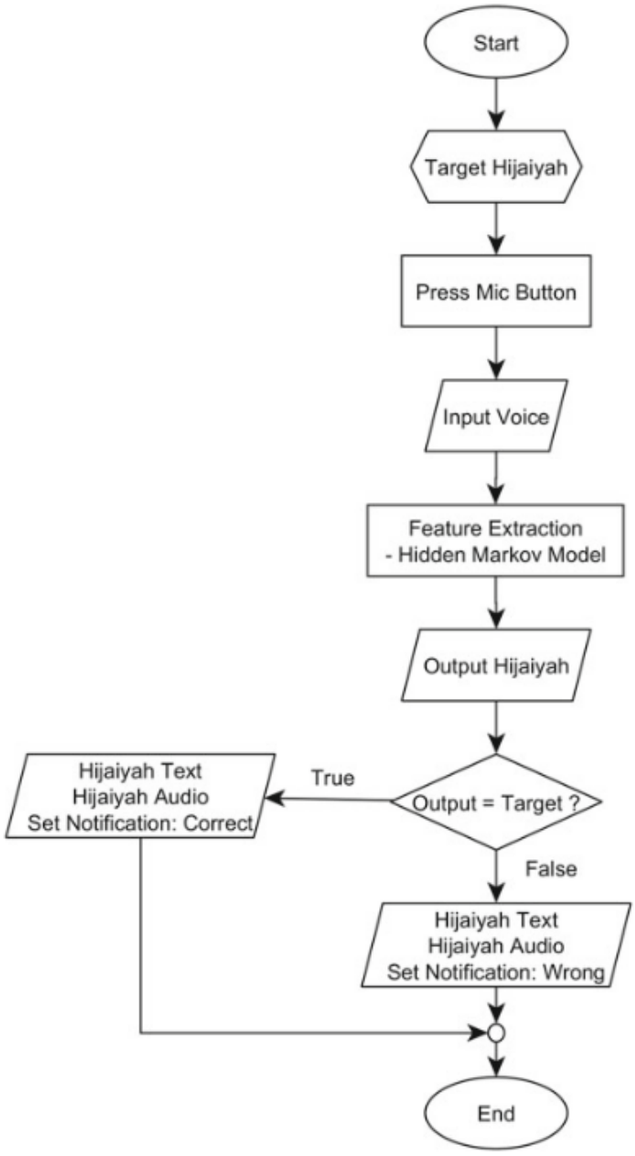

Fig. 2. System workflow

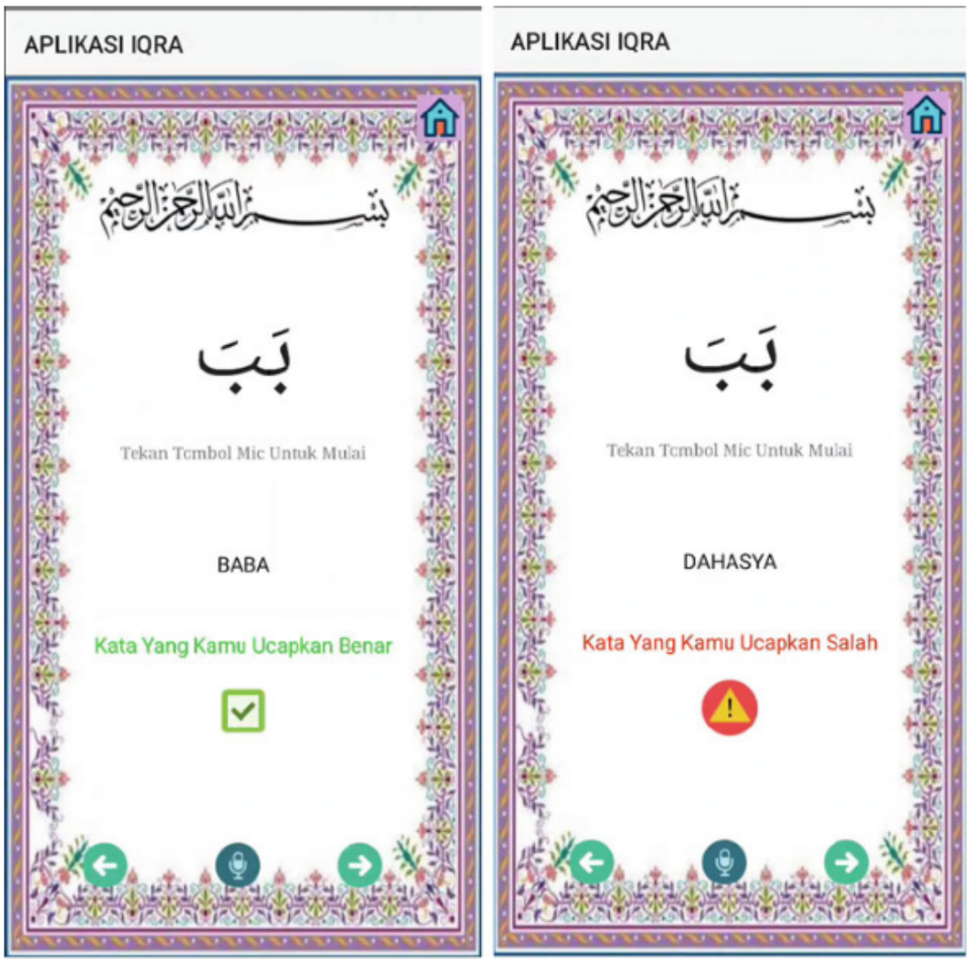

Fig. 3. Mobile application digital Iqra' interface

Table 1. List of 73 Hijaiyah word and pronunciation

No	Word	Pronunciation
1	بَبَ	Baba
2	بَذْ	Badza
3	جَمَ	Jama
4	كَنْ	Kana
5	خَدَ	Khoda
6	يَتَ	Yata
7	يَسْ	Yasya
8	نَفَ	Nafa
9	غَظْ	Ghozo
10	صَحَ	Shoha
11	طَهَ	Thoha
12	قَكَ	Qoka
13	صَدَ	Shoda
14	بَنَنَ	Banana
15	بَنَرَ	Banaro
16	بَدَرَ	Badaro
17	زَهَبَ	Zahaba
18	وَنَدْ	Wanadza
19	نَبَتَ	Nabata
20	يَدَنَ	Yadana
21	نَزَلَ	Nazala
22	بَتَّرَ	Bataro
23	جَعَّلَ	Ja'ala
24	نَبَغَ	Nabagho
25	نَفَلَ	Nafala
26	طَبَقَ	Thobaqo
27	حَكَمَ	Hakama
28	كَدَرَ	Kadaro
29	جَلَلَ	Jalala
30	ظَلَمَ	Zolama
31	بَلَغَ	Balagho
32	كَمَدَ	Kamada
33	قَلَمَ	Qolama
34	كَهَنَ	Kahana
35	سَهَيَ	Sahaya
36	دَهَشَنَ	Dahasya

No	Word	Pronunciation
37	لَهَبَ	Lahaba
38	طَلَعَ	Thola'a
39	مَنَحَ	Manaha
40	سَيَبَ	Sayaba
41	تَبَنَ	Tabana
42	لَضَلَّلَ	Ladholala
43	فَبَلَغَ	Fabalagho
44	لَسَلَكَ	Lasalaka
45	لَنَبَاَ	Lanabaa
46	لَاَكَلَ	Laakala
47	لَلَاَمَ	Lalaama
48	لَاَمَرَ	Laamaro
49	بَلَدِ	Baladi
50	لَزِمَ	Lazima
51	شَهِدَ	Syahida
52	بَطَاِح	Bathoihi
53	خَشِعَ	Khosyi'a
54	قَتَرَتِ	Qotaroti
55	رَضِيَ	Rhodiya
56	حَسَنِ	Hasani
57	نَبَتِ	Nabati
58	نَجِسِ	Najasi
59	فَكِرِحَ	Fakariha
60	حَفِظَ	Hafizo
61	حَسُنَ	Hasuna
62	كَرُمَ	Karuma
63	يَمُنَ	Yamuna
64	بَايِنَ	Baayina
65	لَاحَبَ	Laahaba
66	لِبَاسِ	Libaasi
67	يَكُنُ	Yakuunu
68	يَتُبُ	Yatuubu
69	تَوَابَ	Tawaaba
70	حَاسَدَ	Haasada
71	كَاتِبِ	Kaatibi
72	مَالِكِ	Maaliki
73	وَكِيْلُ	Wakiilu

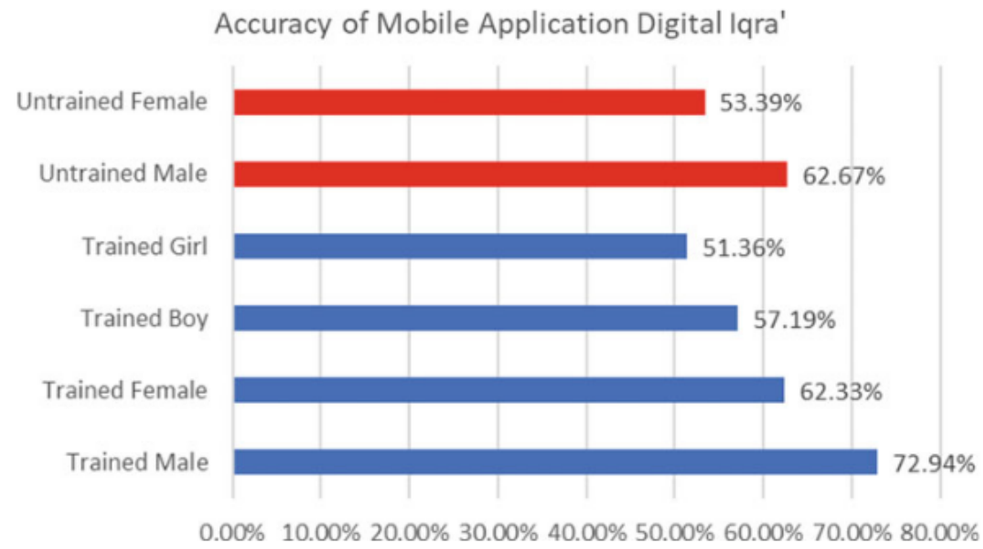

Fig. 4. Accuracy of mobile application digital Iqra'

We conducted usability testing for the mobile application digital iqra'. This testing is conducted to analyze the applicability of the mobile application. There are 6 questioner items being asked to 20 participants with different gender and various range of ages. Three scales are applied as measurements (i.e., Good, Fair, and Bad). The questionnaire items are as shown in Table 2. According to the usability testing result, 18 out of 20 participants perceive that the mobile application is useful and 17 out of 20 participants think that the user interface of the mobile application is good. The average usability testing of the mobile application is 76.67% which means the usability of the mobile phone is quite high.

Table 2. Usability testing of mobile phone application results

Items	Good	Fair	Bad
User interface	17	3	0
Useful	18	2	0
Functionality	15	5	0
Pleasant	12	8	0
Understandable	15	5	0
Recommended to be used	15	5	0
Average	76.67%	23%	0%

4 Conclusion

The result shows that the safe distance from the sound source should be about 5 cm to get the best result. The farther the sound source from the microphone, the smaller the accuracy. The result shows that the application can detect testers' voices with just 10.27% and 8.94% decrease in accuracy compared to the trainers' voice for male and female, respectively. The average usability testing of the mobile application is 76.67%, which means the usability of the mobile phone is quite high.

References

1. Hamed I, Denisov P, Li CY, Elmahdy M, Abdennadher S, Vu NT (2022) Investigations on speech recognition systems for low-resource dialectal Arabic–English code-switching speech. Computer Speech & Language 72:101278. https://doi.org/10.1016/J.CSL.2021.101278
2. Nasution AH (2018) Pivot-based hybrid machine translation to support multilingual communication for closely related languages. World Trans Eng Technol Educ
3. Nasution AH, Murakami Y, Ishida T (2019) Designing a collaborative process to create bilingual dictionaries of Indonesian ethnic languages. In: LREC 2018 - 11th International Conference on Language Resources and Evaluation
4. Nasution AH, Murakami Y, Ishida T (2016) Constraint-based bilingual lexicon induction for closely related languages. In: proceedings of the 10th international conference on language resources and evaluation, LREC 2016
5. Nasution AH, Murakami Y, Ishida T (2017) A Generalized Constraint Approach to Bilingual Dictionary Induction for Low- Resource Language Families. ACM Transactions on Asian and Low-Resource Language Information Processing. https://doi.org/10.1145/3138815
6. Nasution AH, Kadir EA, Murakami Y, Ishida T (2020) Toward Formalization of Comprehensive Bilingual Dictionaries Creation Planning as Constraint Optimization Problem. Optim Based Model Using Fuzzy Other Stat Tech Towards Env Sustainability. 41–54. https://doi.org/10.1007/978-981-15-2655-8_3
7. Nasution AH, Murakami Y, Ishida T (2021) Plan Optimization to Bilingual Dictionary Induction for Low-resource Language Families. Transactions on Asian and Low-Resource Language Information Processing 20. https://doi.org/10.1145/3448215
8. Veri J, Surmayanti S, Andini S (2020) Perancangan dan Pembuatan Media Pembelajaran Berbasis Android pada Mata Pelajaran Iqra. J. PTI (PENDIDIKAN DAN Teknol. INFORMASI) Fak. Kegur. ILMU Pendidik. Univ. PUTRA Indones. YPTK PADANG, pp 1–9
9. Hakim LN, Monika W, Nasution S, Nasution AH (2020) Visualisasi Tematik Al-Qur'an Berbasis Knowledge Graph. Jurnal Linguistik Komputasional (JLK) 3:1. https://doi.org/10.26418/jlk.v3i1.28
10. Rahman MA, Kassim IAA, Rahman TA, Muji SZM (2021) Development of automated Tajweed checking system for children in learning Quran. Evol Electr Electron Eng 2(1)
11. Nahar KMO, Ra'ed M, Moy'awiah A, Malek M (2020) An efficient holy Quran recitation recognizer based on svm learning model. Jordanian J Comput Inf Technol 6(04)
12. Supriyadi T, Julia J, Iswara PD, Abdussalam A (2019) ICT-based Al-Qur'an phonology learning. J Phys Conf Ser 1402(7):77020
13. Fauzan A, Arwani I, Fanani L (2018) Pembangunan Aplikasi Iqro' Berbasis Android Menggunakan Google Speech. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer 2:29–35

14. Zubaidi A, Akbar A, Husodo AY (2019) Implementasi Google Speech API Pada Aplikasi Koreksi Hapalan Al-Qur'an Berbasis Android. Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTIKA) 1:1–8
15. Khan RU, Qamar A, Hadwan M (2019) Quranic reciter recognition: a machine learning approach. Adv Sci Technol Eng Syst J 4(6):173–176
16. Wibowo AS, Darmawan IDMBA (2021) Iqra reading verification with mel frequency cepstrum coefficient and dynamic time warping. J Phys Conf Ser 1722(1):12015
17. Manasa CS, Jeeva Priya K, Gupta D (2019) Comparison of acoustical models of GMM-HMM based for speech recognition in Hindi using PocketSphinx. In: Proceedings of 3rd international conference computing methodologies communication ICCMC, pp 534–539. https://doi.org/10.1109/ICCMC.2019.8819747
18. Le M, Le TM, Duc Bui V, Truong SN (2021) A low-cost speech to sign language converter. IJCSNS Int J Comput Sci Netw Secur 21(3):37 https://doi.org/10.22937/IJCSNS.2021.21.3.5
19. Al-Kaf HAG, Sulong MS, Joret A, Aminuddin NF, Mohammad CA (2021) QVR: Quranic Verses Recitation Recognition System Using PocketSphinx. Journal of Quranic Sciences and Research 2:35–41
20. Lakdawala B, Khan F, Khan A, Tomar Y, Gupta R, Shaikh A (2018) Voice to Text transcription using CMU Sphinx A mobile application for healthcare organization. Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018 749–753. https://doi.org/10.1109/ICICCT.2018.8473305
21. Huggins-Daines D, Kumar M, Chan A, Black AW, Ravishankar M, Rudnicky AI (2006) Pocketsphinx: a free, real-time continuous speech recognition system for hand-held devices. In: ICASSP, IEEE international conference on acoustics speech signal process, vol 1. https://doi.org/10.1109/ICASSP.2006.1659988

Speech Recognition Mobile Application for Learning Iqra'Using PocketSphinx

ORIGINALITY REPORT

8%	5%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Apri Siswanto, Akmar Efendi, Zalian Hasrin, Bustamil Arifin. "Chapter 18 Two-Factor Authentication for Safe Deposit Box Based on Embedded System", Springer Science and Business Media LLC, 2022 Publication	3%
2	www.astesj.com Internet Source	1%
3	Zabulis, Xenophon, Haris Baltzakis, and Antonis Argyros. "Vision-Based Hand Gesture Recognition for Human-Computer Interaction", Human Factors and Ergonomics, 2009. Publication	1%
4	repository.ub.ac.id Internet Source	1%
5	Atallah, L.. "The use of pervasive sensing for behaviour profiling - a survey", Pervasive and Mobile Computing, 200910 Publication	1%
6	Adi Saputro Wibowo, I Dewa Made Bayu Atmaja Darmawan. "Iqra reading verification with mel frequency cepstrum coefficient and dynamic time warping", Journal of Physics: Conference Series, 2021 Publication	1%
7	Submitted to Higher Education Commission Pakistan	1%

Student Paper

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		