A Comparative Usability Study of Voice Input and Touch Interaction for Completing Small Screen Mobile Tasks

Pranav Koirala, Diksha Sharma, Ishan KC
Department of Computer Science and Engineering, Kathmandu University
COMP 341, Human Computer Interaction Research Paper
Under the supervision of Dr. Sushil Shrestha

Abstract—Touch remains the default input method on smartphones, but voice input is increasingly offered as an alternative for moments when touch is inconvenient. It is not yet clear whether voice input is genuinely preferable in those moments, or whether it is simply appealing in theory. This study surveyed 64 smartphone users to find out which method, voice or touch, is more usable for short everyday mobile tasks, and specifically where voice offers a real advantage rather than a marginal one. Respondents judged touch faster (53.1%), more accurate (82.8%), and easier overall (57.8%), and almost unanimously preferred it for sensitive information (95.3%). That preference reversed, though, in the two scenarios where a hand was unavailable: 62.5% preferred voice when one hand was occupied, and 64.1% preferred voice during a simulated temporary hand injury. Accuracy, not speed, was the factor respondents cared about most (53.1%), which helps explain why touch's reliability edge outweighs voice's convenience appeal in most contexts. These findings point to the specific, recurring problems voice input can actually solve for mobile users, including situational or temporary loss of hand use, eyes busy or distracted contexts, and accessibility for users with limited manual dexterity. We propose concrete design changes to address the errors and privacy concerns that currently hold voice input back, and close with a plan for a follow up moderated usability evaluation.

Index Terms—human computer interaction, voice input, touch interaction, mobile usability, speech interfaces, accessibility, situational impairment

I. INTRODUCTION

Mobile users frequently need to complete small, well defined tasks such as searching for a place, filling out a short form, selecting a menu option, or sending a quick message, often in situations where touch is not the most convenient input method. A user might be carrying a bag in one hand, walking through a crowd, or temporarily unable to use a hand because of an injury. Voice input is often proposed as a solution in these moments, since it does not require visual attention or manual dexterity. At the same time, voice recognition is imperfect. Background noise, accent variation, and misrecognition can introduce errors and repeated attempts, and speaking aloud raises privacy concerns in public or sensitive settings.

Despite how widely available voice assistants have become, it is still unclear how often, and under what specific conditions, voice input is actually the better choice compared to touch for short mobile tasks, and what concrete problems it is best suited to solve rather than merely supplement. Much of the existing literature evaluates voice interfaces within a single application rather than comparing the two methods across a broad set of everyday task types and contexts. This study addresses that gap by asking users to compare the two methods across ten realistic task and context scenarios, then uses the resulting pattern to identify where voice input offers genuine, rather than just theoretical, value.

This paper investigates four research questions:

RQ1: Which input method, voice or touch, is faster for short mobile tasks?

RQ2: Which method produces fewer errors?

RQ3: Which method feels easier and more satisfying to use?

RQ4: Does voice interaction help more when users are multitasking, distracted, or unable to use both hands?

Answering these questions matters directly for Human Computer Interaction, since it touches on core HCI concerns such as input device selection, speech and auditory interface design, menu and form design, error handling and feedback, and situational accessibility. The practical payoff of this research is a set of evidence-based recommendations for when a mobile interface should offer voice input as a genuine, problem solving alternative to touch, and how that interface should be designed to overcome the reliability and privacy concerns users currently report.

II. RELATED WORK

Nielsen's usability heuristics, particularly efficiency of use and error prevention, provide the baseline criteria against which any alternative input method should be judged before it is considered a replacement for an established one [1]. Norman's account of how well-designed systems communicate their own state and afford correct action is equally relevant to voice interfaces, which must substitute spoken feedback for the visual affordances users rely on with touch [2].

Prior comparative work is mixed. Wallén conducted a within subjects' comparison of voice and touch interaction for a smartphone radio and podcast application and found that voice was often faster and rated higher than touch, but only once users already knew what the system could do. Users also reported reluctance to use voice commands in public out of concern for how they would appear to others [3]. This social acceptability concern recurs across the voice interface literature and is a central reason touch remains preferred in shared or public spaces even when voice would be technically faster.

Work on encumbrance shows that the relative usability of an input method depends heavily on physical context rather than on the method's intrinsic merit. Techniques that work well when a user's hands are free can perform substantially worse when the user is carrying an object or walking [4]. This motivates treating hands busy scenarios as a distinct condition rather than assuming one preference applies everywhere, and, more constructively, points to hands busy accessibility as the specific

problem space where voice input has the clearest opportunity to help rather than merely compete with touch.

A systematic review of usability questionnaires for voice user interfaces found that the System Usability Scale (SUS) remains the most common evaluation instrument for voice systems, but noted that voice interfaces are still an earlier stage field than touch and keyboard input, with fewer mature, standardized evaluation practices [5]. This supports treating self reported task and context level preference, as used in the present study, as a reasonable first step ahead of a full moderated usability protocol with standardized instruments.

Theoretical Foundation: HCI Principles Applied in This Study This paper is grounded in two classic HCI frameworks introduced above, Nielsen's usability heuristics [1] and Norman's principles of everyday design [2], extended by a third idea drawn from the encumbrance literature [4]: that usability is situational rather than fixed to a method. Table I below makes explicit how each principle maps onto this study's central concept, comparing voice and touch not in the abstract, but across task and context scenarios.

HCI Principle	How It Relates to This Study's Concept
Nielsen: Efficiency of Use [1]	Sets the benchmark for RQ1: completion time and effort are the direct measures used to compare voice and touch (Sections IV C, IV E).
Nielsen: Error Prevention [1]	Operationalized as RQ2; the misrecognition data (Section IV D) motivate the confirmation-before-execution and editable-transcript proposals in Section V D.
Norman: Visibility & Feedback [2]	Because voice has no visual display of system state, this study treats spoken confirmation, waveform cues, and glanceable transcripts (Section V D) as voice's substitute for Norman's feedback principle.
Norman: Affordances [2]	Explains touch's general-purpose edge on ease and accuracy (Section IV C): visible buttons and fields afford correct action before input, which spoken commands cannot replicate.
Situational Usability (extends [2], [4])	Frames RQ4 and Section V directly: usability is not fixed to a method but shifts with physical context, hand availability, motor ability, visual attention, which is the paper's central concept.

TABLE I. Mapping of HCI Principles to This Study's Voice-vs-Touch Concept

Read together, the first two principles, efficiency and error prevention, explain why touch wins under normal conditions, while the latter three, feedback, affordance, and situational usability, explain why that pattern reverses once a hand becomes unavailable. The paper's recurring concept, that voice input should be evaluated as a situational problem solver rather than a general-purpose replacement for touch, is therefore not a standalone observation but a direct application of Norman's situated view of design combined with the encumbrance literature [4], operationalized through Nielsen's efficiency and error-prevention criteria.

III. METHODOLOGY

This study follows a discovery first HCI research process. A structured online survey (Google Forms) was distributed to smartphone users to establish real world preferences, habits, and pain points before any prototype was built. This mirrors the Discovery phase of a broader six phase HCI process (Discovery, Define, Ideation, Design, Evaluate, Finalize). The Design and Evaluate phases, involving a working voice versus touch prototype and a moderated usability test, are planned as the next stage of this research and are described in Section VIII.

A. Participants 64 valid responses were collected. Respondents were predominantly aged 20 to 35 (78.1%), with smaller groups aged 15 to 20 (18.8%) and 35 to 50 (3.1%). All respondents were active smartphone users, and the survey did not restrict participation by device platform.

B. Instrument The survey instrument collected three categories of data: (1) baseline voice input habits, meaning frequency of use, experienced delays, and frequency of misrecognition; (2) preference between the two methods across ten scenarios spanning routine tasks (searching, form filling, menu selection, messaging, contact search) and physical or environmental contexts (one hand occupied, walking, quiet public space, noisy public space, temporary hand injury); and (3) overall comparative usability judgments, meaning which method was faster, more accurate, and easier overall, which method was preferred for sensitive information, which factor mattered most in choosing a method, and how confident respondents were that a voice input task would complete successfully. An open-ended question captured additional qualitative comments.

C. Procedure The survey was distributed and completed online between July 11 and July 13, 2026. Respondents answered based on their existing everyday experience with voice assistants (such as Siri, Google Assistant, or voice typing) and touch interaction, rather than in a controlled lab task. This prioritizes ecological validity and breadth of context over the precision of timed task measurements, which are reserved for the planned moderated evaluation described in Section VIII.

IV. FINDINGS

64 valid survey responses were received. Findings are organized by baseline voice input experience, task and context-based preference, overall comparative usability, and qualitative feedback.

A. Baseline Voice Input Experience Respondents were infrequent voice input users overall: 40.6% reported using voice input "Rarely," 35.9% "Sometimes," 17.2% "Very often," and 6.2% "Never." Reliability concerns were common. 43.8% reported experiencing delays "Often," and 87.5% reported at least occasional misrecognition (60.9% "Sometimes," 26.6% "Often"), with only 4.7% saying their input was never misunderstood. Confidence that a voice task would complete successfully was mostly neutral (60.9%) rather than confident (21.9% confident, 4.7% very confident), which lines up with the mixed reliability experience above.

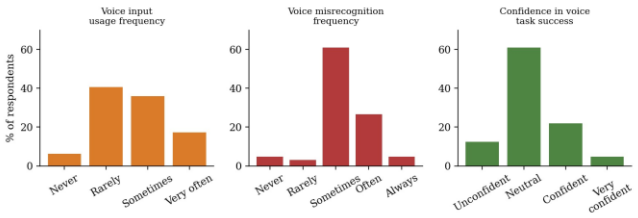


Fig. 1. Baseline voice input usage frequency, misrecognition frequency, and confidence in task success (N = 64).

B. Task and Context Based Preference Touch was preferred for every routine, non-encumbered task, most strongly for quiet public settings (90.6%), noisy public settings (87.5%), and short form filling (87.5%). Preference shifted toward voice only in the two scenarios that removed the use of one or both hands: holding a bag in one hand (62.5% voice) and a simulated temporary hand injury (64.1% voice). Notably, distraction alone (walking) did not produce the same shift. Touch still edged out voice (48.4% vs. 34.4%), which indicates that it is the physical loss of hand use, not distraction in general, that drives the preference reversal.

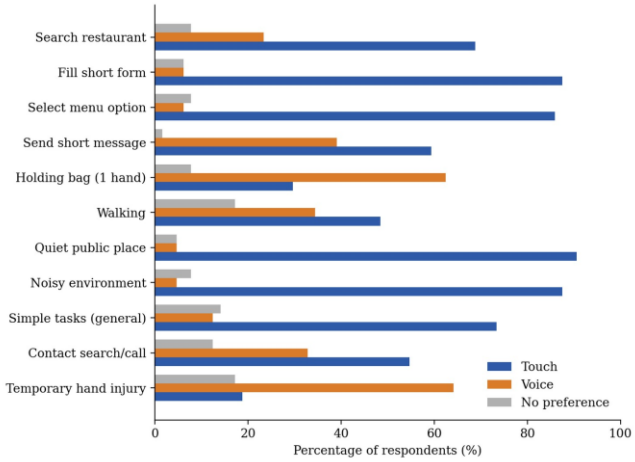


Fig. 2. Preferred interaction method across ten task and context scenarios (N = 64).

Task / Context Scenario	Touch	Voice	No Pref.
Search for a nearby restaurant	68.8%	23.4%	7.8%
Fill out a short form (name, email)	87.5%	6.2%	6.2%
Select an option from a menu	85.9%	6.2%	7.8%
Send a short message	59.4%	39.1%	1.6%
Holding a bag, one hand only	29.7%	62.5%	7.8%
Walking while using phone	48.4%	34.4%	17.2%
Quiet public place (library)	90.6%	4.7%	4.7%
Noisy environment (bus stop)	87.5%	4.7%	7.8%
Completing simple mobile tasks (general)	73.4%	12.5%	14.1%
Searching a contact / making a call	54.7%	32.8%	12.5%
Temporary hand injury (cast/bandage)	18.8%	64.1%	17.2%

TABLE II. Preferred Interaction Method by Task and Context (N = 64)

C. Overall Comparative Usability Touch was judged faster by 53.1% of respondents (29.7% equally fast, 17.2% voice), more accurate by 82.8% (10.9% equally accurate, 6.2% voice), and easier overall by 57.8% (32.8% equally easy, 9.4% voice). For sensitive information such as passwords or banking details, touch was preferred by 95.3% of respondents, reflecting a clear privacy concern around speaking such information aloud. When asked which single factor mattered most, accuracy was the dominant answer (53.1%), followed by convenience (23.4%), privacy (10.9%), speed (7.8%), and accessibility (4.7%), consistent with accuracy producing the widest gap between the two methods.

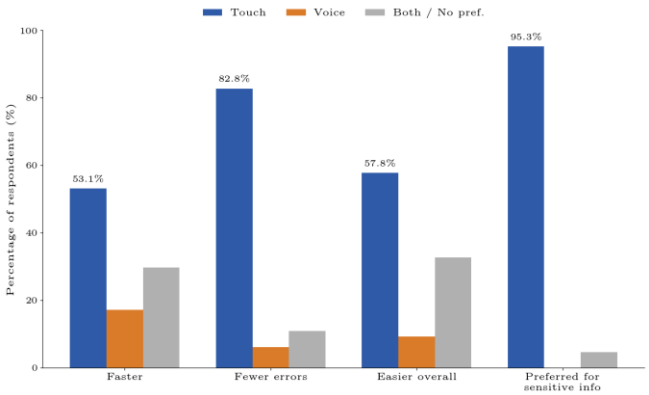


Fig. 3. Overall comparative usability judgments: speed, accuracy, ease of use, and preference for sensitive information tasks (N = 64).

D. Error Rate: Touch vs. Voice Touch produced the lowest self reported error rate of the two methods by a wide margin. 82.8% of respondents judged touch to result in fewer errors, against 6.2% for voice and 10.9% rating the two equals (Fig. 5). This lines up with the baseline data in Section IV A, where 87.5% of respondents reported at least occasional voice misrecognition and only 4.7% reported never experiencing it. Misrecognition, not slowness, is the dominant source of voice's error rate in this sample.

A natural question is whether voice's error rate falls as users gain experience with it, in other words, whether the errors seen here are a temporary, learnable cost rather than a persistent one. The data do not support that expectation. Respondents who use voice input "Very often" reported occasional or worse misrecognition at the same rate as infrequent users, if not slightly higher (100% of frequent users vs. 92.3% of "Rarely" users reported at least occasional misrecognition). What did rise with frequency of use was confidence, not accuracy. Average self rated confidence in task success climbed to levels roughly 70% higher among frequent users than among rare users, despite an equal or higher reported error rate. In other words, experienced users do not appear to encounter fewer voice errors. They appear to develop more tolerance for, and better recovery strategies around, an error rate that stays roughly constant. This matters for design: it argues against assuming voice's reliability problem will resolve itself with adoption, and in favor of the explicit repair mechanisms proposed in Section V D (confirmation before execution, editable transcripts) regardless of how experienced the user is.

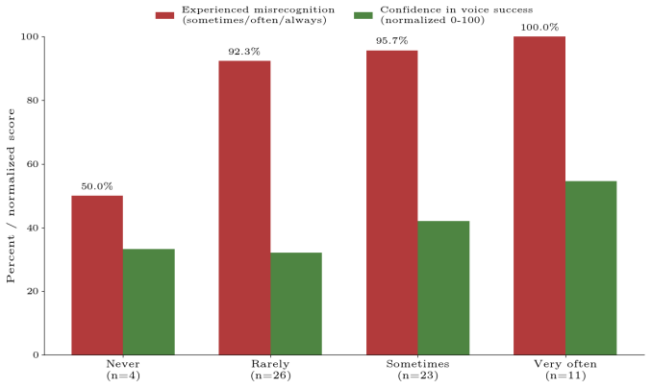


Fig. 4. Misrecognition frequency and confidence in voice task success, by voice input usage frequency (N = 64).

This survey measured self reported error frequency and did not directly count errors per task by physical condition. However, the sharp preference reversal toward voice specifically when a hand is occupied (Table II) is consistent with prior findings that touch's own accuracy depends on a stable, typically two-handed grip, and degrades under encumbrance [4]. Read alongside the accuracy figures above, this suggests touch's low overall error rate (82.8% preferred) likely does not hold uniformly across contexts. It is probably highest in the routine, hands free conditions this survey also shows touch dominating, and lower in the one-handed conditions where respondents shifted toward voice. Confirming this directly is a task for the timed, moderated evaluation in Section VIII, since it requires observing actual error counts under encumbrance rather than recalled preference.

E. Completion Time: Touch vs. Voice Respondents judged touch faster overall (53.1%), with 29.7% rating the two methods equally fast and only 17.2% favoring voice (Fig. 5). This runs against the intuitive expectation that speaking is faster than typing for anything longer than a few words. The most likely explanation sits in the same baseline data discussed above: 43.8% of respondents reported experiencing delays with voice input "Often," and any raw speed advantage voice has in the act of speaking is easily offset by processing latency, misrecognition, and the repair step (repeating or correcting a command) that follows it. Perceived speed, in other words, reflects total task time including error recovery, not just dictation speed.

This pattern lines up with Wallén's finding that voice was only faster than touch once users already knew exactly what commands the system supported [3]. Familiarity removes the hesitation and trial and error that would otherwise erode voice's raw speed advantage. Combined with the confidence without accuracy gain pattern in Section IV D, this suggests frequent voice users may feel faster without necessarily being faster, since rising confidence does not correspond to a falling error rate in this sample.

It is worth noting that this study captures perceived completion time through self report, not measured task time in seconds. Perceived and measured speed can diverge, particularly once delay and error recovery time are counted rather than felt. The moderated evaluation planned in Section VIII will address this directly by timing both methods on the same benchmark tasks, which is the only way to confirm whether touch's perceived speed advantage reflects an actual difference in completion time or a difference shaped by voice's reliability problems.

F. Qualitative Feedback Open ended comments reinforced this pattern. Several respondents described repeating or rephrasing voice commands after a misrecognition and found this frustrating. Others framed the two methods as complementary: voice was described as useful for quick, hands free tasks and for use cases like driving, while touch was preferred for precision, editing, or confidentiality. A recurring request was for improved recognition accuracy across accents and in noisy environments, along with larger, more responsive touch targets, suggesting that users want both methods improved rather than one replaced by the other.

V. HOW VOICE INPUT CAN HELP SOLVE REAL MOBILE INTERACTION PROBLEMS

The results above are sometimes read as a case against voice input, since touch wins on nearly every general-purpose measure. That reading misses the more useful conclusion: voice input is not failing at being touch's replacement. It is succeeding at solving a narrower, real set of problems that touch cannot solve well. The task is to design specifically for those problems rather than for general purpose parity with touch.

A. Problem 1: Temporary and Situational Loss of Hand Use
The clearest, data backed win for voice is exactly the condition where a user's hands are physically unavailable: one hand occupied by a bag (62.5% preferred voice) and a simulated temporary hand injury (64.1% preferred voice). These are not edge cases. They are common, recurring moments in daily life, such as carrying groceries or a child, a sprained wrist, wearing gloves, or cooking, where touch is not merely less convenient but sometimes not possible at all. Voice input removes the hard dependency on manual dexterity in these moments, turning a task that would otherwise require putting an item down, or being unable to act at all, into one that can be completed hands free.

B. Problem 2: Accessibility for Users with Motor Impairments
The temporary hand injury scenario in this survey is a proxy for a broader and more permanent population: users with arthritis, tremor, limited fine motor control, or upper limb disability, for whom small touch targets and multi step form filling are a persistent barrier rather than an occasional inconvenience. For this population, voice input is not a convenience feature but an accessibility requirement. It can remove a barrier to independent smartphone use that touch only design leaves unaddressed. Framing voice input as an accessibility feature, rather than a novelty, also justifies investing in its reliability even in task categories where it is not the faster or more accurate option for the general population.

C. Problem 3: Eyes Busy and Attention Constrained Contexts
Although walking alone did not shift preference toward voice as strongly as hand occupation did in this sample, voice input still solves a distinct problem in genuinely eyes busy contexts such

as driving, cooking, or exercising, where looking at and tapping a screen is unsafe or impossible even though the hands may be technically free. The 39.1% of respondents who preferred voice for sending a short message, the highest voice preference figure among the non-encumbered routine tasks, hints at this same appeal for quick, low precision tasks where reading and tapping a full reply takes more effort than simply speaking one.

D. Addressing the Errors and Privacy Concerns This Study Found For voice input to solve the problems above without introducing new ones, the reliability and privacy issues respondents reported need direct design responses rather than being treated as unavoidable trade offs:

- Misrecognition, reported by 87.5% of respondents at least sometimes, can be eased with a brief on screen confirmation of what the system heard before an action is executed, rather than silent auto execution, and by allowing a quick touch-based correction rather than requiring the entire voice command to be repeated.
- Delays, reported "Often" by 43.8%, can be reduced by processing short, common commands on the device rather than relying on a network round trip, and by giving immediate listening feedback (such as a waveform or pulse) so users are not left uncertain whether the system registered their voice at all.
- Privacy concerns, with 95.3% preferring touch for sensitive information, can be addressed by having the interface automatically decline spoken input for password, PIN, or banking fields and instead prompt a switch to touch for that field only, rather than asking the user to remember not to speak sensitive data themselves.
- Social acceptability concerns in public settings, with only 4.7% preferring voice in both quiet and noisy public places, can be reduced, though not eliminated, by supporting a whispered or low volume input mode and a visible, glanceable transcript so a user can silently verify what was heard without speaking again.

Together, these changes reframe voice input's current weaknesses as solvable design problems rather than inherent limitations, and align the technology with the specific situations, hand unavailability, motor accessibility needs, and eyes busy moments, where this study shows it already has a genuine, user confirmed advantage.

VI. DISCUSSION

Taken together, the findings answer the four research questions directly. Touch is faster (RQ1), more accurate (RQ2), and easier overall (RQ3) for the general population across most everyday tasks. RQ4, however, is confirmed in a narrower but meaningful way: voice input's advantage depends on whether a hand is physically available, not on distraction or multitasking in general. Section V reframes this conditional advantage constructively. Rather than treating voice as an inferior general-purpose input method, the data point to it as a targeted solution for hand unavailability and accessibility needs, provided its reliability and privacy weaknesses are addressed through the design changes above.

A practical interface recommendation follows from this: rather than offering voice as a parallel, always available input mode, a mobile interface could surface it contextually, for example when sensors suggest one handed or encumbered use, or when a user has indicated a motor accessibility need, while keeping touch as the default, more accurate path for form filling, menu navigation, and any task involving sensitive data.

VII. LIMITATIONS

This study reports self reported preference and recalled experience rather than measured task time or error counts from a controlled test. The sample skews toward the 20 to 35 age group and was recruited through a convenience sample rather than a probability sample, so the preference figures should be read as directional rather than population representative. Voice input experience also varied widely across respondents, meaning perceptions of voice usability are shaped by differing prior exposure to voice assistants rather than a single standardized set of trials. The planned moderated evaluation below is intended to address these limitations directly.

VIII. EVALUATION PLAN (NEXT PHASE)

This survey is a Discovery phase study and is not a substitute for controlled usability testing. As an initial Design phase step, we sketched the simple screen shown in Fig. 7: a standard registration form with a voice option placed alongside touch, rather than hidden in a settings menu. The next phase will turn this sketch into a working prototype and run moderated sessions measuring completion time, error and repetition count, and a post task ease of use rating.

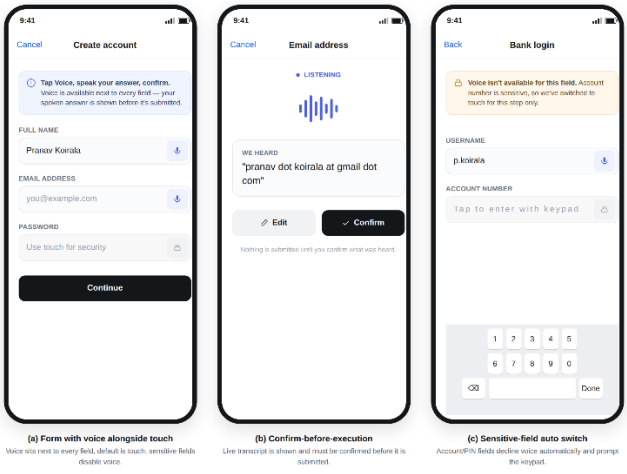


Fig. 7. Proposed mobile screen offering voice as an alternative to touch for the same form filling task.

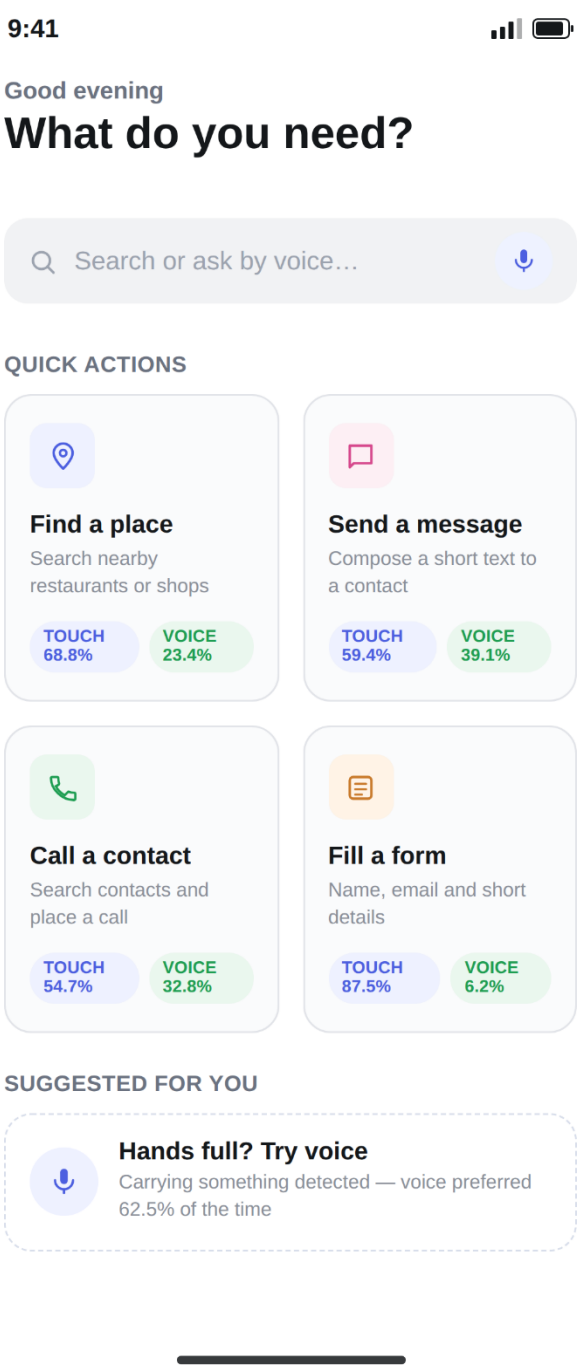


Fig. 7. Proposed mobile screen offering voice as an alternative to touch for the same form filling task.

The design is intentionally simple. Touch remains the default, matching this study's finding that touch is preferred for most everyday tasks. The Voice button sits right next to it, not behind an extra menu, so it is available the moment a user's hands are full without requiring them to search for it. This directly answers RQ4, where voice was preferred once a hand was occupied. A short prompt, "Tap Voice, speak your answer, confirm," sets the expectation that the spoken answer will be shown back before it is submitted, which is the confirmation step Section V D identifies as necessary given how often respondents reported misrecognition.

IX. CONCLUSION AND FUTURE WORK

This study surveyed 64 smartphone users to compare voice input and touch interaction across ten everyday mobile task and context scenarios. Touch was preferred overall for speed, accuracy, ease of use, and privacy, and accuracy was the single most important factor respondents used to choose an interaction method. Voice input's advantage was narrower but genuine and actionable: it reversed the overall pattern specifically when a hand was unavailable, pointing to hand unavailability and motor accessibility as the problems voice input is best positioned to solve, provided its recognition error and privacy weaknesses are designed around rather than ignored. Future work includes building and testing the prototype described in Section VIII, and extending the context set to other common hands busy scenarios such as cooking, driving, and exercising.

REFERENCES

[1] J. Nielsen, "Enhancing the explanatory power of usability heuristics," in Proc. ACM CHI Conf. Human Factors in Computing Systems, 1994, pp. 152 to 158.

[2] D. A. Norman, The Design of Everyday Things, revised ed. New York, NY, USA: Basic Books, 2013.

[3] F. Wallén, "Comparing voice and touch interaction for smartphone radio and podcast application," M.S. thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2017.

[4] Investigating the Effect of Encumbrance on Gaze and Touch Based Target Acquisition on Handheld Mobile Devices, arXiv preprint, 2025.

[5] A. L. Iñiguez-Carrillo, L. S. Gaytán-Lugo, M. A. Garcia-Ruiz, and R. Maciel-Arellano, "Usability questionnaires to evaluate voice user interfaces," IEEE Latin America Transactions, vol. 19, no. 9, pp. 1468 to 1477, 2021.