International Journal of Advanced Research in Computer Science and Engineering (IJARCSE)
ISSN (Online): 3071-0154
Volume-2 Issue-3 || Jul-Sep 2026 || PP. 1 – 7 www.ijarcse.org

Usability Analysis of Voice-Based Interfaces in Mobile Applications

Dr. Akankasha Kathuria

Assistant Professor

Amity University Rajasthan, Jaipur

Pin-303002

ORCID Id: https://orcid.org/0000-0002-7288-2490

www.ijarcse.org || **Vol. 2 No. 3 (2026): September Issue**

Date of Submission: 07-08-2026	**Date of Acceptance:** 25-08-2026	**Date of Publication:** 05-09-2026

Abstract— Voice-based interaction has moved from a peripheral accessibility feature to a mainstream input modality in mobile applications, yet its usability is still evaluated largely with instruments designed for graphical interfaces. This study reports a controlled, within-subjects usability evaluation comparing voice-based and touch-based interaction across six representative mobile tasks performed by 120 participants under two acoustic conditions (quiet and ambient noise). Effectiveness, efficiency, error recovery, perceived usability (System Usability Scale) and perceived workload (NASA-TLX) were measured. Voice interaction achieved a mean task completion rate of 80.0% against 95.2% for touch, but was markedly faster in quiet conditions (24.6 s vs 31.8 s per task). A significant modality × noise interaction indicated that the efficiency advantage of voice collapsed under ambient noise, where completion time rose to 33.8 s and completion fell to 71.3%. Mean SUS for the voice condition was 68.4 compared with 79.6 for touch, and perceived workload was significantly higher for voice. Analysis of 1,152 repair episodes showed that recognition errors accounted for 41.1% of breakdowns, while discoverability and feedback ambiguity together accounted for a further 27.5%. The findings indicate that the usability ceiling of mobile voice interfaces is determined less by raw recognition accuracy than by command discoverability, confirmation design and environmental robustness, and that hybrid multimodal designs offer the most defensible interaction strategy.

Keywords— voice user interface, mobile usability, System Usability Scale, NASA-TLX, speech recognition, multimodal interaction, human-computer interaction

Introduction

Speech has become one of the most widely deployed non-visual input channels in mobile computing. Contemporary mobile operating systems embed voice assistants at the system level, and application developers increasingly expose voice shortcuts for search, navigation, messaging and transactional workflows. The commercial argument is straightforward: speech is hands-free, eyes-free, faster than typing for many inputs, and lowers the interaction cost of deeply nested menu structures on small screens.

The usability argument is considerably less settled. Unlike graphical interfaces, voice interfaces are invisible. They offer no persistent affordances, no menu from which the command vocabulary can be inferred, and no spatial memory to support learning. Feedback is transient and serial rather than persistent and parallel. Errors are frequently silent: a misrecognised utterance may produce a plausible but incorrect action rather

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

than an obvious failure. Furthermore, mobile use is situated, and the acoustic environment of a train platform or a busy street differs fundamentally from the laboratory conditions under which most speech systems are benchmarked.

These characteristics raise a measurement problem. Standard usability instruments assume the user can see the system, revisit prior states and rely on recognition rather than recall. Applying them uncritically to voice may either overstate or understate the real interaction cost. This study therefore evaluates voice-based mobile interaction against a matched touch baseline using converging objective and subjective measures, and examines how acoustic context moderates the outcome.



Literature Review

Early formative work on conversational systems established that the dominant usability problem is expectation mismatch rather than technical failure. Luger and Sellen (2016) interviewed users of commercial conversational agents and found a persistent gulf between what users assumed the system understood and what it actually processed, with users routinely underestimating the narrowness of the supported command space. Cowan et al. (2017) reported converging evidence from infrequent users, who described interaction as effortful, socially awkward in public settings, and constrained by uncertainty about system capability.

The discoverability problem has been isolated repeatedly as the central structural weakness of voice interfaces. Corbett and Weber (2016) examined a mobile voice interface designed for accessibility and identified command discovery and learning as the principal barriers to sustained use, proposing design implications centred on progressive disclosure of the command vocabulary. Furqan, Myers and Zhu (2017) extended this with adaptive discovery tools, showing that learnability improves when the system surfaces candidate commands contextually rather than requiring users to recall them. Myers et al. (2018) provided the most granular account, classifying obstacles encountered by users of an unfamiliar multimodal voice interface and analysing the tactics used to recover from them. Their key observation was that natural-language processing errors were the most frequent obstacle class, but were not the most frustrating; obstacles relating to unclear system state and unsupported intents produced disproportionate confusion, and users tended to guess new phrasings rather than consult available visual aids.

Ethnographic work situates these findings in real use. Porcheron et al. (2018) analysed voice interface use in domestic settings and argued that interaction is better understood as embedded in ongoing social activity than as isolated dialogue turns, which has direct implications for how success is defined and measured. Murad et al. (2019) reviewed the state of speech interaction guidelines and concluded that existing HCI design heuristics translate only partially to speech, calling for guidance grounded in the specific properties of the modality.

A parallel literature addresses measurement. The System Usability Scale (Brooke, 1996) remains the most widely used perceived-usability instrument, with normative interpretation established by Bangor, Kortum and Miller (2008) and a consolidated methodological review by Lewis (2018). Its applicability to speech systems has been questioned: Ghosh et al. (2018) assessed the utility of SUS for voice-based interfaces and found that several items map awkwardly onto non-visual interaction. Zwakman, Pal and Arpnikanondt (2021) tested SUS with 62 voice-assistant users and, through exploratory factor analysis, reported structural weaknesses that motivated a dedicated Voice Usability Scale. Kocaballi, Laranjo and Coiera (2019) compared six questionnaires used for conversational systems and concluded that no single instrument covers the relevant experience dimensions adequately, recommending combined administration. Domain-specific instruments such as SASSI (Hone and Graham, 2000) remain relevant for capturing speech-specific constructs including perceived accuracy, habitability and annoyance.

Finally, equity of recognition performance is an established concern. Koenecke et al. (2020) documented substantial disparities in commercial automatic speech recognition error rates across speaker groups, demonstrating that recognition quality is not uniformly distributed across the user population. Pal et al. (2020) similarly found that adoption of voice-based smart products is shaped by perceived reliability and prior experience rather than novelty alone. Taken together, the literature identifies four recurring determinants of voice

OPEN ACCESS
@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

usability — recognition reliability, command discoverability, feedback and confirmation design, and environmental or speaker variability — but comparatively few studies quantify their joint effect against a matched graphical baseline in mobile conditions. This study addresses that gap.

Research Objectives and Hypotheses

The study pursued four objectives: to quantify the effectiveness and efficiency of voice interaction relative to touch interaction for equivalent mobile tasks; to determine whether ambient noise moderates that relationship; to characterise the distribution of interaction breakdowns in voice mode; and to assess perceived usability and workload using established instruments.

Three hypotheses were tested. H1: task completion rate is significantly lower for voice interaction than for touch interaction. H2: there is a significant interaction between modality and acoustic condition for task completion time, such that the efficiency advantage of voice is present in quiet conditions and absent under noise. H3: perceived usability is significantly lower, and perceived workload significantly higher, for voice interaction than for touch interaction.

Methodology

A counterbalanced within-subjects experimental design was used, with two independent variables: interaction modality (voice, touch) and acoustic condition (quiet, ambient noise). Every participant completed all six tasks in both modalities and both acoustic conditions, producing 24 trials per participant and 2,880 trials in total, of which 1,440 were voice trials.

One hundred and twenty participants were recruited through purposive sampling from university and workplace populations. Participants were aged 18–62 (M = 31.4, SD = 9.8); 62 identified as male and 58 as female. All were regular smartphone users with at least two years of experience and self-reported conversational or higher proficiency in the language of interaction. Participant characteristics are summarised in Table 1.

Table 1. Participant characteristics (N = 120)

Characteristic	Category	n	%
Age	18–25	33	27.5
	26–35	46	38.3
	36–45	27	22.5
	46 and above	14	11.7
Prior voice assistant use	Daily	34	28.3
	Weekly	42	35.0
	Monthly	26	21.7
	Rarely or never	18	15.0
Device platform	Android	81	67.5
	iOS	39	32.5

A single test application was instrumented to expose identical functionality through both a voice interface and a conventional touch interface, eliminating confounds arising from feature asymmetry between commercial apps. Six tasks were drawn from three common mobile domains: navigation (set a destination; add an en-route stop), messaging (send a message to a named contact; dictate and send a twenty-word reply), and commerce (search for a product applying two filters; check the status of a recent order).

The quiet condition was conducted in a laboratory at approximately 35 dB(A). The noise condition used calibrated recorded street and transit ambience presented at approximately 65 dB(A) at the participant's position. Task order was randomised and modality order counterbalanced using a Latin square to control learning effects. A five-minute familiarisation session preceded data collection, and a three-minute washout separated modality blocks.

Four dependent measures were collected. Task completion was scored binarily by two independent observers against predefined success criteria, with disagreements resolved by session video review; inter-rater agreement was high (Cohen's $\kappa = 0.89$). Task completion time was logged automatically from task initiation to goal state. Repair attempts were counted as any user-initiated retry, rephrasing, correction or manual fallback within a trial. After each modality block, participants completed the System Usability Scale and the NASA Task Load Index; voice-specific breakdowns were additionally coded post hoc from session recordings using a five-category taxonomy derived from prior work on voice interaction obstacles.

Analysis used two-way repeated-measures ANOVA for completion time and completion rate, paired-samples t-tests for SUS and NASA-TLX comparisons, independent-samples t-tests for experience subgroups, and Pearson correlation for associations between objective and subjective measures. Internal consistency of the SUS was acceptable in both conditions (Cronbach's $\alpha = 0.86$ voice, 0.84 touch). Sphericity assumptions were satisfied for all two-level factors, and the significance threshold was set at $\alpha = 0.05$. Participation was voluntary and informed consent was obtained; recordings were anonymised before coding.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

International Journal of Advanced Research in Computer Science and Engineering (IJARCSE)
ISSN (Online): 3071-0154
Volume-2 Issue-3 || Jul-Sep 2026 || PP. 1 – 7 www.ijarcse.org

Results

Aggregate performance differed substantially by modality. Voice interaction produced a mean task completion rate of 80.0% against 95.2% for touch, while mean completion time was lower for voice (29.2 s) than for touch (32.3 s). Task-level results are presented in Table 2.

Table 2. Task-level performance by modality (N = 120, means across acoustic conditions)

Task	Voice completion (%)	Voice time (s)	Touch completion (%)	Touch time (s)
Set navigation destination	89.6	22.4	97.1	30.5
Add en-route stop	78.3	30.1	95.0	34.2
Send message to named contact	86.7	24.9	96.7	27.8
Dictate twenty-word reply	82.5	28.6	94.2	41.3
Product search with two filters	71.3	33.7	94.6	28.4
Check order status	71.7	35.5	93.8	31.6
Overall	**80.0**	**29.2**	**95.2**	**32.3**

The pattern is directional rather than uniform. Voice outperformed touch on time for the four tasks dominated by text entry or single-intent commands, most strongly for dictated replies (28.6 s vs 41.3 s, a reduction of 30.8%). Touch outperformed voice on the two tasks requiring parameter refinement, where filters had to be specified and inspected.

Acoustic condition moderated these outcomes sharply, as shown in Table 3.

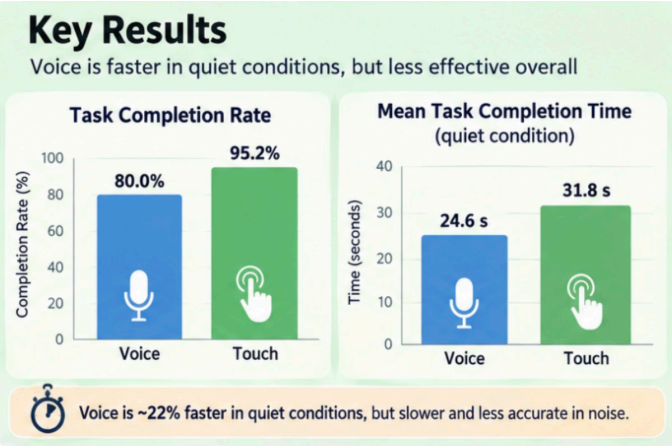
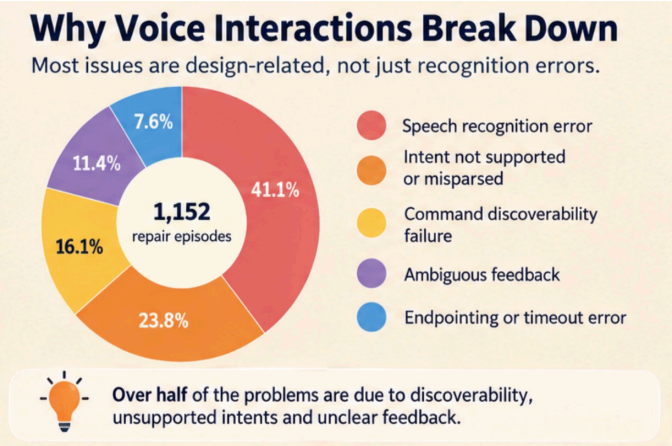


Table 3. Performance by modality and acoustic condition

Measure	Voice, quiet	Voice, noise	Touch, quiet	Touch, noise
Task completion rate (%)	88.6	71.3	95.8	94.6
Mean completion time (s)	24.6 (SD 7.9)	33.8 (SD 11.2)	31.8 (SD 8.4)	32.8 (SD 8.9)
Repair attempts per trial	0.42	1.18	0.11	0.14



Two-way repeated-measures ANOVA on completion rate returned significant main effects of modality, $F(1, 119) = 112.40$, $p < .001$, $\eta p^2 = .486$, and acoustic condition, $F(1, 119) = 64.70$, $p < .001$, $\eta p^2 = .352$, together with a significant interaction, $F(1, 119) = 51.90$, $p < .001$, $\eta p^2 = .304$. H1 was supported. For completion time, the main effect of modality

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

was significant, $F(1, 119) = 9.87$, $p = .002$, $\eta p^2 = .077$, as was acoustic condition, $F(1, 119) = 46.35$, $p < .001$, $\eta p^2 = .280$, with a strong interaction, $F(1, 119) = 38.72$, $p < .001$, $\eta p^2 = .245$. Simple-effects analysis confirmed that voice was significantly faster than touch in quiet conditions (mean difference 7.2 s, $p < .001$) but not in noise (mean difference 1.0 s, $p = .39$), supporting H2. Touch performance was effectively unaffected by noise.

Coding of the 1,152 repair episodes observed in voice trials produced the distribution shown in Table 4.

Table 4. Distribution of voice interaction breakdowns (1,152 repair episodes)

Breakdown category	Episodes	% of total
Speech recognition error	474	41.1
Intent not supported or misparsed	274	23.8
Command discoverability failure	185	16.1
Ambiguous feedback or confirmation	131	11.4
Endpointing or timeout error	88	7.6

Recognition errors dominated numerically but accounted for less than half of all breakdowns. The three categories relating to interface design rather than signal processing — unsupported intent, discoverability and feedback ambiguity — jointly accounted for 51.3% of episodes.

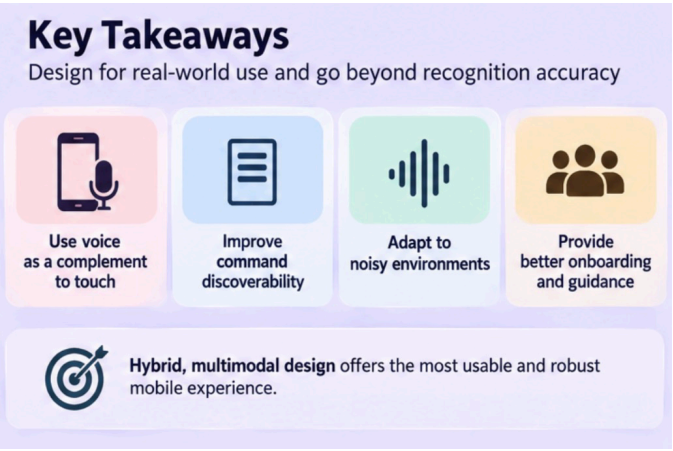
Subjective measures aligned with the objective pattern. Mean SUS was 68.4 (SD 12.6) for voice and 79.6 (SD 10.2) for touch, a significant difference, $t(119) = 8.14$, $p < .001$, $d = 0.74$. Mean NASA-TLX workload was 41.7 (SD 13.5) for voice and 33.2 (SD 11.8) for touch, $t(119) = 6.02$, $p < .001$, $d = 0.55$, with the mental demand and frustration subscales contributing most of the gap. H3 was supported. Prior experience mattered: frequent users (daily or weekly, $n = 76$) rated voice usability at 72.0 while infrequent users ($n = 44$) rated it at 62.2, $t(118) = 4.02$, $p < .001$. Within the voice condition, SUS correlated negatively with repair attempts, $r = -0.63$, $p < .001$, and positively with completion rate, $r = 0.58$, $p < .001$.

Discussion

The results describe a modality with a genuine but conditional advantage. Where the task reduces to a single expressible intent or to free-text production, voice is measurably more efficient than touch, and the dictation result in particular quantifies a benefit that small-screen keyboards cannot match. Where the task requires iterative refinement of parameters and visual verification of intermediate state, the advantage disappears and reverses. This aligns with the observation of Myers et al. (2018) that users cope poorly when the acceptable utterance structure is uncertain, and with Corbett and Weber's (2016) framing of discoverability as the governing constraint.

The most consequential finding is the modality × noise interaction. A voice interface benchmarked only in quiet conditions would appear to deliver a 22.6% efficiency gain over touch; the same interface in realistic ambient noise delivers no efficiency gain at all and loses 17.3 percentage points of effectiveness. Because mobile use is by definition situated, laboratory-only evaluation systematically overstates voice usability. Environmental robustness should therefore be treated as a usability variable rather than an engineering footnote.

The breakdown taxonomy carries a similar implication. Improving recognition accuracy addresses at most 41.1% of observed repair episodes. The majority of breakdowns originate in interface design decisions — the absence of a visible command vocabulary, silent handling of unsupported intents, and confirmations that do not make system state unambiguous. These are tractable through design rather than through model improvement, which supports the argument advanced by Murad et al. (2019) that speech interaction requires its own design guidance rather than inherited graphical heuristics.



The measurement picture is more ambiguous. A voice SUS of 68.4 sits close to the conventional mid-point of the scale, which would ordinarily be read as marginal acceptability. Yet the strong negative correlation between SUS and repair attempts suggests the instrument is tracking something real, while the concerns raised by Ghosh et al. (2018) and Zwakman et al. (2021) about item fit remain plausible. The convergence between SUS, NASA-TLX and objective measures observed

OPEN ACCESS
@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

here supports the recommendation of Kocaballi et al. (2019) that multiple instruments be administered rather than any single score being treated as definitive.

The experience effect is also notable. The 9.8-point SUS gap between frequent and infrequent users reproduces, in quantitative form, the expectation mismatch described by Luger and Sellen (2016): infrequent users have not yet internalised the boundaries of the command space and therefore encounter more breakdowns per task.

Design Implications

Four implications follow directly from the data. First, voice should be positioned as a complement to touch rather than a replacement, with seamless handoff to the graphical interface at the point of breakdown; the 11.4% of episodes attributable to confirmation ambiguity in particular argue for persistent visual echo of recognised intent. Second, command vocabularies should be surfaced contextually and progressively rather than left to recall. Third, interfaces should detect degraded acoustic conditions and adapt — by raising confirmation thresholds, shortening expected utterances, or proactively offering the touch path — rather than failing silently. Fourth, onboarding should be treated as a usability intervention, given the measured performance gap between experienced and inexperienced users.

Limitations

Several limitations qualify these findings. The study used a single instrumented test application, which improves internal validity but limits generalisation to the behaviour of specific commercial assistants. The noise manipulation was controlled and reproducible but does not capture the social inhibition associated with speaking aloud in genuinely public settings. Participants were predominantly urban, educated smartphone users, and the sample did not permit systematic analysis of accent or dialect variation — a significant omission given the disparities documented by Koenecke et al. (2020). Finally, the design measured first-encounter and short-term performance; longitudinal study would be required to determine how far discoverability failures attenuate with sustained use.

Conclusion

This study quantified the usability of voice-based interaction in mobile applications against a matched touch baseline across 2,880 task trials. Voice interaction was significantly less effective overall (80.0% vs 95.2% completion) yet significantly more efficient under quiet conditions, with the efficiency advantage eliminated by ambient noise. Perceived usability and workload favoured touch, and breakdown analysis showed that a majority of interaction failures stemmed from design-level issues of discoverability, intent coverage and feedback rather than from recognition error alone. The practical conclusion is that mobile voice interfaces should be designed and evaluated as one channel within a multimodal system, tested under realistic acoustic conditions, and instrumented to recover gracefully when speech fails. Future work should extend this design longitudinally, incorporate speaker diversity as an explicit factor, and validate voice-specific usability instruments against the objective measures reported here.

References

1. Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594.
2. Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189–194). Taylor & Francis.
3. Corbett, E., & Weber, A. (2016). What can I say? Addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)* (pp. 72–82). ACM.
4. Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., & Bandeira, N. (2017). "What can I help you with?": Infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)* (pp. 1–12). ACM.
5. Furqan, A., Myers, C., & Zhu, J. (2017). Learnability through adaptive discovery tools in voice user interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)* (pp. 1617–1623). ACM.
6. Ghosh, D., Foong, P. S., Zhang, S., & Zhao, S. (2018). Assessing the utility of the System Usability Scale for evaluating voice-based user interfaces. In *Proceedings of the Sixth International Symposium of Chinese CHI (ChineseCHI '18)* (pp. 11–15). ACM.
7. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). North-Holland.
8. Hone, K. S., & Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering*, 6(3–4), 287–303.
9. Kocaballi, A. B., Laranjo, L., & Coiera, E. (2019). Understanding and measuring user experience in conversational interfaces. *Interacting with Computers*, 31(2), 192–207.
10. Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14), 7684–7689.
11. Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. *International Journal of Human-Computer Interaction*, 34(7), 577–590.
12. Luger, E., & Sellen, A. (2016). "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)* (pp. 5286–5297). ACM.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org

13. Murad, C., Munteanu, C., Cowan, B. R., & Clark, L. (2019). Revolution or evolution? Speech interaction and HCI design guidelines. *IEEE Pervasive Computing*, 18(2), 33–45.
14. Myers, C., Furqan, A., Nebolsky, J., Caro, K., & Zhu, J. (2018). Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)* (pp. 1–7). ACM.
15. Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)* (pp. 249–256). ACM.
16. Pal, D., Arpnikanondt, C., Funilkul, S., & Chutimaskul, W. (2020). The adoption analysis of voice-based smart IoT products. *IEEE Internet of Things Journal*, 7(11), 10852–10867.
17. Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)* (pp. 1–12). ACM.
18. Zwakman, D. S., Pal, D., & Arpnikanondt, C. (2021). Usability evaluation of artificial intelligence-based voice assistants: The case of Amazon Alexa. *SN Computer Science*, 2(1), Article 28.

OPEN ACCESS
@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.ijarcse.org