



Alibbaa': Jurnal Pendidikan Bahasa Arab

Special Issue Vol. 01 2026

P-ISSN: 2721-1606 | E-ISSN: 2716-4985

doi: <https://doi.org/10.19105/ajpba.vi.24450>

Local Dialect Interference and AI-Based Arabic Speech Recognition: Evidence from Madurese Learners

Indi Hulwun Nasyithah¹, Habibur Rahman², Maulana Abdillah³

^{1,2,3} *Universitas Islam Negeri Madura, Indonesia*

Corresponding E-mail: indihulwunakademik@gmail.com

Abstract

This study investigates the influence of Madurese phonetic interference on the accuracy of AI-based Arabic Speech Recognition systems. The research focuses on vowel-shift deviations produced by Madurese native speakers in Arabic pronunciation and evaluates the performance of Google Speech-to-Text and OpenAI Whisper in recognizing dialect-influenced Arabic speech. Using a mixed-method approach, speech data were collected from 13 sixth-semester students of the Arabic Language Education Department at UIN Madura through direct voice recording involving isolated words, continuous Arabic sentences, and Qur'anic recitations. Acoustic observation using Praat identified systematic vowel shifts, particularly the transitions of /a/ to [e], /i/ to [e], and /u/ to [o], which generated acoustic variability in Arabic speech production. The findings reveal that OpenAI Whisper achieved higher recognition accuracy (84.0%) with a lower Word Error Rate (16.0%) compared to Google Speech-to-Text, which obtained an accuracy of 61.8% and a WER of 38.2%. The dominant errors included substitution, deletion, insertion, and segmentation errors, indicating instability in phoneme mapping caused by dialect-induced vowel transitions. The study concludes that current AI-based Arabic Speech Recognition systems remain sensitive to low-resource dialect interference and that acoustic variability significantly affects recognition stability in multilingual speech environments. These findings highlight the importance of developing more adaptive and linguistically inclusive ASR models for Arabic language learning contexts.

Keywords: *AI-Based Speech Recognition; Arabic Pronunciation; Phonetic Interference; Madurese Dialect; Automatic Speech Recognition*

Abstrak

Penelitian ini bertujuan untuk menganalisis pengaruh interferensi fonetik dialek Madura terhadap akurasi sistem AI-based Arabic Speech Recognition. Fokus penelitian diarahkan pada pergeseran vokal yang dihasilkan oleh penutur asli bahasa Madura dalam pelafalan bahasa Arab serta evaluasi performa Google Speech-to-Text dan OpenAI Whisper dalam mengenali ujaran Arab yang terinterferensi dialek lokal. Penelitian menggunakan pendekatan mixed methods dengan pengumpulan data melalui perekaman suara langsung terhadap 13 mahasiswa semester enam Program Studi Pendidikan Bahasa Arab UIN Madura yang membaca kata, kalimat Arab mengalir, dan teks Al-Qur'an. Observasi akustik menggunakan Praat menunjukkan adanya pola vowel shift yang sistematis, khususnya pergeseran /a/ menuju [e], /i/ menuju [e], dan /u/ menuju [o], yang menghasilkan acoustic variability pada pelafalan bahasa Arab. Hasil penelitian menunjukkan bahwa OpenAI Whisper memiliki tingkat akurasi lebih tinggi (84.0%) dengan Word Error Rate sebesar 16.0% dibandingkan Google Speech-to-Text yang memperoleh akurasi sebesar 61.8% dengan WER sebesar 38.2%. Bentuk kesalahan yang dominan meliputi substitution error, deletion error, insertion error, dan segmentation error yang menunjukkan ketidakstabilan pemetaan fonem akibat transisi vokal berbasis dialek. Penelitian ini menyimpulkan bahwa sistem AI-based Arabic Speech Recognition saat ini masih sensitif terhadap interferensi dialek low-resource dan bahwa variasi akustik lokal berpengaruh signifikan terhadap stabilitas pengenalan ujaran pada lingkungan multilingual. Temuan ini menegaskan pentingnya pengembangan model ASR yang lebih adaptif dan inklusif terhadap keragaman linguistik dalam pembelajaran bahasa Arab berbasis teknologi.

Kata Kunci: *AI-Based Speech Recognition; Pelafalan Bahasa Arab; Interferensi Fonetik; Dialek Madura; Automatic Speech Recognition*

Introduction

The integration of Artificial Intelligence (AI) in language learning has evolved beyond simple text recognition toward precise phonetic evaluation through Automatic Speech Recognition (ASR) systems. This technology is increasingly expected to function as an autonomous learning assistant capable of providing immediate feedback on the pronunciation accuracy of non-native speakers. However, global studies have highlighted that the effectiveness of ASR is frequently constrained by phonetic variability resulting from users' first-language (L1) backgrounds, in which AI algorithms often fail to recognize subtle

acoustic distortions in speech.¹ Therefore, examining the accuracy of AI systems in responding to specific phonetic variations has become essential to ensure the validity of these technologies within language education contexts.

Madurese speakers possess distinctive phonological characteristics that significantly influence vowel production in second-language speech, including Arabic. Within the Madurese phonological system, vowels tend to undergo shifts in sound quality depending on the preceding consonants, particularly aspirated and unaspirated consonants, resulting in variations between open and close vowel realizations. Sociolinguistic studies have demonstrated that such phonetic interference is persistent and forms recognizable pronunciation patterns among Madurese speakers when learning foreign languages.² This distinctive sound system presents a considerable challenge for AI-based speech recognition systems, which are generally trained using standardized Arabic vowel representations derived from fusha pronunciation norms.

Empirically, many students of the Arabic Language Education Department at UIN Madura exhibit systematic vowel-shift phenomena across nearly all Arabic vowel markers during pronunciation activities. This interference does not occur only in isolated phonetic contexts but appears comprehensively, in which the fathah vowel /a/ tends to shift toward [e], kasrah /i/ approaches a lower [e]-like realization, and dhammah /u/ shifts toward [o]. Preliminary observations indicate that these pronunciations do not completely transform into different vowels; rather, they occupy transitional phonetic spaces that are difficult for non-local listeners to distinguish, yet potentially problematic for AI-based accuracy assessment systems. The emergence of these systematic vowel-shift patterns creates a strong need to investigate the extent to which AI systems are capable of detecting highly nuanced phonetic deviations produced by Madurese speakers.

Previous studies have shown that Artificial Intelligence (AI)-based Automatic Speech Recognition (ASR) technologies have

¹ Sarwadi, Sarwadi. "Artificial Intelligence Integration in Second Language Pronunciation Training." *Pioneer: Journal of Language and Literature* 17, no. 1 (June 2025): 80. <https://doi.org/10.36841/pioneer.v17i1.6329>.

² Khuluq, Muchsinul, and Siti Zulaikha. "Tadakhkhulu Al-Lughatu al-Umm Fî Mahârati al-Kalâm Bimarkazi al-Lughah al-'Arabiyah Fî Ma'hadi al-'Inâyah Sumber Batu Lilbanât Pegantenan Pamekasan." *Alibbaa'*: Jurnal Pendidikan Bahasa Arab 1, no. 1 (July 2020): 1–23. <https://doi.org/10.19105/ajpba.v1i1.3511>.

developed into important instruments for second-language (L2) pronunciation learning, particularly through their integration into Computer-Assisted Language Learning (CALL) environments. Cucchiarini and Strik argue that ASR systems are capable of providing rapid pronunciation evaluation and automated feedback, making them effective tools for L2 pronunciation training.³ These findings are reinforced by the meta-analysis conducted by Thi-Nhu Ngo et al., which demonstrates that ASR integration generally contributes positively to improvements in pronunciation accuracy and learners' phonological awareness.⁴ Martin further emphasizes that advances in modern AI models have substantially enhanced speech recognition performance in processing human speech across diverse language-learning contexts.⁵ Nevertheless, such effectiveness does not remain equally stable across all speaker characteristics, as ASR performance continues to depend heavily on the quality of acoustic models and the diversity of speech data used during system training.

On the other hand, several studies have begun to demonstrate that ASR systems still face significant challenges when processing non-standard accents and dialectal variations. Hinsvark et al.'s study on accented speech recognition revealed that AI models tend to experience decreased accuracy when dealing with speech containing phonetic characteristics that do not fully correspond to the dominant acoustic representations within the system's training corpus.⁶ This issue becomes even more complex in the context of Arabic, as Arabic Automatic Speech Recognition exhibits high sensitivity to variations in vowel quality and phoneme articulation. Dhouib et al.'s review of Arabic ASR further indicates that Arabic speech recognition systems continue to

³ Cucchiarini, Catia, and Helmer Strik. "Automatic Speech Recognition for Second Language Pronunciation Training." In *The Routledge Handbook of Contemporary English Pronunciation*, 556–69. First edition. | Abingdon, Oxon ; New York, NY : Routledge, 2017. <https://doi.org/10.4324/9781315145006-35>.

⁴ Thi-Nhu Ngo, Thuy, Howard Hao-Jan Chen, and Kyle Kuo-Wei Lai. "The Effectiveness of Automatic Speech Recognition in ESL/EFL Pronunciation: A Meta-Analysis." *ReCALL* 36, no. 1 (January 2024): 4–21. <https://doi.org/10.1017/S0958344023000113>.

⁵ Martin, Joshua L., and Kelly Elizabeth Wright. "Bias in Automatic Speech Recognition: The Case of African American Language." *Applied Linguistics* 44, no. 4 (August 2023): 613–30. <https://doi.org/10.1093/applin/amac066>.

⁶ Hinsvark, Arthur, Natalie Delworth, Miguel del Rio, Quinten McNamara, Joshua Dong, Ryan Westerman, Michelle Huang, et al. *Accented Speech Recognition: A Survey*. June 2, 2021.

encounter limitations in processing dialectal speech and non-standard phonetic variations, particularly within low-resource speech environments.⁷ These findings are consistent with Issa's study on the linguistic bias of OpenAI Whisper toward non-native Arabic speech, which demonstrated that even modern AI models still experience recognition instability when processing Arabic pronunciation produced by non-native speakers.⁸

Although the existing literature on AI and phonetic interference is already extensive, only a limited number of studies have specifically examined the precision of AI systems when dealing with dialects characterized by closely related vowel realizations such as Madurese. Most previous AI-related studies have focused primarily on English speakers or speakers of standard Indonesian, resulting in a scarcity of data regarding AI accuracy in detecting vowel-shift phenomena, such as the transition of the Arabic fathah vowel toward [e]-like realizations. A systematic review of the literature indicates that current research remains largely concentrated on the development of general-purpose algorithms rather than on the specific evaluation of ethnically grounded phonetic distortions. This gap in the literature forms the basis for the present study, which seeks to examine the extent to which AI systems are capable of distinguishing subtle phonetic nuances between standard Arabic vowels and vowels influenced by the Madurese dialect.

This study was conducted to address the lack of empirical data concerning the effectiveness of AI technologies in detecting specific phonetic interference among Madurese dialect speakers. As identified earlier, existing AI systems tend to generalize speech patterns based on standardized speaker models, thereby overlooking the transitional vowel variations commonly produced by students of the Arabic Language Education Department at UIN Madura. By focusing on the shifts of fathah, kasrah, and dhammah vowels toward intermediate vowel realizations, such as /a/ toward [e], this study provides a concrete depiction of the current limitations of AI-based speech recognition

⁷ Dhoubib, Amira, Achraf Othman, Oussama el Ghoul, Mohamed Koutheair Khribi, and Aisha al Sinani. "Arabic Automatic Speech Recognition: A Systematic Literature Review." *Applied Sciences* 12, no. 17 (September 2022): 8898. <https://doi.org/10.3390/app12178898>.

⁸ Issa, Elsayed, Mahmoud Ali, and Kevin Hirschi. "Measuring Linguistic Bias in ASR: Whisper Large-v3 on Non-Native Speech versus Human Perception." *Procedia Computer Science* 275 (2026): 692–99. <https://doi.org/10.1016/j.procs.2026.01.080>.

systems. Therefore, the primary objective of this research is to map the extent of AI accuracy and failure in systematically recognizing such phonetic distortions.

The novelty of this study lies in its approach to AI evaluation, which does not merely examine general pronunciation errors but instead employs highly specific local phonetic interference parameters that have not been extensively explored in previous literature. Unlike earlier studies that focused primarily on broad pronunciation inaccuracies, the present study investigates subtle phonetic distinctions between acoustically similar yet phonologically different sounds, which constitute a unique technical challenge for AI algorithms. The use of participants from the Arabic Language Education Department at UIN Madura, who possess a strong Madurese dialect background, provides authentic primary data for the future development of technology-based language evaluation models. Practically, the findings of this study are expected to contribute significantly to the development of more inclusive technology-assisted Arabic language learning curricula that are responsive to the linguistic diversity of Indonesian regional dialects.

Method

This study employed a mixed-methods design by integrating quantitative and descriptive qualitative approaches to analyze the influence of Madurese dialect phonetic interference on the accuracy of AI-based Arabic Speech Recognition systems. The quantitative approach was used to measure the accuracy of speech recognition systems through the calculation of transcription error percentages and Word Error Rate (WER), while the qualitative approach was employed to describe phonetic interference patterns and the forms of speech recognition failures produced by AI systems. In addition, the study adopted a comparative analytical approach by comparing the performance of two speech recognition platforms, namely Google Speech-to-Text API and OpenAI Whisper, in recognizing Arabic pronunciation produced by Madurese dialect speakers. The primary focus of the research was directed toward identifying transitional vowel-shift patterns and examining their impact on phoneme-mapping stability within Automatic Speech Recognition (ASR) systems.

The participants consisted of 13 sixth-semester students from the Arabic Language Education Department at UIN Madura who were selected using purposive sampling techniques. Participant selection was

based on several criteria: they were native speakers of Madurese, actively used the Madurese dialect in daily communication, and had completed at least five semesters of formal Arabic language instruction. Sixth-semester students were selected because they were considered to possess adequate foundational competence in Arabic pronunciation while still exhibiting relatively consistent phonetic interference in Arabic vowel realization. With these characteristics, the participants were considered representative for observing the relationship between local phonetic interference and AI performance in Arabic speech recognition.

Data collection was conducted through direct voice recording and digital documentation of automatic transcription outputs. In the initial stage, the researchers constructed a speech corpus consisting of word lists, short sentences, continuous Arabic texts, and selected Qur'anic recitation passages designed to elicit the realization of the vowels /a/, /i/, and /u/ in various phonetic contexts. All linguistic instruments were purposively selected based on their potential to trigger vowel-shift phenomena among Madurese speakers. The recording process was conducted in a low-noise environment using high-quality recording equipment to maintain acoustic signal clarity and minimize noise interference that could affect ASR decoding processes. Each participant was instructed to read the entire speech corpus naturally, using their normal speaking tempo and intonation without corrective intervention from the researchers.

The collected audio data were subsequently uploaded to two AI systems, Google Speech-to-Text API and OpenAI Whisper, to obtain automatic transcription outputs. The transcription results from both platforms were documented and classified according to speech type, vowel-shift patterns, and phoneme recognition error patterns. To strengthen the phonetic validity of the data, acoustic observations were also conducted using Praat software through spectrogram analysis and formant frequency (F1 and F2) measurements in order to verify the vowel-shift characteristics produced by the participants. This analysis was intended to ensure that the observed vowel changes genuinely reflected Madurese phonetic interference patterns rather than merely individual articulation errors.

Data analysis was conducted in several stages. First, the researchers manually transcribed all participant utterances as baseline references for comparison with the AI-generated outputs. Second, the transcription results from Google Speech-to-Text and OpenAI Whisper were compared with the target utterances to identify phoneme

recognition errors. These errors were then categorized into substitution errors, deletion errors, insertion errors, and segmentation errors. Third, the frequency of each error category was calculated, and Word Error Rate (WER) was measured to quantitatively evaluate the accuracy of both ASR systems. The WER formula was used to calculate the ratio of total errors to the total number of target words by combining substitution, deletion, and insertion errors.

In the final stage, the quantitative findings were qualitatively interpreted using an acoustic-phonetic approach to examine the relationship between Madurese vowel-shift patterns and AI speech recognition failures. The analysis focused particularly on vowel-quality transitions toward intermediate acoustic realizations, such as shifts from /a/ toward [e], /i/ toward [e], and /u/ toward [o], as well as their effects on the stability of Arabic phoneme mapping within ASR systems. All findings were subsequently interpreted through the perspectives of first-language phonetic interference and acoustic variability in speech recognition in order to explain the limitations of AI systems in processing local dialectal phonetic variations within technology-assisted Arabic language learning contexts.

Results and Discussion

Analysis of the speech data produced by sixth-semester students of the Arabic Language Education Department at UIN Madura revealed consistent patterns of phonetic interference in Arabic pronunciation, particularly in vowel realization during continuous speech production. Based on acoustic observations conducted using Praat, vowel shifts did not occur solely in isolated word pronunciation but became more dominant in sentence reading and continuous Arabic texts. These shifts primarily appeared in final vowel endings (harakat al-i'rab), word-linking processes (washal), and long vowels (mad). In general, the most frequently observed interference patterns involved vowel transitions from /a/ toward [e], /i/ toward [e], and /u/ toward [o]. In addition, excessive emphasis on plosive consonants was identified, which affected the overall articulatory stability of speech production.

Table 1. Acoustic Characteristics of Vowel Shift in Arabic Sentence Reading

Linguistic Element	Phonetic Phenomenon	Madurese Realization	Frequency
Fathah Vowel	Shift from /a/ → [e]	<i>Iyyaka</i> → <i>Iyyeka</i>	88%
Kasrah Vowel	Shift from /i/ → [e]	<i>Bismillah</i> → <i>Besmellah</i>	82%
Dhammah Vowel	Shift from /u/ → [o]	<i>Alhamdu</i> → <i>Alhamdo</i>	94%
Diphthong/Harakat	Excessive consonant emphasis	Plosive consonants become more dominant	75%

The data presented in Table 1 indicate that the vowel shift from /u/ toward [o] constituted the most dominant interference phenomenon, occurring in 94% of the utterances. This was followed by the shift from /a/ toward [e] at 88% and /i/ toward [e] at 82%. These findings demonstrate that phonetic interference from the Madurese dialect in Arabic pronunciation was not sporadic but instead formed systematic transitional vowel patterns. In addition to changes in vowel quality, relatively strong consonantal emphasis was observed among most participants, particularly in plosive consonants. In several utterances, these vowel-quality changes did not produce complete phoneme substitution but instead generated intermediate acoustic realizations positioned between standard Arabic vowel categories. Such acoustic variability potentially affected the stability of speech recognition within AI-based speech recognition systems.

To measure the impact of phonetic interference on AI system performance, this study compared the transcription accuracy of Google Speech-to-Text API and OpenAI Whisper in processing Arabic speech produced by the participants across various speech contexts.

Table 2. Comparative Accuracy of AI Transcription across Speech Contexts

AI Platform	Short Sentence Accuracy (%)	Long Sentence Accuracy (%)	Qur'anic Text Accuracy (%)	Overall Average
Google Speech-to-Text	68.0%	55.5%	62.0%	61.8%
OpenAI Whisper	85.5%	78.0%	88.5%	84.0%

The results in Table 2 demonstrate that OpenAI Whisper consistently achieved higher accuracy rates than Google Speech-to-Text across all speech categories. In short sentences, Whisper achieved an accuracy rate of 85.5%, whereas Google STT obtained only 68.0%. The performance gap became more significant in long-sentence recognition, where Google STT accuracy declined to 55.5%, while Whisper maintained a relatively high accuracy rate of 78.0%. The highest recognition accuracy was identified in Qur'anic text recitation processed by Whisper, reaching 88.5%. These findings indicate that OpenAI Whisper possessed greater stability in processing continuous speech containing non-standard phonetic variations. Nevertheless, both systems still experienced performance degradation when confronted with strong vowel-shift interference, particularly in the realization of the vowels [o] and [e] resulting from Madurese dialect interference. The high accuracy achieved by Whisper in Qur'anic recitation also suggests differences in AI sensitivity toward specific speech types, particularly utterances characterized by more standardized rhythmic and phonological structures.

Further analysis was conducted to identify the forms of speech recognition errors appearing in the transcription outputs of both AI systems. The findings indicate that transcription errors did not occur randomly but instead followed specific linguistic patterns associated with changes in vowel quality and segmentation stability.

Table 3. Error Patterns in Arabic Speech Transcription

Target (Standard)	Sentence	Student Realization	Speech	AI Transcription Output	Error Type
<i>Bismillahirrahmani rrahim</i>		<i>Besmellaherrahmane rrahem</i>		بسم الله الرحمن رحيم	Segmentation Error
<i>Iyyaka Na 'budu</i>		<i>Iyyeka Na 'bodo</i>		إياك نعبد / نعبود	Vowel Substitution
<i>Alhamdulillah</i>		<i>Alhamdolellahe</i>		الحمد لله / لله	Loss of Phonetic Detail
<i>Shirathal Mustaqim</i>		<i>Shirothol Mostaqem</i>		صراط المستقيم	Omission / Deletion

Table 3 demonstrates that the most dominant error types in both AI systems were substitution errors and segmentation errors. Segmentation errors occurred when the systems failed to separate word structures accurately within continuous speech, particularly in utterances characterized by rapid rhythm and strong vowel shifts. Meanwhile, substitution errors appeared when the systems replaced transitional vowels resulting from Madurese interference with different standard Arabic phoneme representations. In several cases, the systems also exhibited loss of phonetic detail, reflected in the omission of certain vowel elements during the decoding process. Furthermore, omission or deletion errors were identified through the disappearance of alif and long-vowel representations in the transcription outputs. To quantitatively evaluate the impact of phonetic interference on AI system performance, this study employed the Word Error Rate (WER) approach through calculations involving substitution, deletion, and insertion errors.

Table 4. Word Error Rate (WER) Comparison between Google STT and OpenAI Whisper

Evaluation Component	Google STT	OpenAI Whisper
Total Words (N)	100	100
Substitution (S)	28	10
Deletion (D)	7	4
Insertion (I)	3	2
Total Errors (S+D+I)	38	16
WER (%)	38.2%	16.0%
Accuracy (%)	61.8%	84.0%

The data presented in Table 4 reveal that Google Speech-to-Text produced substantially higher error rates than OpenAI Whisper. Out of 100 evaluated words, Google STT generated 38 total errors, dominated by 28 substitution errors. In contrast, Whisper produced only 16 total errors with comparatively lower error distribution across all categories. These findings indicate that substitution errors constituted the most dominant form of recognition failure when AI systems encountered vowel shifts caused by Madurese dialect interference. Overall, the results demonstrate that phonetic interference from the Madurese dialect generated acoustic variability that directly affected phoneme-mapping stability and the accuracy of AI-based Arabic Speech Recognition systems.

Discussion

The findings of this study indicate that phonetic interference from the Madurese dialect significantly affects the stability of speech recognition in AI-based Arabic Speech Recognition systems. The consistently occurring vowel shifts in the realization of /a/, /i/, and /u/ generated acoustic variability that contributed to phoneme-mapping instability within ASR systems. In this context, changes in vowel quality did not always produce entirely new phonemes but instead generated transitional acoustic spectra positioned between standard Arabic vowel categories. Such acoustic variability presents a substantial challenge for AI systems because speech recognition models operate based on the correspondence between acoustic frequency patterns and previously learned acoustic representations.⁹ When vowel quality moves toward transitional regions, the systems experience a decline in their ability to decode speech accurately.

The findings further demonstrate that OpenAI Whisper exhibited more stable performance than Google Speech-to-Text in recognizing Arabic speech influenced by Madurese phonetic interference. This difference is clearly reflected in the Word Error Rate (WER) values, where Whisper produced an error rate of 16.0%, while Google STT reached 38.2%. The dominance of substitution errors in both systems indicates that the primary source of ASR failure lies in the inability of the systems to map transitional vowels into standard Arabic phoneme

⁹ Rukwong, Niyada, and Sunee Pongpinigpinyo. "An Acoustic Feature-Based Deep Learning Model for Automatic Thai Vowel Pronunciation Recognition." *Applied Sciences* 12, no. 13 (June 2022): 6595. <https://doi.org/10.3390/app12136595>.

representations. In Google STT, substitution errors appeared more frequently when the system encountered vowel-quality shifts toward [e] and [o]. This finding suggests that dictionary-oriented speech recognition models tend to experience difficulties when speech sounds do not fully conform to the standard phonetic patterns available within the training corpus.¹⁰

In contrast, OpenAI Whisper demonstrated better adaptability in processing continuous speech containing non-standard phonetic variations.¹¹ The transformer-based architecture enabled the system to maintain linguistic context more consistently, resulting in lower deletion and segmentation error rates compared to Google STT. Nevertheless, Whisper still experienced a decline in accuracy when confronted with strong vowel-shift interference, particularly in the realization of /u/ toward [o]. In standard Arabic orthography and phonology, the sound [o] does not possess an independent phonemic representation,¹² causing AI systems to experience instability when attempting to map such sounds into Arabic phonological structures. These findings suggest that improvements in AI model capacity have not yet fully overcome the limitations of acoustic representation in handling dialect-based speech variability.

From an acoustic-phonetic perspective, the high frequency of substitution errors in both systems was associated with changes in formant frequency patterns (F1 and F2) within participant speech. The shift of L1 vowel positions toward intermediate acoustic spectra created acoustic overlap between target phoneme representations and the speakers' actual sound realizations.¹³ Under such conditions, ASR

¹⁰ Thomas, Tessy M., Aditya Singh, Latané P. Bullock, Daniel Liang, Cale W. Morse, Xavier Scherschligt, John P. Seymour, and Nitin Tandon. "Decoding Articulatory and Phonetic Components of Naturalistic Continuous Speech from the Distributed Language Network." *Journal of Neural Engineering* 20, no. 4 (August 2023): 046030. <https://doi.org/10.1088/1741-2552/ace9fb>.

¹¹ Graham, Calbert, and Nathan Roll. "Evaluating OpenAI's Whisper ASR: Performance Analysis across Diverse Accents and Speaker Traits." *JASA Express Letters* 4, no. 2 (February 2024). <https://doi.org/10.1121/10.0024876>.

¹² Moussa, Touahmi, Ouamane Abdelmalik, and Chouchane Ammar. "A Guide to Cost-Effective Collection of Colloquial Algerian Arabic Speech Data." *2024 8th International Conference on Image and Signal Processing and Their Applications (ISPA)*, April 21, 2024, 1–6. <https://doi.org/10.1109/ISPA59904.2024.10536822>.

¹³ Perkins, Kyle, and Lawrence Jun Zhang. "The Effect of First Language Transfer on Second Language Acquisition and Learning: From Contrastive Analysis to Contemporary Neuroimaging." *RELC Journal* 55, no. 1 (April 2024): 162–78. <https://doi.org/10.1177/00336882221081894>.

systems experienced difficulties in determining the phoneme category most appropriate for the received acoustic signals. These findings are consistent with acoustic phonetics theory, which explains that digital speech recognition processes depend heavily on the stability of vowel-frequency patterns and the consistency of acoustic distribution within training datasets.¹⁴ Consequently, phonetic interference from the Madurese dialect not only affected the articulatory aspects of Arabic pronunciation but also directly influenced the computational processes underlying AI-based speech recognition.¹⁵

The results of this study also revealed that the highest recognition accuracy was achieved in Qur'anic text recitation, particularly with OpenAI Whisper, which reached an accuracy rate of 88.5%. This finding indicates that AI performance is influenced not only by speakers' phonetic characteristics but also by the type and structure of the processed speech. Qur'anic texts possess relatively standardized rhythmic structures, intonation patterns, and recitation forms compared to general Arabic speech,¹⁶ thereby enabling the systems to maintain greater decoding stability. In addition, the extensive exposure of Qur'anic digital corpora during AI model training may also have contributed to Whisper's ability to recognize repetitive murottal recitation patterns and standardized phonological structures. In other words, corpus familiarity constitutes an important factor influencing ASR sensitivity toward particular speech types.¹⁷

The findings of this study support numerous previous studies stating that regional accents and first-language interference significantly

¹⁴ Nasr, Seham, Rehab Duwairi, and Muhannad Quwaider. "End-to-End Speech Recognition For Arabic Dialects." *Arabian Journal for Science and Engineering* 48, no. 8 (August 2023): 10617–33. <https://doi.org/10.1007/s13369-023-07670-7>.

¹⁵ Khwaileh, Tariq, Eiman Mustafawi, and Yusuf Albustanji. "A Linguistically-Driven Response Categorisation Protocol for Arabic Nouns and Verbs: Clinical and Research Applications." *Clinical Linguistics & Phonetics* 34, no. 9 (September 2020): 861–77. <https://doi.org/10.1080/02699206.2019.1708464>.

¹⁶ Garadaghli, V. "The Problem of Standardizing the Phonetic Transcription for the Qur'an Readings of Asim and Nafi." *The Oriental Studies* 2024, no. 94 (December 2024). <https://doi.org/10.15407/skhodoznavstvo2024.94.233>.

¹⁷ Bogach, Natalia, Elena Boitsova, Sergey Chernonog, Anton Lamtev, Maria Lesnichaya, Iurii Lezhenin, Andrey Novopashenny, et al. "Speech Processing for Language Learning: A Practical Approach to Computer-Assisted Pronunciation Teaching." *Electronics* 10, no. 3 (January 2021): 235. <https://doi.org/10.3390/electronics10030235>.

affect speech recognition performance in second-language learning.¹⁸ However, the present study extends earlier research by demonstrating that such issues occur not only at the level of general accents but also in highly specific and systematic transitional vowel-shift phenomena among Madurese speakers.¹⁹ Within the context of Arabic Speech Recognition, such local phonetic variability presents a more complex challenge because Arabic possesses high sensitivity to changes in vowel quality.²⁰ Therefore, these findings demonstrate that the limitations of AI in Arabic language learning are associated not only with algorithmic sophistication but also with limitations in the linguistic and acoustic representations contained within system training corpora.

Practically, the findings indicate that the use of AI as an instrument for Arabic pronunciation assessment should be implemented cautiously, particularly within multilingual environments strongly influenced by local dialects. Although ASR systems are capable of supporting independent pronunciation practice and providing rapid feedback, speech recognition outputs do not always accurately represent actual pronunciation quality when the systems encounter non-standard phonetic variations. In some cases, AI systems may accept vowel-shifted pronunciations as correct utterances, while in other situations the systems may produce excessive corrections toward accent variations that remain linguistically intelligible. Therefore, within Arabic language learning contexts, AI-based pronunciation assessment should be positioned more appropriately as an assistive tool rather than as a complete replacement for human phonetic evaluation.

This study has several limitations that should be acknowledged. The number of participants was limited to students from the Arabic Language Education Department at UIN Madura and therefore does not fully represent the diversity of Madurese dialect variations across different regions. In addition, the study examined only two AI platforms,

¹⁸ Aissiou, M. "A Genetic Model for Acoustic and Phonetic Decoding of Standard Arabic Vowels in Continuous Speech." *International Journal of Speech Technology* 23, no. 2 (June 2020): 425–34. <https://doi.org/10.1007/s10772-020-09694-y>.

¹⁹ Sulisty, Danang Arbian. "LSTM-Based Machine Translation for Madurese-Indonesian." *Journal of Applied Data Sciences* 4, no. 3 (September 2023): 189–99. <https://doi.org/10.47738/jads.v4i3.113>.

²⁰ Tibi, Sana, Ashley A. Edwards, Christopher Schatschneider, and John R. Kirby. "Predicting Arabic Word Reading: A Cross-Classified Generalized Random-Effects Analysis Showing the Critical Role of Morphology." *Annals of Dyslexia* 70, no. 2 (July 2020): 200–219. <https://doi.org/10.1007/s11881-020-00193-y>.

limiting broader generalization regarding the performance of other speech recognition systems. Future studies are recommended to employ larger speech corpora, involve more diverse regional dialect variations, and develop more in-depth acoustic analyses concerning formant distribution and spectrogram characteristics of multilingual speakers' utterances. Consequently, future developments in AI-based Arabic Speech Recognition may be directed toward more adaptive models that are sensitive to users' phonetic and linguistic diversity.

Conclusion

This study demonstrates that phonetic interference from the Madurese dialect significantly affects the accuracy of AI-based Arabic Speech Recognition systems in recognizing the Arabic pronunciation of students from the Arabic Language Education Department at UIN Madura. The findings reveal that vowel-shift patterns involving the transitions of /a/ toward [e], /i/ toward [e], and /u/ toward [o] generated acoustic variability that influenced phoneme-mapping stability within Automatic Speech Recognition (ASR) systems. These phonetic variations were proven to produce substitution errors, deletion errors, insertion errors, and segmentation errors in AI transcription outputs. Comparatively, OpenAI Whisper exhibited more stable performance than Google Speech-to-Text, achieving an accuracy rate of 84.0% with a Word Error Rate (WER) of 16.0%, whereas Google STT achieved only 61.8% accuracy with a WER of 38.2%. These findings confirm that the capability of AI systems in Arabic speech recognition is determined not only by algorithmic sophistication but also by the sensitivity of acoustic models toward the phonetic variability of multilingual speakers and local dialects. Furthermore, the high accuracy observed in Qur'anic text recitation indicates that corpus familiarity and the degree of phonological standardization also influence ASR system performance in processing Arabic speech.

Nevertheless, this study still has several limitations. The number of participants was limited to students from the Arabic Language Education Department at UIN Madura and therefore does not fully represent the diversity of Madurese dialect variations across different regions. In addition, this study examined only two AI platforms, preventing broader generalization regarding the performance of other ASR models. Moreover, the acoustic analysis in this study remained focused on observing vowel-shift patterns and transcription error distributions without conducting more extensive experimental

exploration of formant dynamics and spectrogram characteristics. Therefore, future research is recommended to involve larger speech corpora, incorporate more diverse regional dialect variations, and employ more detailed acoustic-phonetic analysis approaches in order to strengthen understanding of the relationship between local phonetic interference and AI-based speech recognition. The development of ASR models that are more adaptive to low-resource dialects is also essential to ensure that AI technologies in Arabic language learning become more inclusive and sensitive to users' linguistic diversity.

REFERENCES

- Abidah, Sa'idatul, and Suci Ramadhanti Febriani. "Application of Clustering Method in Arabic Learning to Improve Speaking Skills for High School Levels." *Tanwir Arabiyyah: Arabic As Foreign Language Journal* 2, no. 2 (December 2022): 109–22. <https://doi.org/10.31869/aflj.v2i2.3456>.
- Alshehri, Awad, and Abdulrahman AlShabeb. "Exploring Attitudes, Identity, and Linguistic Variation among Arabic Speakers: Insights from Acoustic Landscapes." *International Journal of Arabic-English Studies*, ahead of print, November 5, 2023. <https://doi.org/10.33806/ijaes.v24i2.587>.
- an-Naqah, Mahmud Kamil. *Ta'lim al-Lughah al-'Arabiyyah li al-Nathiqin bi Lughat Ukhra*. Makkah: Jami'ah Umm al-Qura, 1985.
- Cervetti, Gina N., Tanya S. Wright, and HyeJin Hwang. "Conceptual Coherence, Comprehension, and Vocabulary Acquisition: A Knowledge Effect?" *Reading and Writing* 29, no. 4 (April 2016): 761–79. <https://doi.org/10.1007/s11145-016-9628-x>.
- Ellis, Rod. "Corrective Feedback and Teacher Development." *L2 Journal* 1, no. 1 (April 2009). <https://doi.org/10.5070/L2.V1I1.9054>.
- Farid, Edi Kurniawan, Mamluatun Ni'mah, and Nur Arifah. "The Teaching of The Speaking Skill Using The Audio-Oral Method for The Students of Islamic Senior High School of Zainul Hasan Genggong Probolinggo." *Mantiqutayr: Journal of Arabic Language* 1, no. 1 (January 2021): 1–10. <https://doi.org/10.25217/mantiquitayr.v1i1.1197>.
- Gazali, A., Nihayatur Rahmah, Roychan Yasin, Muhammad Ridwan, and Alya Raihana Sari. "Arabic for Specific Purposes in Islamic Higher Education: Systemic Learning Challenges among Non-Pesantren Students." *Alibbaa': Jurnal Pendidikan Bahasa Arab* 7, no. 1 (February 2026): 370–91. <https://doi.org/10.19105/ajpba.v7i1.23592>.
- Hanafiah, Waode, Muhammad Aswad, Harlinah Sahib, Abdul Hakim Yassi, and Movahede Sadat Mousavi. "The Impact of CALL on Vocabulary Learning, Speaking Skill, and Foreign Language Speaking Anxiety: The Case Study of Indonesian EFL Learners." *Education Research International* 2022 (January 2022): 1–13. <https://doi.org/10.1155/2022/5500077>.

- Ilmiani, Aulia Mustika, Nurul Wahdah, and Mahfuz Rizqi Mubarak. "The Application of Albert Bandura's Social Cognitive Theory: A Process in Learning Speaking Skill." *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban* 5, no. 2 (December 2021). <https://doi.org/10.15575/jpba.v5i2.12945>.
- Kashmiri, Hayat Ahmad. "Communication Challenges: Saudi EFL Speaking Skills and Strategies to Overcome Speaking Difficulties." *Arab World English Journal*, no. 267 (December 2020): 1–61. <https://doi.org/10.24093/awej/th.267>.
- Kosim, Nanang, Ade Arip Ardiansyah, Hilda Saripatul Hikmah, and Yusuf Ali Shaleh Atha. "The Use of The Task-Base Language Teaching (TBLT) Method to Improve Learning Outcomes of Arabic Language Skills." *Alibbaa': Jurnal Pendidikan Bahasa Arab* 5, no. 2 (July 2024): 144–65. <https://doi.org/10.19105/ajpba.v5i2.14804>.
- Kuddus, Khushboo. "Artificial Intelligence in Language Learning: Practices and Prospects." In *Advanced Analytics and Deep Learning Models*, 1–17. Wiley, 2022. <https://doi.org/10.1002/9781119792437.ch1>.
- Kundu, Arnab, and Tripti Bej. "AI in School EFL Learning: A Systematic Review of Impact Pathways for Engagement, Achievement, and Satisfaction." *Journal of Language and Education* 11, no. 4 (December 2025): 131–48. <https://doi.org/10.17323/jle.2025.22083>.
- Littlemore, Jeannette, and Graham Low. "Metaphoric Competence, Second Language Learning, and Communicative Language Ability." *Applied Linguistics* 27, no. 2 (June 2006): 268–94. <https://doi.org/10.1093/applin/aml004>.
- Mahmudah, Menik, Nurhanifansyah Nurhanifansyah, and Syarif Muhammad Syaheed bin Khalid. "Psycholinguistic Approaches to Enhancing Arabic Speaking Proficiency through Comic Strips." *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2 (November 2024): 805–26. <https://doi.org/10.29240/jba.v8i2.11349>.
- Miles, Matthew B., A. Michael Huberman, and Johnny Saldaña. *Qualitative Data Analysis: A Methods Sourcebook*. 3rd ed. Thousand Oaks, CA: SAGE Publications, 2014.
- Milankov, Vesela, Slavica Golubović, Tatjana Krstić, and Špela Golubović. "Phonological Awareness as the Foundation of Reading Acquisition in Students Reading in Transparent Orthography." *International Journal of Environmental Research*

- and Public Health* 18, no. 10 (May 2021): 5440.
<https://doi.org/10.3390/ijerph18105440>.
- Mokhtar, Mohamed Mohd Ieruan. "Lower Secondary Students' Arabic Speaking Anxiety: A Foreign Language Literacy Perspective." *International Journal of Education and Literacy Studies* 8, no. 4 (October 2020): 33. <https://doi.org/10.7575/aiac.ijels.v.8n.4p.33>.
- Novak, Gary. "History of the Behavior Analysis of Child Development." In *The Encyclopedia of Child and Adolescent Development*, 1–11. Wiley, 2020.
<https://doi.org/10.1002/9781119171492.wecad049>.
- Putri, Tarisha, and Sahkholid Nasution. "Gamified Digital Tools in Arabic Language Instruction: The Case of Froggy Jumps in an Indonesian Secondary Islamic School." *Alibbaa': Jurnal Pendidikan Bahasa Arab* 6, no. 2 (August 2025): 412–27.
<https://doi.org/10.19105/ajpba.v6i2.21379>.
- Saldaña, Johnny. *The Coding Manual for Qualitative Researchers*. 3rd ed. London: SAGE Publications, 2016.
- Sapawi, Mior Syazril Mohamed, and Nik Mohd Rahimi Nik Yusoff. "Integrating Technology into the Arabic Language Curriculum: A Systematic Review of Trends, Strategies and Cultural Dimensions." *Social Sciences & Humanities Open* 12 (2025): 101974. <https://doi.org/10.1016/j.ssaho.2025.101974>.
- Sedghi, Hamed, Soheyla Mohseny Nejad, Seyyed Adnan Eshkevari, Poursan Rezaei Choshali, and Hossein Talebzadeh. "Identifying the Speaking Proficiency Level of Arabic Learners in Accordance with the International Guidelines of ACTFL for Assessing Proficiency in Foreign Languages." *Research in Arabic Language* 12, no. 25 (January 2020): 1–22.
- Singh, Gundeep, Sahil Sharma, Vijay Kumar, Manjit Kaur, Mohammed Baz, and Mehedi Masud. "Spoken Language Identification Using Deep Learning." *Computational Intelligence and Neuroscience* 2021, no. 1 (January 2021).
<https://doi.org/10.1155/2021/5123671>.
- Soelistianto, Farida Arinie, Dony Andrasromo, Yusriati Yusriati, Mardiati Mardiati, and Aldi Bastiatul Fawait. "The Impact of Adaptive Learning Technology on Improving Students' Concept Understanding." *Journal of Computer Science Advancements* 2, no. 3 (July 2024): 158–75.
<https://doi.org/10.70177/jsca.v2i3.1176>.

- Thu'aimah, Rusydi Ahmad. *Ta'lim al-'Arabiyyah li Ghair al-Nathiqin Biha: Manahijuhu wa Asalibuhu*. Rabat: ISESCO, 1989.
- Vo, Thi Kim Anh, and Huong Nguyen. "Generative Artificial Intelligence and ChatGPT in Language Learning: EFL Students' Perceptions of Technology Acceptance." *Journal of University Teaching and Learning Practice* 21, no. 6 (April 2024). <https://doi.org/10.53761/fr1rkj58>.
- Wang, Yiyin, Tiancheng Zhang, Le Yao, and Paul Seedhouse. "A Scoping Review of Empirical Studies on Generative Artificial Intelligence in Language Education." *Innovation in Language Learning and Teaching*, June 5, 2025, 1–28. <https://doi.org/10.1080/17501229.2025.2509759>.
- Zaimah, Nely Rahmawati, Risti Kamila Wening Estu, Syarifatul Fitri Hidayah, Syamsul Hadi, and Aiden Button. "Harnessing Gemini for Arabic Mastery: Educators' and Learners' Views." *Alibbaa': Jurnal Pendidikan Bahasa Arab* 5, no. 2 (July 2024): 166–88. <https://doi.org/10.19105/ajpba.v5i2.14808>.
- Zamzami, Rifa'atus Shoolihah, Nurdyansyah Nurdyansyah, and Taufiq Churrahman. "Implementation of the Arabic Language Habituation Program in Improving Academic Achievement of Students in SMP Muhammadiyah 9 Tanggulangin Boarding School." *Proceedings of The ICECRS* 6 (June 2020). <https://doi.org/10.21070/icecrs2020379>.
- Zawacki-Richter, Olaf, Victoria I. Marín, Melissa Bond, and Franziska Gouverneur. "Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators?" *International Journal of Educational Technology in Higher Education* 16, no. 1 (December 2019): 39. <https://doi.org/10.1186/s41239-019-0171-0>.