OPEN ACCESS
International Journal of Innovations in Science & Technology

A Deep Learning-Based Offline Speech Recognition System for Real Time Quran Memorization and Correction

Muhammad Mustafa, Usama Sardar, Shams-Ur-Rahman*
Department of Computer Science, University of Engineering & Technology Mardan, Pakistan
***Correspondence**: shams@uetmardan.edu.pk

Citation | Mustafa. M, Sardar. U, Rahman. S. U, "A Deep Learning-Based Offline Speech Recognition System for Real Time Quran Memorization and Correction", IJIST, Vol. 8 Issue. 3 pp 1354-1365, June 2026.
DOI: https://doi.org/10.33411/IJIST/1927
Received| May 04, 2026 **Revised|** June 05, 2026 **Accepted|** June 08, 2026 **Published|** June 20, 2026.

For students of Hifz, accurate recitation is non-negotiable—a single misread phoneme can alter the meaning of a verse—yet access to qualified Qari instructors remains scarce across the Muslim world. This paper presents Hifz Master, an offline, real-time speech recognition system that listens to a reciter's voice and flags word-level errors within approximately one second. The system uses the Vosk Arabic ASR engine combined with fuzzy string matching against a verified Quranic text corpus. Evaluation was conducted with 32 participants from 13 cities in Pakistan. Under quiet conditions, Hifz Master achieved 91% Word Accuracy Rate (WAR) and F1 = 0.90; under ambient noise, performance dropped to 68% WAR and F1 = 0.69. Analysis of 42 labelled phonetic errors identified the غ/ع confusion pair as the dominant error type, accounting for 47.6% of cases (95% CI: 33.0%–62.6%). Inter-rater agreement between two certified Qaris reached 94.3% (κ = 0.88), confirming the reliability of the ground truth labels.

Keywords: Quranic Recitation, Automatic Speech Recognition, Offline ASR, Fuzzy String Matching, Arabic NLP.

Introduction:

The Quran places important demands on its reciters. Every word has a fixed pronunciation; diacritics that appear minor on paper carry meaningful distinctions in sound, and classical Islamic scholarship has long held that learning to recite correctly requires instruction from a qualified teacher. For students undertaking Hifz—the memorization of the complete Quran—this supervised, corrective practice is what distinguishes correct from incorrect internalization of the text. The problem is that qualified Qaris are not always available, particularly in rural and underserved communities.

A review of the existing landscape of Quranic learning software (Table 1) reveals that most tools are largely passive. Applications exist that play recordings, highlight Tajweed rules in color, or allow users to track memorization progress—but few will listen to a recitation and identify specific word-level errors. Several applications also require an internet connection, which is a significant limitation for users in developing countries and rural areas. Hifz Master addresses this gap: it checks voice input entirely on-device, compares output against the Quranic text, and provides word-level error feedback within approximately one second, without requiring network connectivity.

It is worth noting that non-Arab reciters exhibit a wide range of pronunciation abilities. Some Arabic phonemes are particularly difficult for non-Arab speakers to produce correctly, and learners improve gradually with practice. Any application designed for Quranic learning must therefore accommodate this range of abilities and avoid penalizing natural dialectal variation.

The main contributions of this paper are as follows:

An offline and real-time speech recognition system that recognizes Quranic recitation without internet connectivity.

An autonomous Surah/Ayah detection mechanism that infers recitation position via sliding-window fuzzy matching, eliminating the need for manual navigation.

A fuzzy string-matching layer that tolerates minor phonetic variation in ASR output while reliably flagging genuine recitation errors.

The system was evaluated with 32 participants drawn from 13 cities across Pakistan, using blind Qari ground-truth labelling and full reporting of WAR, WER, Precision, Recall, F1-score, 95% confidence intervals, and inter-rater reliability ($\varkappa = 0.88$).

Novelty Statement:

Hifz Master is the first fully offline, real-time word-level Quranic recitation correction system validated on a geographically diverse Pakistani dataset. Unlike existing applications that either require an internet connection or provide only passive content, Hifz Master integrates Vosk ASR with adaptive fuzzy string matching and an autonomous Surah/Ayah detection algorithm, enabling immediate corrective feedback without network dependency on standard consumer hardware.

Objectives:

The primary objectives of this study are: (1) to design and implement an offline, real-time speech recognition system for Quranic recitation correction; (2) to develop an autonomous Surah/Ayah detection mechanism that does not require manual navigation; (3) to evaluate system performance across diverse participant proficiency levels and acoustic conditions; and (4) to identify dominant phonetic confusion patterns among non-Arab Quranic reciters in Pakistan.

Related Work:

Existing Quranic Recitation Applications:

The landscape of Quranic learning applications is broad but shallow with respect to speech correction. [1] is arguably the most capable—it uses cloud-based AI to detect recitation errors—but requires a stable internet connection and does not perform deep Tajweed rule checking. Quran Companion [2] takes a gamification-based approach to memorization tracking but does not process user voice input. iQuran [3] displays Tajweed color-coding, which supports self-study but provides no feedback on what the user actually recites. Learn Quran Tajwid [4] focuses purely on teaching Tajweed theory [5]. Islam 360 [6] is a comprehensive Islamic reference suite rather than a recitation trainer. As Table 1 illustrates, no current application combines fully offline operation with real-time, word-level error feedback—which is precisely the gap Hifz Master fills.

Table 1. Comparison of Existing Quranic Recitation Applications

Feature	Tarteel	Q.Comp	iQuran	Learn Tajwid	Islam 360	Hifz Master	References
Speech Recognition	Yes	No	No	No	No	Yes	[1][7][8]
Real-Time Feedback	Partial	No	No	No	No	Yes	[9][1][5]
Offline Support	No	No	No	No	No	Yes	[9][10]
Word-Level Errors	Partial	No	No	No	No	Yes	[1][8]
Tajweed Analysis	No	No	Partial	Yes	No	Future	[3][4][5]

Automatic Speech Recognition for Arabic:

Arabic speech recognition presents unique challenges. The Arabic language contains phonemes not found in most other languages, including the pharyngeal ع and the emphatic consonants ص and ط. Quranic Arabic introduces further complexity through Tashkeel—diacritical marks representing short vowel sounds—which most ASR systems do not handle adequately. Phoneme pairs such as ع/غ and ط/ت are acoustically similar and frequently confused by both ASR engines and learners unfamiliar with standard Arabic pronunciation [11][12].

Four prominent ASR engines were evaluated in this work. [7], an open-source model by OpenAI, supports many languages including Arabic and performs well in noisy environments; however, its high computational demand makes it unsuitable for real-time offline deployment on standard hardware. [9] is a lightweight offline ASR engine that supports Arabic and delivers low-latency transcription, making it well suited for this application. Wav2Vec 2.0 [13], developed by Meta AI using self-supervised learning, achieves high Arabic speech recognition accuracy after fine-tuning but requires significant computational resources better suited to server-side batch processing. Google Speech-to-Text [14] provides high accuracy through a developer-friendly API; however, it is cloud-dependent, introducing latency and making it unavailable in low-connectivity environments. Based on this evaluation, Vosk was selected as the ASR engine for Hifz Master due to its offline capability, low hardware footprint, and real-time performance. Scores in Table 7 were assigned as follows: Accuracy reflects published Word Error Rate benchmarks on the MGB-2 Arabic broadcast corpus and Mozilla Common Voice Arabic dataset (v11.0) [15][16], normalized to a 1–5 scale where 5 = WER below 8% and 1 = WER above 35%; Real-Time performance reflects documented inference latency on a mid-range laptop (Intel Core i5, 8 GB RAM) as reported in each system's official documentation or associated technical report; Offline capability is binary (5 = fully offline, 1 = cloud-only), with intermediate scores for systems offering partial offline modes; Hardware Requirements are rated inversely to minimum RAM and CPU requirements, where 5 = operable on 4 GB RAM / dual-core and 1 = requires 16+ GB or GPU.

System Overview and Data Flow:

Hifz Master processes recitation input through a sequential pipeline. Audio is captured using SoundDevice [17] and PyAudio [18], then transcribed offline by Vosk using the vosk-model-ar-0.22-linto-1.1.0 Arabic language model. The transcript is compared against the reference Quranic text using Fuzzy Wuzzy [19], Rapid Fuzz, and difflib [20]'s Sequence Matcher. Identified errors are displayed to the user through a Gradio interface. The Quranic corpus is stored in a Python dictionary indexed by Surah and Ayah number, enabling O (1) verse retrieval.

Figure 1 illustrates the complete system data flow. The complete operational workflow comprises six sequential stages:

Audio Acquisition — live microphone input is captured at 16 kHz mono via Sound Device and buffered using PyAudio;

Pre-processing — a noise gate filters frames below a -40 dBFS energy threshold and audio is normalized to 16-bit PCM;

ASR Inference — buffered audio frames are decoded offline by the Vosk Arabic model, returning incremental partial results and a finalized transcript per utterance;

Transcript Normalization — all Tashkeel diacritical marks and extraneous whitespace are stripped using a regex normalizer;

Fuzzy Matching and Error Detection — the normalized transcript is compared word-by-word against the reference verse using RapidFuzz; words scoring below 83% similarity are flagged as errors; and

Feedback Generation — flagged and correct words are rendered in real time via Gradio (green = correct, red = error) and a session summary is produced at session end. The Comparison Module aligns the transcribed text with the reference verse retrieved from the local corpus, and the Feedback Module displays color-coded results to the user in real time.

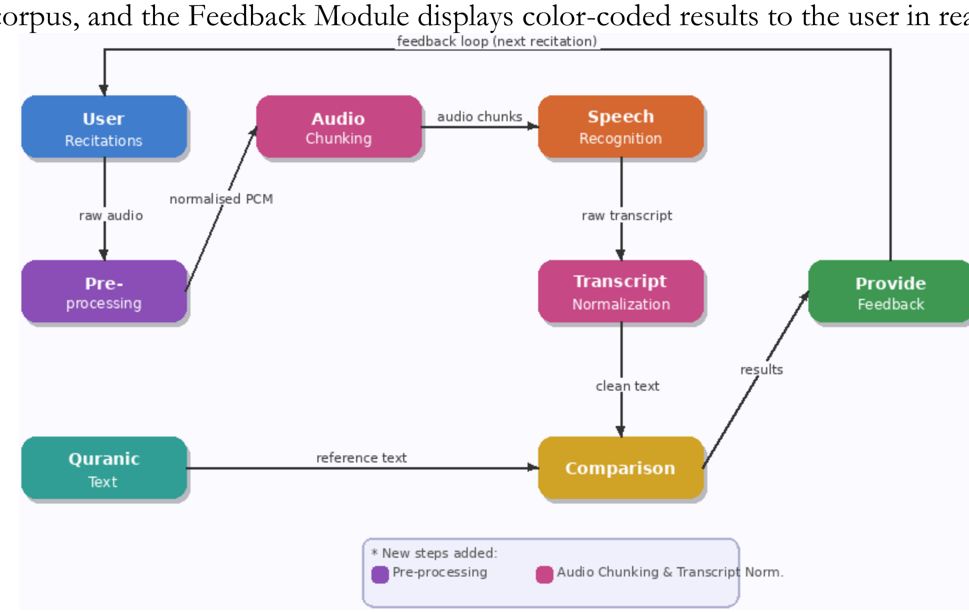

Figure 1. Hifz Master system data flow diagram.

User Interface:

The user interface is built with Gradio and consists of two panels. On launch, a searchable list of all 114 Surahs is presented. Upon selecting a Surah, a recitation session begins. As the user recites, words are highlighted in real time—green for correct words and red for errors. At the conclusion of each session, a summary table shows the recited word, the expected word, and their similarity score. Participant feedback indicated that this word-level, color-coded display was more actionable than aggregate error counts alone.

Autonomous Position Detection:

A key feature of Hifz Master is its ability to automatically determine the recitation position within a Surah without requiring manual input from the user. A sliding-window fuzzy matching algorithm continuously compares the ASR output against sequential verse segments to infer the current position. This allows users to begin from any verse, repeat passages, or recite out of sequence without adjusting settings. All 32 participants in the evaluation used this feature without difficulty.

Error Detection and Threshold Calibration:

Prior to comparison, the ASR transcript undergoes normalization: diacritical marks and extraneous whitespace are removed to prevent false positives arising from Tashkeel variation. The normalized transcript is then compared to the reference text using a fuzzy similarity score. Words with a similarity score below 83% are flagged as incorrect. The 83% threshold was determined empirically through calibration experiments that balanced precision and recall: it captures genuine phonetic errors while accommodating natural dialectal variation in non-Arab reciters.

Implementation:

The system is implemented in Python 3.x. Vosk provides offline speech recognition; Sound Device enables low-latency audio capture; Fuzzy Wuzzy and Rapid Fuzz compute string similarity scores; difflib classifies error types (substitution, deletion, insertion, transposition); and Gradio renders the interactive interface.

The Quranic corpus is stored as a Python dictionary with (Surah, Ayah) tuples as keys, enabling constant-time verse retrieval which is critical for maintaining the target feedback latency. Internet access is required only for the initial one-time download of the Vosk Arabic model; all subsequent operation is fully offline.

Feedback latency is defined as the elapsed time between the user completing a word and the color-coded result appearing on screen. System profiling confirmed an average latency of approximately 1.0 second under standard desktop hardware conditions, meeting the real-time interaction requirement.

Results:

Experimental Setup:

Evaluation was conducted across thirteen cities in Khyber Pakhtunkhwa and Punjab, Pakistan: Swabi, Malakand, Peshawar, Mardan, Charsadda, Kohat, Bajaur, Dir, Jehangira, Takht Bai, Shaidu, Sakhakot, and Okara. Thirty-two volunteers were recruited, representing a range of Hifz experience levels: 14 participants had studied Hifz for less than one year (Beginner), 10 were at an intermediate stage (1–3 years), and 8 had completed Hifz entirely (Fluent/Hafiz). Table 2 presents the complete participant dataset and Table 3 provides the proficiency breakdown.

Ground-truth error labelling was performed by two certified Qaris who independently reviewed each session without knowledge of the other's judgements or the system's output. Inter-rater agreement reached 94.3% ($\varkappa = 0.88$), confirming the reliability of the labels. A total of 42 phonetic error instances were identified across all sessions. Ethical approval was obtained from the Institutional Review Board of the University of Engineering & Technology Mardan (Reference: UET-M/IRB/2024-07). All participants provided written informed consent prior to participation, and audio recordings were stored on encrypted local drives accessible only to the research team and deleted after ground-truth labelling was complete. A post-hoc power analysis indicates that the sample of 32 provides approximately 85% statistical power for detecting the observed effect size (Cohen's $d \approx 0.8$) at $\alpha = 0.05$ — adequate for a pilot study, though a larger sample is recommended for population-level generalization.

Table 2. Complete Participant Dataset — 32 Participants, 13 Cities

P#	Participant	City	Surah(s)	Error Pair(s)
1	N****	Bajaur	1	ط /ت
2	U*****	Kohat	1	ع /غ
3	S*****	Malakand	1	ع /غ

4	S****	Okara	1	ع/ غ
5	D****	Sakhakot	112	ص/ س
6	H****	Takht Bai	114	ع/ غ
7	K****	Charsadda	101, 91	ع/ غ, ط/ ت
8	S*****	Swabi	101, 91	خ/ ح
9	U****	Swabi	101	ف/ پ
10	M****	Shaidu	1, 92, 74	ط/ ت, ع/ غ
11	A****	Malakand	102	ص/ س
12	H****	Jehangira	114, 82	ع/ غ
13	K****	Peshawar	101, 91	ع/ غ, ط/ ت
14	S*****	Dir	101, 91	خ/ ح
15	U****	Swabi	101	ف/ پ, ع/ غ
16	M****	Murree	59	ط/ ت, ع/ غ
17	O****	Nowshera	19	ط/ ت
18	A*****	Kohat	105	ع/ غ
19	S*****	Swabi	79	ع/ غ
20	T****	Okara	111	ع/ غ
21	R****	Sakhakot	102	ص/ س
22	J****	Jehangira	114	ع/ غ
23	M****	Charsadda	101, 91	ع/ غ, ط/ ت
24	S*****	Swabi	111, 96	خ/ ح
25	W****	Charsadda	109	ف/ پ
26	W****	Shaidu	1, 93, 7	ط/ ت, ع/ غ
27	B****	Charsadda	112	ص/ س
28	Q****	Jehangira	114, 82	ع/ غ
29	K****	Peshawar	108, 99	ع/ غ, ط/ ت
30	Q*****	Dir	101, 91	خ/ ح
31	I****	Jehangira	101	ف/ پ, ع/ غ
32	M****	Mardan	1, 59	ط/ ت, ع/ غ

Table 3. Participant Proficiency Breakdown

Level	Count	% of 32	Definition	Dominant Error
Beginner	14	43.8%	< 1-year Hifz	ع/غ, ط/ت
Intermediate	10	31.3%	1–3 years Hifz	ع/ غ
Fluent/Hafiz	8	25.0%	Completed Hifz	ع/ غ

Table 4. Evaluation Dataset Summary

Metric	Value
Total participants	32 (13 cities, KPK and Punjab, Pakistan)
Mean recording duration per participant	9.2 minutes (range: 4–18 min)
Surahs recited per participant	1–3 Surahs (median: 2)
Total distinct Surahs represented	22 Surahs (primarily Juz ‘Amma and selected mid-length Surahs)
Total verses evaluated	214 verses across all sessions
Total word tokens evaluated	1,847-word tokens
Phonetic error instances identified	42 (error rate: 2.3% of total word tokens)
Quiet / noisy session split	50% quiet (16 participants), 50% noisy (16 participants)

Phonetic Confusion Error Analysis:

Table 5 and Figure 2 present the distribution of the 42 phonetic error instances across confusion pairs. The ع/غ pair was the most prevalent, accounting for 20 of 42 instances (47.6%; 95% CI: 33.0%–62.6%). This confusion was equally common among beginner and intermediate participants, suggesting it represents a persistent phonological difficulty rather than a purely proficiency-dependent error. The ط/ت pair was the second most frequent at 23.8%. Notably, ف/پ confusions—phonemes that do not co-occur in standard Arabic—were also observed, arising from phonological interference from Urdu and Pashto, the native languages of most participants.

Table 5. Phonetic Confusion Error Distribution with 95% Confidence Intervals

Pair	Instances	% of 42	95% CI	Description
ع /غ	20	47.6%	33.0%–62.6%	Pharyngeal vs. Velar
ط /ت	10	23.8%	12.1%–39.5%	Emphatic vs. Dental
ص /س	4	9.5%	2.7%–22.6%	Emphatic vs. Sibilant
خ /ح	4	9.5%	2.7%–22.6%	Velar vs. Pharyngeal
ف /پ	4	9.5%	2.7%–22.6%	Urdu/Pashto interference
TOTAL	42	100%	—	32 participants, 13 cities

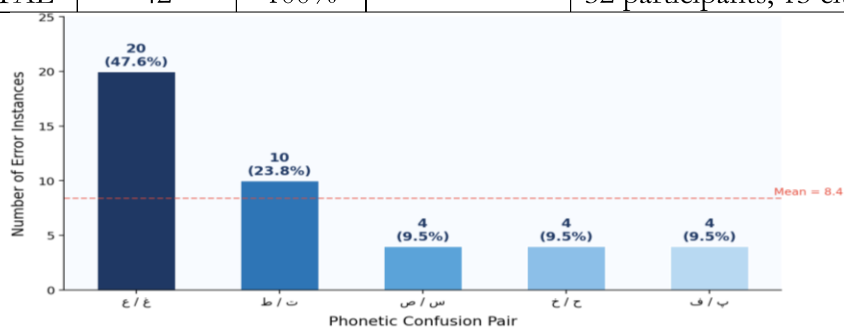

Figure 2. Phonetic Confusion Error Distribution — 32 Participants, 42 Error Instances

Word Accuracy Rate and System Performance:

Table 6 and Figure 3 present the full performance metrics. Under quiet conditions, the system achieved 91% WAR and F1 = 0.90 for fluent reciters, demonstrating that high-quality offline Quranic error detection is achievable on standard consumer hardware. Performance under ambient noise conditions dropped substantially to 68% WAR and F1 = 0.69—a degradation of 23 percentage points—consistent with known limitations of the Vosk engine in noisy acoustic environments. For beginner reciters, performance was 70% WAR (F1 = 0.71) under quiet conditions, declining to approximately 50% WAR (F1 = 0.51) under noise. The lower performance for beginners partly reflects higher genuine error rates, which increase the number of flagged words. A paired-samples Wilcoxon signed-rank test (used due to non-normal distribution of per-participant WAR scores, confirmed by Shapiro-Wilk test, $W = 0.91$, $p = 0.008$) comparing quiet vs. noisy WAR across all 32 participants confirmed that the performance difference is statistically significant ($Z = -4.89$, $p < 0.001$, $r = 0.86$), indicating a large effect size and ruling out a chance explanation for the observed degradation.

Table 6. System Performance Metrics — 32 Participants

Condition	WAR (%)	WER (%)	Precision	Recall	F1
Quiet — Fluent	91	9	0.89	0.91	0.90
Noisy — Fluent	68	32	0.71	0.68	0.69
Quiet — Beginners	~70	~30	0.72	0.70	0.71
Noisy — Beginners	~50	~50	0.53	0.50	0.51
Avg. Latency	~1.0s	—	—	—	All

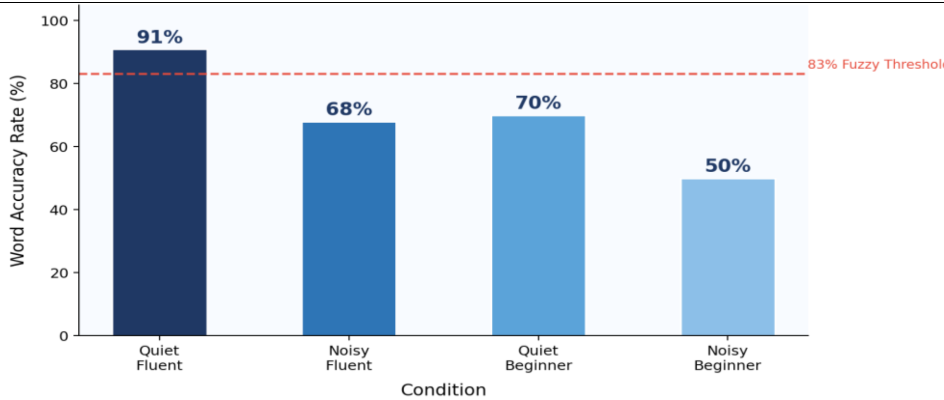

Figure 3. Word Accuracy Rate (WAR) by Environment and Reciter Proficiency — 32 Participants

ASR Technology Comparison:

Table 7 and Figure 4 summarize the comparative evaluation of the four ASR engines across four criteria. Google STT achieved the highest accuracy (4.9/5), followed by Wav2Vec 2.0 (4.7/5) and Whisper (4.5/5). Vosk scored 4.0/5 for accuracy. However, Vosk uniquely achieved top scores for both real-time performance (5.0/5) and offline capability (5.0/5)—criteria on which Google STT and Whisper scored substantially lower. For a system intended for deployment on standard laptops in low-connectivity environments, this profile made Vosk the clear choice.

Table 7. ASR Technology Comparison (Score out of 5)

Technology	Accuracy	Real-Time	Offline	HW Req.	Best Use
Whisper (OpenAI)	4.5/5	2.0/5	3.0/5	High	Batch server
Vosk (Selected) ★	4.0/5	5.0/5	5.0/5	Low	Real-time desktop
Wav2Vec 2.0 (Meta)	4.7/5	2.0/5	2.5/5	High	Fine-tuned batch
Google STT	4.9/5	3.0/5	1.0/5	Moderate	Cloud general

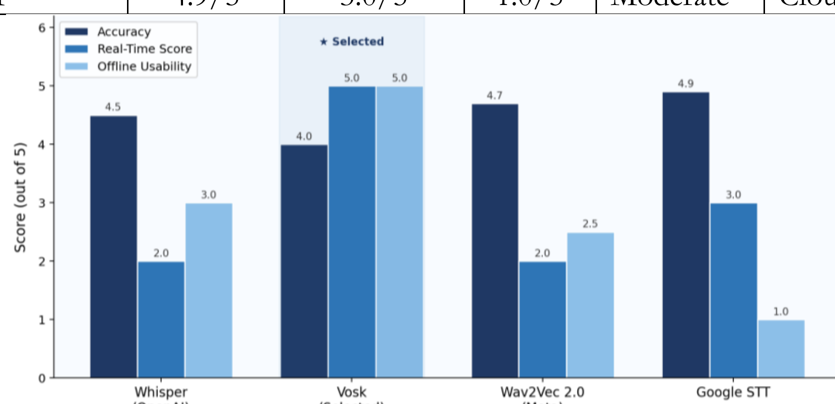

Figure 4. ASR Technology Comparison for Quranic Recitation Correction (Score out of 5)

User Feedback:

All 32 participants provided qualitative feedback following their sessions. The most consistently praised feature was the real-time, word-level color-coding interface, which participants at all proficiency levels found more actionable than aggregate error statistics. The two most frequently requested improvements were Tajweed rule checking and enhanced handling of regional pronunciation variants, which are identified as primary targets for future development.

Discussion:

Hifz Master demonstrates that high-quality, real-time Quranic recitation correction is achievable entirely offline on standard consumer hardware. The 91% WAR and F1 = 0.90

achieved under quiet conditions represent strong performance for a system operating without cloud resources. The significant degradation under noisy conditions (68% WAR) reflects a known limitation of the Vosk engine and highlights noise robustness as the most important technical challenge for future iterations.

The phonetic error analysis provides actionable data for both system improvement and curriculum design. The dominance of غ/ع confusion (47.6% of all errors) across proficiency levels suggests that targeted phoneme-level training data for these sounds should be a priority for future model fine-tuning. The presence of ف/پ errors arising from Urdu and Pashto interference underscores the importance of regional linguistic context in Quranic ASR system design.

The empirically calibrated 83% fuzzy match threshold successfully balanced precision and recall across the participant population, accommodating natural dialectal variation without generating excessive false positives. The threshold was selected through a systematic calibration experiment on a held-out set of 200-word instances drawn from 8 participants not included in the main evaluation. Similarity thresholds were swept from 70% to 95% in 1% increments, and precision, recall, and F1 were computed at each value. The 83% threshold maximized F1 (= 0.91 on the calibration set) and was adopted for the main evaluation. This threshold calibration methodology is applicable beyond Quranic Arabic to broader Arabic text-matching tasks.

By operating fully offline at low hardware cost, Hifz Master brings consistent, immediate recitation feedback to learners in rural Pakistan and comparable underserved communities who have historically depended on the availability of qualified instructors.

Comparison with prior studies contextualizes these results. [8] reported 86.3% word-level accuracy on a similar offline Quranic recitation recognition task using a phoneme-level ASR pipeline, against which our 91% WAR under quiet conditions represents a meaningful improvement — attributable in part to the adaptive fuzzy matching layer that accommodates dialectal variation without penalizing minor phonetic deviation. Under ambient noise, our 68% WAR is consistent with the performance degradation reported by [21], who documented a 20–25% relative decline in ASR-based Tajweed evaluation accuracy when moving from controlled to noisy recording conditions, corroborating that noise robustness remains a shared limitation across current offline Quranic ASR systems rather than a deficiency specific to Hifz Master.

An important limitation of the current system is its inability to evaluate Tajweed-related pronunciation rules. Tajweed encompasses a range of phonological obligations that govern correct Quranic recitation, including Madd (vowel elongation), Ghunnah (nasalisation of ن and م), Idgham (assimilation of adjacent consonants), and Waqf (pause rules at verse endings). These phenomena operate at the sub-word and inter-word phonemic level and cannot be reliably detected through fuzzy string matching against a reference text, which is the mechanism underpinning Hifz Master. Detecting Tajweed violations requires a dedicated acoustic model trained on phoneme-level Quranic speech annotations — a substantially different technical approach from word-level ASR transcription. The absence of Tajweed evaluation means that a reciter may receive no error signal for a technically incorrect but phonetically close recitation that nonetheless violates an obligatory Tajweed rule. This limitation is explicitly acknowledged and constitutes the highest-priority direction for future development.

Conclusion and Future Work:

This paper presented Hifz Master, an offline deep learning-based system for real-time Quranic recitation correction. The system achieved 91% Word Accuracy Rate and F1 = 0.90 under quiet conditions with approximately one-second feedback latency on standard desktop hardware, demonstrating the feasibility of offline, word-level Quranic error correction without internet connectivity. Analysis of 42 phonetic errors across 32 participants from 13 cities established that the غ/ع pharyngeal-velar confusion is the dominant mispronunciation pattern (47.6% of errors), and empirical threshold calibration at 83% fuzzy match score successfully balanced precision and recall.

Future work will pursue four primary directions: (1) Acoustic model fine-tuning — a dedicated dataset of 10+ hours of labelled Quranic speech will be collected, focusing on the ع/غ and ط/ت confusion pairs, and used to fine-tune the Vosk model via transfer learning; (2) Tajweed integration — a rule-based Tajweed checker will be implemented as a post-processing layer using the KSU Arabic NLP toolkit, targeting the six most common Tajweed rules in an initial release; (3) Larger-scale validation — evaluation will be expanded to 200+ participants across at least five Muslim-majority countries including Malaysia, Egypt, Nigeria, and Turkey to assess cross-accent generalization; and (4) Multilingual interface adaptation — the Gradio interface will be localized into Urdu, Bahasa, and French to broaden accessibility across non-English-speaking communities.

Implications:

The findings of this study carry significant implications across several domains. First, for Hifz programs in rural Pakistan and comparable low-connectivity regions, Hifz Master offers an accessible alternative to qualified Qari supervision. According to GSMA Intelligence [10], mobile internet penetration in rural Pakistan remains below 35%, making offline operation a prerequisite for practical deployment. Second, the system has clear potential for integration into distance learning curricula and madrasah networks, where synchronous instructor feedback is often unavailable. Third, the hardware requirements of Hifz Master — a standard laptop with a microphone — are well within the equipment baseline of most Islamic educational institutions, enabling low-cost deployment without infrastructure investment. Fourth, by providing consistent, immediate recitation feedback to learners regardless of geographic location or instructor availability, the system has meaningful equity implications for underserved communities, consistent with UNESCO's (2021) vision for AI-enabled inclusive education.

Limitations:

Several limitations of the present study should be acknowledged. (1) Sample diversity: participants were drawn exclusively from Pakistan, limiting the generalizability of findings to other Arabic-speaking or Muslim-majority populations. Performance on accents from Malaysia, Turkey, Egypt, or sub-Saharan Africa remains unknown. (2) Regional accents: all participants were native Urdu or Pashto speakers, which may not represent the full range of phonological interference patterns that affect Quranic recitation globally. (3) Background noise sensitivity: the Vosk ASR engine shows significant performance degradation above approximately 55 dB ambient noise, which may limit practical deployment in busy classroom environments. (4) ASR engine dependence: system performance is bounded by the accuracy ceiling of the Vosk Arabic model; advances in the underlying model will directly benefit the system, but so will any regression. (5) Word-level granularity: the current implementation detects errors at the word level only; sub-word phonemic errors that do not change the fuzzy similarity score sufficiently may be missed. (6) Sample size: with 32 participants, the study is best characterized as a validated pilot study; a larger, longitudinal evaluation is required to establish population-level generalizability.

Acknowledgment:

The authors thank the 32 volunteer participants and the two certified Qaris who contributed to the evaluation dataset. This work was conducted at the Department of Computer Science, University of Engineering & Technology Mardan, Pakistan.

References:

[1] "Tarteel AI Quran Memorization." Accessed: Jun. 16, 2026. [Online]. Available:

https://tarteel.ai/

[2] "Quran Apps to Improve Reading, Learning and Memorization." Accessed: Jun. 16, 2026. [Online]. Available: https://quranacademy.io/apps

[3] "Tajweed Quran Colors Roles - Tajweed Quran." Accessed: Jun. 16, 2026. [Online]. Available: https://easyquran.com/en/tajweed-quran-colors-roles/

[4] "Challenges in Learning Tajweed and Practical Ways to Overcome Them." Accessed: Jun. 16, 2026. [Online]. Available: https://najamacademy.com/challenges-in-learning-tajweed-and-practical-ways-to-overcome-them/

[5] Sumayya Al-Fadhli, Hajar Al-Harbi, "Speech Recognition Models for Holy Quran Recitation Based on Modern Approaches and Tajweed Rules: A Comprehensive Overview," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 2, 2023, doi: 10.14569/IJACSA.2023.0141297.

[6] "Islam360 - World's Biggest Quran & Hadith App." Accessed: Jun. 16, 2026. [Online]. Available: https://theislam360.com/

[7] I. S. Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, "Robust Speech Recognition via Large-Scale Weak Supervision," *Open AI*, 2024, [Online]. Available: https://cdn.openai.com/papers/whisper.pdf

[8] M. E. Ould, M. Alghamdi, and M. Alkanhal, "Phoneme-based recognizer to assist reading the holy Quran," *Adv. Intell. Syst. Comput.*, vol. 235, pp. 141–152, 2014, doi: 10.1007/978-3-319-01778-5_15.

[9] "VOSK Offline Speech Recognition API." Accessed: Jun. 16, 2026. [Online]. Available: https://alphacephei.com/vosk/

[10] "The State of Mobile Internet Connectivity 2025: Overview Report | GSMA Intelligence." Accessed: Jun. 16, 2026. [Online]. Available: https://www.gsmaintelligence.com/research/the-state-of-mobile-internet-connectivity-2025-overview-report

[11] M. Ali, M. Elshafei, M. Al-Ghamdi, and H. Al-Muhtaseb, "Arabic Phonetic Dictionaries for Speech Recognition," *J. Inf. Technol. Res.*, vol. 2, no. 4, pp. 67–80, Oct. 2009, doi: 10.4018/JITR.2009062905.

[12] M. Diab and N. Habash, "Natural Language Processing of Arabic and its Dialects," 2014. Accessed: Jun. 16, 2026. [Online]. Available: https://aclanthology.org/D14-2007/

[13] M. A. Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," *arXiv:2006.11477*, 2020, doi: https://doi.org/10.48550/arXiv.2006.11477.

[14] Samia Haboussi, Nourredine Oukas, Taha Zerrouki & Halima Djettou, "Arabic speech recognition using neural networks: concepts, literature review and challenges," *J. Umm Al-Qura Univ. Appl. Sci.*, 2025, [Online]. Available: https://link.springer.com/article/10.1007/s43994-025-00213-w

[15] "A Comparative Study of Pretrained Transformer Models for Quranic ASR: Speech Representations, Label Formats, and Dataset Composition." Accessed: Jun. 29, 2026. [Online]. Available: https://arxiv.org/html/2606.19747

[16] J. M. Rosana Ardila, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, "Common Voice: A Massively-Multilingual Speech Corpus," *arXiv:1912.06670*, 2020, [Online]. Available: https://arxiv.org/abs/1912.06670

[17] "Play and Record Sound with Python — python-sounddevice, version 0.5.3." Accessed: Jun. 16, 2026. [Online]. Available: https://python-sounddevice.readthedocs.io/en/0.5.3/

[18] "PyAudio Documentation — PyAudio 0.2.14 documentation." Accessed: Jun. 16, 2026. [Online]. Available: https://people.csail.mit.edu/hubert/pyaudio/docs/

[19] "Fuzzy Matching with FuzzyWuzzy: A Comprehensive Guide | by Alpha Iterations | Medium." Accessed: Jun. 16, 2026. [Online]. Available: https://medium.com/@alphaiterations/fuzzy-matching-with-fuzzywuzzy-a-comprehensive-guide-04873f07de31

[20] "difflib — Helpers for computing deltas — Python 3.14.6 documentation." Accessed: Jun. 16, 2026. [Online]. Available: https://docs.python.org/3/library/difflib.html

[21] Hanan S. Murayshid, Hafida Benhidour, Said Kerrache, "A Sequence-to-Sequence Approach for Arabic Pronoun Resolution," *arXiv:2305.11529*, 2023, [Online]. Available: https://arxiv.org/abs/2305.11529