



OPEN ACCESS

EDITED BY

Carlos Duarte,
University of Lisbon, Portugal

REVIEWED BY

Maikel Lázaro Pérez Gort,
Ca' Foscari University of Venice, Italy
Qiwei Li,
California State University, Fresno,
United States

*CORRESPONDENCE

Hui Huang
✉ 1132419014@qq.com

RECEIVED 07 July 2025

ACCEPTED 06 October 2025

PUBLISHED 13 November 2025

CITATION

Huang H and Hou G (2025) The impact of
usage experience and input modality on trust
experience and cognitive load in older adults.
Front. Comput. Sci. 7:1659594.
doi: 10.3389/fcomp.2025.1659594

COPYRIGHT

© 2025 Huang and Hou. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

The impact of usage experience and input modality on trust experience and cognitive load in older adults

Hui Huang^{1*} and Guanhua Hou²

¹Yibin Hospital Affiliated to Children's Hospital of Chongqing Medical University, Yibin, China,

²Southeast University, Nanjing, China

Trust experience plays a pivotal role in human–computer interaction, particularly for older adults, where it serves as a critical psychological threshold for technology adoption and sustained usage. Against the backdrop of increasingly diverse intelligent interaction modalities, trust directly influences older adults' initial acceptance and long-term reliance on technological systems. This study focuses on the interactive effects of users' experience and input modality on trust experience and cognitive load in the elderly. Employing a 2 (prior experience: experienced vs. inexperienced) × 3 (input modality: touch, speech, eye control) mixed experimental design. Following each task, participants completed NASA-TLX scales and trust perception questionnaires, supplemented by eye-tracking data to quantify cognitive load and behavioral patterns. The results showed that (1) Experience-dependent divergence in trust perception: Experienced older adults exhibited higher trust in touch input, attributable to established press-response mental models from prior device usage, while inexperienced users preferred speech input due to its alignment with natural conversational paradigms. (2) Cognitive load mediation effect: Although voice input reduces the learning cost of user interfaces for inexperienced elderly users (NASA-TLX is 24% lower than touch), recognition errors can cause a sharp drop in trust; This study reveals that the trust experience of elderly users is influenced by both usage experience and input methods, with cognitive load being a key mediating factor. In terms of design, the touch physical metaphor should be retained for experienced elderly users, and the voice fault tolerance mechanism should be strengthened for inexperienced elderly users, while reducing technical anxiety by enhancing operational visibility.

KEYWORDS

input modality, human–computer interaction, older adult interaction, user trust experience, cognitive load

1 Introduction

The advancement of AI and computer vision technologies has provided strong support for the creation of more natural and efficient human–computer interaction scenarios, as well as multimodal forms of input, but it has also raised many concerns about the trust experience. The trust experience is the most profound feeling that users have when using a product, and trust is critical in regulating the relationship between humans and automated systems, as well as the acceptance, adoption, and continued use of interactive systems (Gefen et al., 2003; Lee and See, 2004). The input modality (speech, gesture, eye gaze, and touch) that a user uses to communicate information to a mobile self-service terminal is particularly significant because it allows the user to access personal data and perform tasks like viewing personal information

and conducting online transactions (Vildjiounaite et al., 2006; Lee and See, 2004). However, because different input modes require different ways of operating, this is especially challenging for older users who are already dealing with the challenges of the digital divide. At the moment, older adults have relatively poor trust experiences and low acceptance of smart technologies due to lower transparency and interpretability in human-computer interaction. Given the widespread use of smart products in mobile self-service terminals, it is critical to investigate the effect of various input modalities on the trust experience of older users.

Elderly people's cognitive modes are heavily influenced by their prior experiences and knowledge; thus, interaction systems that correspond to their experiences are more likely to be understood and accepted. Research has shown that an individual experiential knowledge has a significant impact on the trust experience in human-computer interaction (HCI), influencing not only the user's trust in the system but also their understanding and satisfaction with the system. A complete human-computer interaction process consists of two major components: user input mode and system output feedback. The user implements control behaviors using input commands and interprets their responses using system feedback. Different input modalities have a significant impact on the user experience.

Based on this, this study focuses on the various input modes used by elderly users when interacting with smart bodies, as well as how differences in usage experience influence elderly user trust experience. In addition, this study will investigate whether usage experience modifies older trust experience. Furthermore, the study aims to determine whether using different input modes increases the cognitive load of older adults and whether usage experience plays a moderating role in the process.

Therefore, this study aims to answer the following research questions:

1. What are the differences in task performance and cognitive load of older adults when using different input modalities (touch, speech, and eye control)?
2. Do older adults with different usage experiences (experienced vs. inexperienced) differ in input modalities (touch, voice, eye control), trust experience, and cognitive load?
3. Is there an interaction between user trust experience, cognitive load, and task performance?

2 Related work

2.1 User trust experience

The critical role of trust in technology adoption has been well-documented across various domains, particularly in the acceptance of novel interactive systems (Hoff and Bashir, 2015). Studies on human-automation interaction consistently identify trust as a pivotal determinant of user acceptance (Gefen et al., 2003), serving as a psychological mediator between users and technological systems (Ghazizadeh et al., 2012). Trust significantly influences users' willingness to engage with technology, especially in contexts requiring risk mitigation (Hengstler et al., 2016). In the realm of self-service terminals, trust formation varies

substantially across input modalities (e.g., speech, touch, eye control) and is further modulated by age-related differences. Prior research suggests that older adults' trust in touch interfaces often stems from schema transfer from legacy devices (e.g., feature phones, ATMs), whereas their trust in speech interfaces reflects alignment with natural communication paradigms (Oviatt, 2022). However, excessive cognitive load—common in complex modality interactions—can erode trust by depleting metacognitive resources necessary for confidence calibration (Parasuraman, 2000).

Trust is a multidimensional construct defined as "the attitude that a system will fulfill user goals amid uncertainty" (Lee and See, 2004). Its calibration depends on three factors: (1) modality familiarity (e.g., older users' predisposition toward tactile interfaces) (Claypoole et al., 2016), (2) transparency of system feedback (Mikulski, 2014), and (3) error tolerance (e.g., voice recognition errors disproportionately reduce trust in novice users) (Kaye et al., 2018). Mismatches in trust calibration—such as overreliance on flawed touch systems or distrust of efficient speech interfaces—can lead to either misuse or disuse of self-service technologies (Pop et al., 2015). While Lee and See (2004) emphasize that trust builds through cognitive and affective pathways, its dynamic nature means it fluctuates with user experience.

Despite its centrality, trust in multimodal self-service systems remains understudied, particularly for older adults. Existing work seldom addresses how age-specific factors (e.g., cognitive decline, technophobia) interact with modality-dependent trust formation (Choi and Ji (2015)). This gap is critical because older users—facing steeper learning curves with touch or eye control—may default to speech despite its potential cognitive overhead (Basu, 2021). A systematic integration of trust dynamics into self-service design could mitigate adoption barriers across age groups.

2.2 Input modality

2.2.1 Utility of touch input modality in self-service terminals

The low cost of hardware means that touch input devices are a popular input modality for self-service terminals installed in public places, offering services such as hospitals (Shang et al., 2020), internet access (Guo et al., 2007), information (Slay et al., 2006), city guides (Johnston and Bangalore, 2004), banks (Paradi and Ghazarian-Rock, 1998) and many more. These self-service terminals reduce the need for costly staff, and contents can be changed in real time. Compared to traditional physical-button-based interaction, touch input has the advantage of being concise and convenient, and was reported as having high usability and being more preferred by young users (Lee et al., 2020). Most touch-based self-service terminals are based on absolute positioned virtual buttons which are difficult to locate without any tactile, audible or visual cues (Sandnes et al., 2012). Older adults in particular may have struggle searching and clicking on targets due to reduced information processing, precise movement and timely control (Fisk et al., 2020; Leonardi et al., 2010). Chène et al. (2016) also found that they have a harder time using clicks than younger adults do because of invalid touches. In conclusion, touch input is the dominant input modality for self-service terminals and is popular with young people for its

simplicity of operation. However, it may not be as user-friendly for older adults when performing actions such as searching and manipulating. Therefore, it is necessary to explore the accessibility of touch input for different age groups when using medical self-service terminals.

2.2.2 Utility of touchless interaction for older adults

The user’s touchless input modalities such as voice, gesture, or eye control are commonly utilized as a means to transmit information to the system. Recently, the input modalities that gain a lot of attention are voice and eye-control input (Kim et al., 2021). As voice input device become a mature technology, it has emerged as a popular touchless input modality. Voice input allows users to utter a command that can be recognized by the system to execute an operation (e.g., return) or enter certain information (e.g., outpatient charges) (Zhang et al., 2023). The use of voice input technology can go some way to improving the usability and user experience of self-service terminals (Sajić et al., 2021). It increases the confidence of older adults and improves their acceptance of self-service terminals (Chi et al., 2020). Kaufman et al. (1993) showed that voice input provided users with introductory interaction to increase usability. Manzke et al. (1998) investigated the use of self-service terminals by visually impaired users, in which the use of voice input was evaluated. For visually impaired users, voice input showed significant performance advantages. This also benefits older adults with reduced vision. Voice input is one of the touchless input modalities, which improves the accessibility of self-service terminals. However, does voice input benefit older adults in using self-service terminals and improve their task performance and usability of operations? These questions are yet to be thoroughly researched and explored.

Eye-control input relies on gaze behavior for computer interaction. Users can use eye gaze to select products (Kim et al., 2015). Determining the appropriate dwell time for the eye-control input becomes critical when the eye-control input is used to simulate a click and confirm operation (Hansen et al., 2001). Researchers have tried to improve user experience by setting different gaze times to complete information selection (Pfeuffer et al., 2021). Regarding eye-control input, the best dwell time is 600 which is recorded in milliseconds (ms) in the trigger system (Ya-feng et al., 2022). Niu et al. (2019) also showed that when the gaze dwell time is set to 600 ms, the efficiency of the interaction is the highest, and the task load of users is minimal as well. However, prolonged gaze dwells time and eye fatigue can result in an excessive cognitive load (Sato et al., 2018). In addition, eye-control research has focused on young individuals (Rozado et al., 2012). It is questionable whether it is suitable for older people to use eye-control input to operate self-service terminals and cause the high cognitive load.

3 Methods

3.1 Participants

This study recruited 40 participants and a vision and listening ability test was conducted among the participants by using the Chinese

TABLE 1 Information for participants.

Information	Participants data
Gender	20 (male), 20 (female)
Age	60 years and older (Mean ± SD: 64 ± 3.127)
Experience	20 (experienced), 20 (inexperienced)

All the participants had a visual acuity of 1.0 LogMAR, and a hearing ability of 25 dB or less (Roth et al., 2011; De Raedemaeker et al., 2022).

version of the Functional Visual Screening Questionnaire (FVSQ) (1991). Each participant was able to walk easily and able to complete tasks independently. Table 1 shows a detailed breakdown of participant demographics.

Experienced users were defined as those who reported using smart devices ≥3 times per week over the past year and could independently perform complex operations (e.g., online payments, app installations). Inexperienced users were defined as those who used smart devices <1 time per week or required assistance with basic operations (e.g., download software, register account). This dichotomous approach helped minimize within-group variance and ensured clearer detection of experimental effects. The older adults were active or retired school employees. Informed consent was obtained from all the participants before the experiment was conducted, and each participant was paid 50 RMB for their participation.

3.2 Experimental material

Using both web-based and field research, this study analyzed the interfaces of the medical self-service systems in China and used a standard interface of a medical self-service system as the experimental material. Diagram of experimental material consisting of the home page and six other pages for each step in the task to complete a registration (see Figure 1; English version in Figure 2).

3.3 Experimental design

This study employed a mixed-design experiment with two factors: a three-level within-subjects factor of Input Modality (Touch, Speech, Eye Control) and a two-level between-subjects factor of User Type (Experienced, Inexperienced). The design was implemented to systematically evaluate the impact of different interaction modes and prior experience on key dimensions of user experience. Specifically, each participant interacted with all three input modalities, while being assigned to one of the two user type groups. The dependent variables encompassed a multi-dimensional set of metrics, primarily including subjective perceptions (e.g., user trust and cognitive load) and objective behavioral performance (e.g., task completion time), as detailed in the table below. The study recruited a total of 40 participants, comprising an experienced group (n = 20) and an inexperienced group (n = 20). Task completion time was measured in seconds (s) and was recorded for every trial starting from the start cue up until the participants submitted their responses to end each trial. The system cannot tolerate skipping each step, as shown in Table 2.



FIGURE 1
Experimental material diagram (corresponding to the experimental task).

3.4 Measurement method

3.4.1 Trust experience questionnaire

Research has demonstrated that trust scales can be a straightforward and efficient way to assess the level of trust between humans and computers. These scales are based on the perception of the individual who is placing their trust and are both user-friendly and adaptable to large-scale applications (Ma et al., 2025). Jian et al. (2000) classified user trust into three dimensions: motivation, operation, and utility, based on the order of interaction. They developed and validated a user trust scale that is highly usable (Cronbach's $\alpha = 0.92$). This experiment is divided into three trust dimensions: motivation, which is the initial trust generated by the user's use of the input modality; operation, which is the real-time trust in the user's behavior when carrying out the input modality; and utility, which is the ex-post trust formed by the final control effect of such an input modality. The questionnaire for this study contains three trust factors and twelve measurement items on a seven-point Likert scale (least agree, strongly disagree, disagree, neutral, agree, strongly agree, most agree) (Likert, 1917).

3.4.2 NASA-TLX scales

Cognitive load was measured using the NASA-TLX scale developed by Hart and Staveland (1988). The NASA-TLX scale items were rated on a 20-point scale (0 = low, 20 = high). The mental demand, physical demand, temporal demand, performance, effort,

and frustration subscales were combined to create a composite NASA-TLX workload score (scaled to 0 = low, 100 = high) (Lowndes et al., 2020).

3.4.3 Eye-tracking data measurement

Eye-tracking fixation data is a measure of visual perceptual engagement. For this experiment, visual workload was evaluated using two metrics: (1) the total fixation duration on the areas of interest (AOIs), and (2) the number of fixations. These metrics were collected and preprocessed for subsequent analysis. Eye-tracking studies used predefined areas of interest (AOIs) based on relevant regions and targets. Fixation duration is an attention distribution indicator that measures how long the eye stays in the area of interest (AOI) (Eckstein et al., 2017). A longer fixation duration indicates difficulty in extracting information (Just and Carpenter, 1976). The number of fixations is the number of user gaze points located within the target AOIs. A longer fixation duration indicates difficulty in extracting information (Just and Carpenter, 1976). The number of fixations is the number of user gaze points located within the target AOIs. The higher the number of fixations, the more difficult it is to identify the target in the search task and the greater the cognitive load (Poole and Ball, 2006). Fixation is the relatively static state of the eye in a certain period. This makes the foveal vision stable in a certain location so that the visual system can obtain the details of an object. Measurable fixation must have a minimum duration of 60 which is recorded in milliseconds (ms), and the gaze velocity should not exceed 30°/s (Olsen, 2012). Among them, "ms" is the abbreviation of "millisecond,"

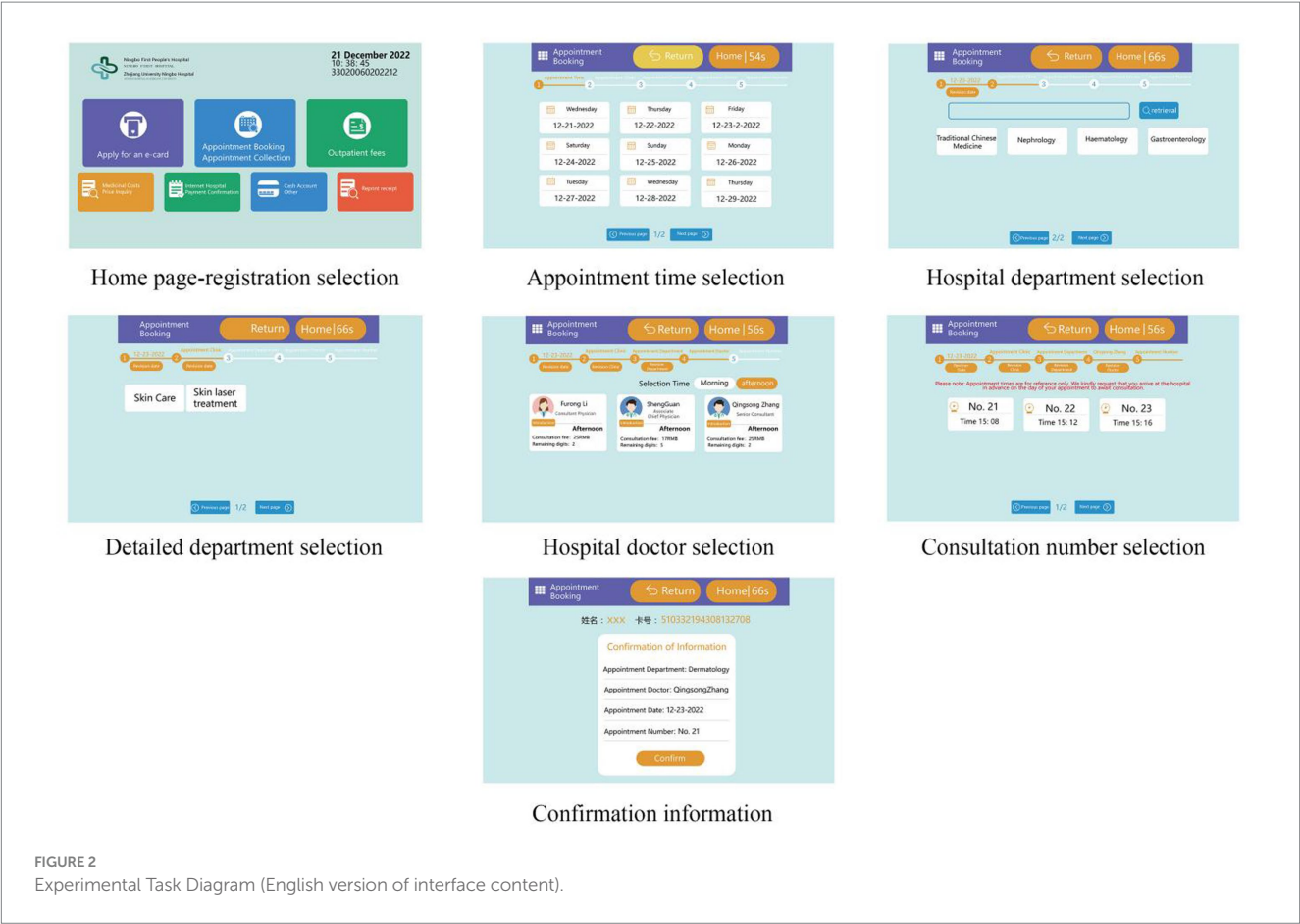


FIGURE 2
Experimental Task Diagram (English version of interface content).

TABLE 2 Experimental design.

Variable type	Variable name	Levels/measurement method	Design type (within-/between-subjects)
Independent variable 1	Input modality	Touch, speech, eye control (3 levels)	Within-subjects
Independent variable 2	User type	Experienced, inexperienced (2 levels)	Between-subjects
Dependent variable	Trust experience questionnaire	Subjective scale measurement	–
	Cognitive load	Subjective scale measurement	–
	Task completion time	Objective recording (seconds)	–
	Eye-tracking data	The total fixation duration	–
		The number of fixations	–

TABLE 3 Input modality operation.

Input modality	Hardware equipment	Operating method
Touch	Touch screen PC	Touch screen
Speech	Touch screen PC	Voice Calls, e.g., 'Select Dr. XX.'
Eye-control	The Tobii Pro spectrum	Gaze trigger time of 600 ms

which means “millisecond.” It is a unit of time, and 1 ms is equal to 10^{-3} s (i.e., one-thousandth of a second). “°/s” is the abbreviation of “degree per second,” which means “degree per second.” It is a unit of angular velocity and is used to express the angle that an object rotates through per second.

4 Experimental equipment

The operation of the three input methods involves the utilization of hardware devices, software programming, and operating methods, as demonstrated and elucidated in Table 3.

4.1 Hardware environment of the system construction

The Tobii Pro Spectrum eye-tracking device was used in this study for eye-control input, as shown in Figure 3A. The eye-control input device is particularly suited for scientific research involving the observation of eye movements across various experimental settings. This device accommodates a wide range of head movements, enabling participants to record data with high accuracy and precision. It can be mounted on a monitor, laptop, or other compatible devices to facilitate eye-controlled interactions. As depicted in Figure 3B, a 24-inch IPS bezel-less touchscreen monitor is utilized for touch input. This monitor features a Full HD 1080p display with a maximum resolution of 1920×1080 and an aspect ratio of 16:9, providing clear and detailed visuals. Aliyun's intelligent voice interaction technology can be integrated with a touchscreen computer, a spatial audio speaker, and a two-channel microphone array to support audio input and output. This combination enables seamless voice-controlled interactions and enhances the overall user experience.



4.2 The software environment of the system construction

This study employed the following software tools for the development of the voice interaction system:

- Visual Studio 2022: Used for C# programming, compilation, and debugging.
- Unity3D: Utilized for the development of the input system.
- Adobe Icon Design Tools (e.g., Adobe Photoshop, Adobe Illustrator): Employed for icon creation and interface design.

All software tools were deployed on a Windows 10 operating system environment. Voice materials were synthesized using a text-to-speech mini program, which was based on the Aliyun Text-to-Speech Platform API.¹ A neutral female synthetic voice (system identifier: Zhitian) was configured with a speech rate of approximately 5 words per second while maintaining default prosodic parameters (fundamental frequency = 0 Hz; amplitude modulation = 0 dB).

5 Procedure

The participants were asked to answer a questionnaire regarding their age, educational background, and experience using self-service system before the start of the experiment. In addition, a visual and auditory ability test was used to screen participants. Prior to the formal experiment, a practice session was implemented to familiarize participants with the experimental procedure. This session also served to verify that their actual operational ability aligned with the experience level identified during pre-screening. Each participant was required to practice three modalities by completing hospital department search tasks, which were different from the experimental task. Training continued until the participants were familiar with and correctly completed the task for each input modality. Then, the Tobii Pro Spectrum device started eye calibration and recorded data. Participants followed the instructions on the display and completed seven medical registration tasks using the input modality, as shown in Table 4. At the end of each level of testing,

¹ <https://ai.aliyun.com/nls/tts>

TABLE 4 Experimental task content arrangement.

Task	Touch	Speech	Eye control
Registration selection	Appointment booking	Appointment booking	Appointment booking
Appointment time selection	May-17	May-13	May-21
Hospital department selection	Orthopedics	Dermatology	Gastrointestinal surgery
Detailed department selection	Orthopedic care	Skin laser	Gastrointestinal surgery care
Hospital doctor selection	Dr. Zheyang Wang	Dr. Hongming Zhu	Dr. Jianming Xie
Consultation number selection	No. 36 in the morning	No. 48 in the morning	No. 31 in the morning
Verifying information	Confirmation	Confirmation	Confirmation



FIGURE 4
AOI divisions for different input modes. (A) shows the areas of interest for the touch input interface, (B) shows the areas of interest for the voice input interface, and (C) shows the areas of interest for the eye-control input interface.

participants would complete the trust experience questionnaire and NASA to record feedback on their experience. The time it took for each individual to complete all the tests under uniform screen brightness and ambient light ranged from 20 to 30 min, with a 2–5 min delay between each of the three variants of the experiment.

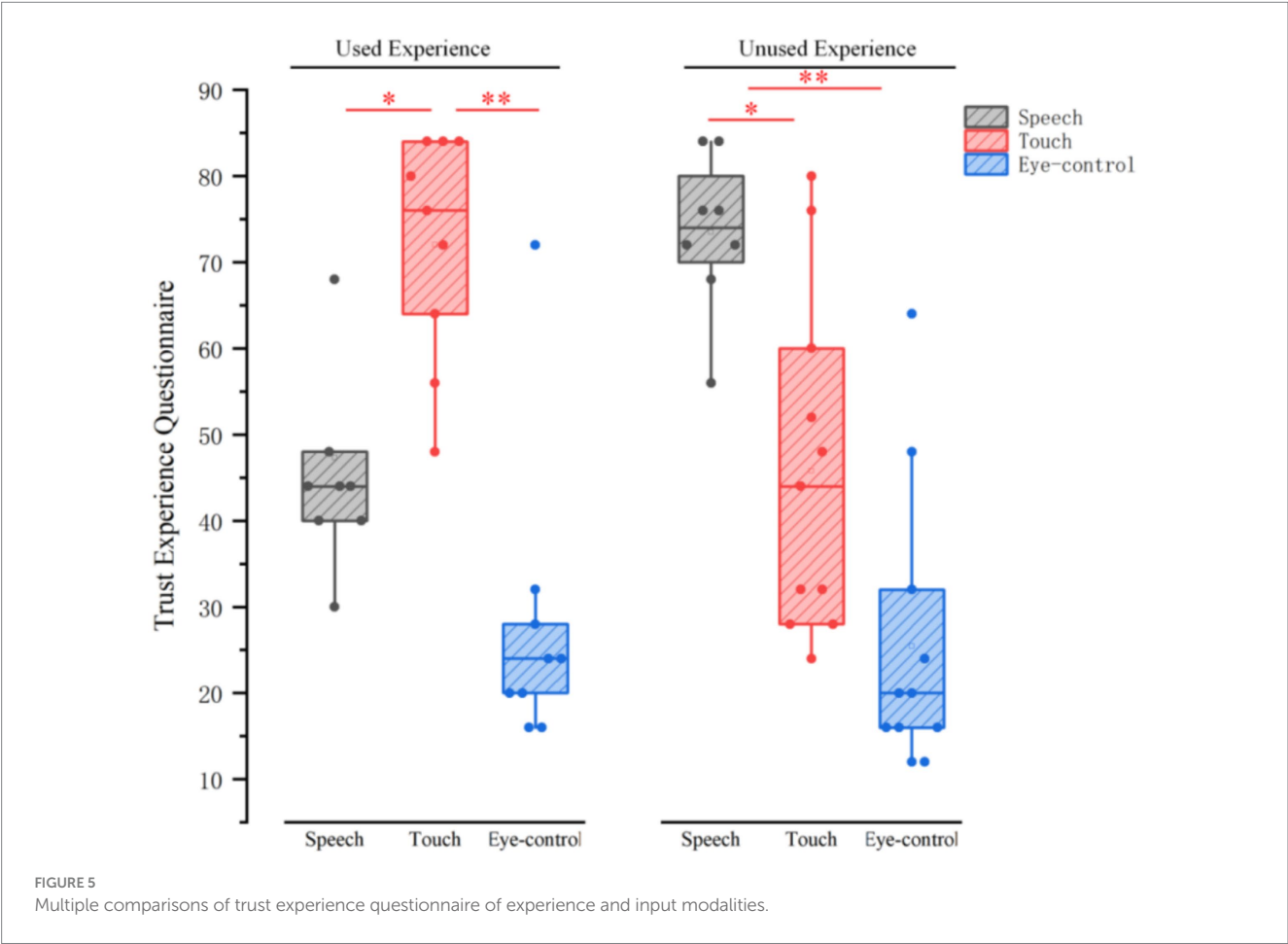
6 Data collection and analysis

A total of 120 video data and scales were collected for this experiment. Before conducting further analysis, the AOI for each experimental video was defined according to the tasks (see Figure 4). In Figure 4, A refers to the touch input interface, where the yellow area

TABLE 5 Analysis of variance (ANOVA) for six experimental conditions.

Dependent variables	Factors	df	<i>F</i>	<i>p</i>	η_p^2
TS	Experience	1	0.100	0.760	0.012
	Input modality	2	13.295	0.000***	0.624
NASA	Experience	1	33.172	0.000***	0.806
	Input modality	2	3.848	0.043*	0.325
TDOF	Experience	1	6.611	0.030*	0.424
	Input modality	2	47.478	0.000***	0.841
NOF	Experience	1	3.227	0.106	0.264
	Input modality	2	67.865	0.000***	0.883
TCT	Experience	1	19.980	0.002**	0.698
	Input modality	2	55.489	0.000***	0.860

Asterisks denote statistical significance: **p* < 0.05, ***p* < 0.01, ****p* < 0.001.



is the task AOI; B represents the speech input interface, where the purple area is the task AOI; and C represents the eye-control input interface, where the green area is the task AOI. The individuals' processing levels with respect to various AOIs were investigated by comparing their total duration of fixation and total number of fixations for these AOIs. All data were analyzed through repeated-measures analysis of variance. Repeated measures ANOVA examines data collected from the same subjects across different time points or conditions by partitioning variance components and employing an

F-statistic to evaluate treatment effects against error, thereby controlling for individual differences and assessing significance among measurements. IBM SPSS Statistics 19 software was used to analyze the aforementioned results, with *p* < 0.05 set as the significance level. Before analysis of variance (ANOVA) was performed, the normal data distribution was examined for each condition. In addition, Mauchly's spherical test was conducted to correct the results of the repeated-measures ANOVA for different input modalities and different user types.

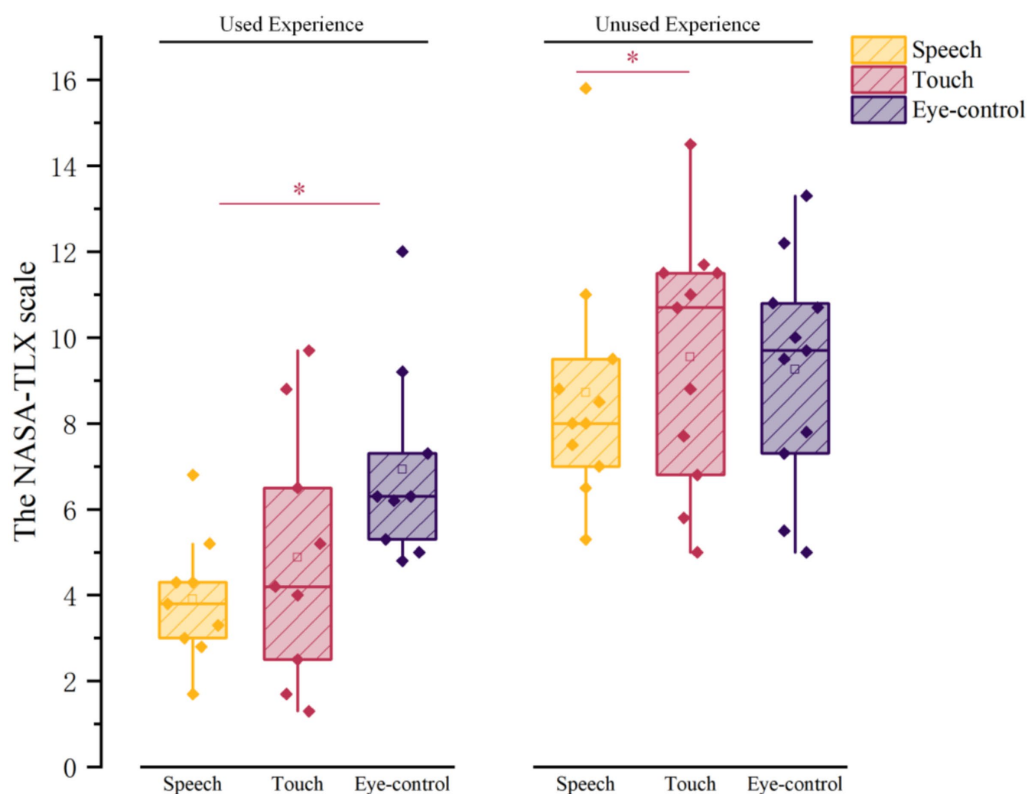


FIGURE 6
Multiple comparisons of the subjective cognitive loads of users and input modalities.

7 Results

The six experimental conditions (2 user experiences and 3 input modalities) were assessed using five measures: the User Trust Scale, the NASA-TLX Scale, the Total Duration of Fixation, the Total number of Fixations, and the Task Completion Time. Table 5 summarizes the main effects from the analysis of variance (ANOVA) for the six experimental conditions. To facilitate the presentation, we abbreviate the names of the five measures as follows:

- TS: Mean score on the user trust experience questionnaire.
- NASA: NASA-TLX Scale aka cognitive load scale mean score.
- TDOF: Total Duration of Fixation of interest in the AOI for the user.
- NOF: Total number of Fixations of interest in the AOI for the user.
- TCT: Time for user to complete tasks.

7.1 Trust experience questionnaire

The results of the ANOVA showed that there was no significant difference between the older adults' experience of use ($F = 0.10$, $p = 0.760$, $\eta^2 = 0.012$) on the user trust experience. Input modality made a significant difference in user trust experience ($F = 13.295$, $p = 0.000$, $\eta^2 = 0.624$) (see Table 1). Touch input ($M = 60.889$, $SD = 3.683$) had a higher user trust experience than speech input

($M = 58.667$, $SD = 3.037$) and eye-control input ($M = 27.333$, $SD = 5.637$). Multiple comparisons (see Figure 5) revealed that experienced older adults had a higher trust experience with touch input ($M = 72$, $SD = 4.42$) than speech input ($M = 46.22$, $SD = 4.9$) and eye-control input ($M = 28$, $SD = 5.77$); inexperienced older adults had the highest trust experience with speech input ($M = 71.11$, $SD = 4.04$), which was significantly higher than touch input ($M = 49.78$, $SD = 6.5$) and eye-control input ($M = 26.67$, $SD = 6.03$).

7.2 The NASA-TLX scale

Experience of use ($F = 33.172$, $p = 0.000$, $\eta^2 = 0.806$) and input modality ($F = 3.848$, $p = 0.043$, $\eta^2 = 0.325$) had a differentially significant effect on cognitive load (see Table 1). Older adults with use experience had a lower cognitive load ($M = 5.241$, $SD = 0.380$) compared to those without use experience ($M = 9.119$, $SD = 0.518$). Speech input ($M = 5.794$, $SD = 0.373$) had the lowest cognitive load, significantly less than touch input ($M = 7.522$, $SD = 0.751$) and eye control input ($M = 8.222$, $SD = 0.623$). All older adults had the lowest cognitive load using speech input; experienced older adults had a significantly lower cognitive load using speech input ($M = 3.911$, $SD = 0.495$) than eye-control input ($M = 6.93$, $SD = 0.775$); and inexperienced older adults had a significantly lower cognitive load using speech input ($M = 7.678$, $SD = 0.425$) than touch input ($M = 10.167$, $SD = 0.954$) (see Figure 6).

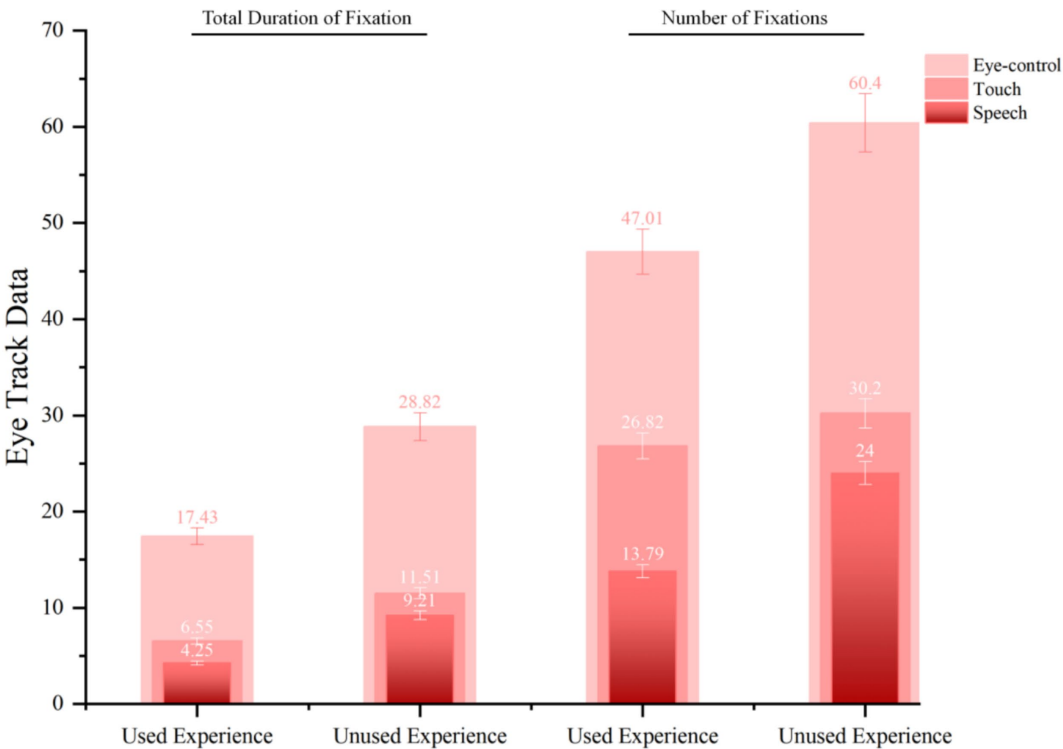


FIGURE 7 Multiple comparisons of the total number of fixations according to users and input modalities.

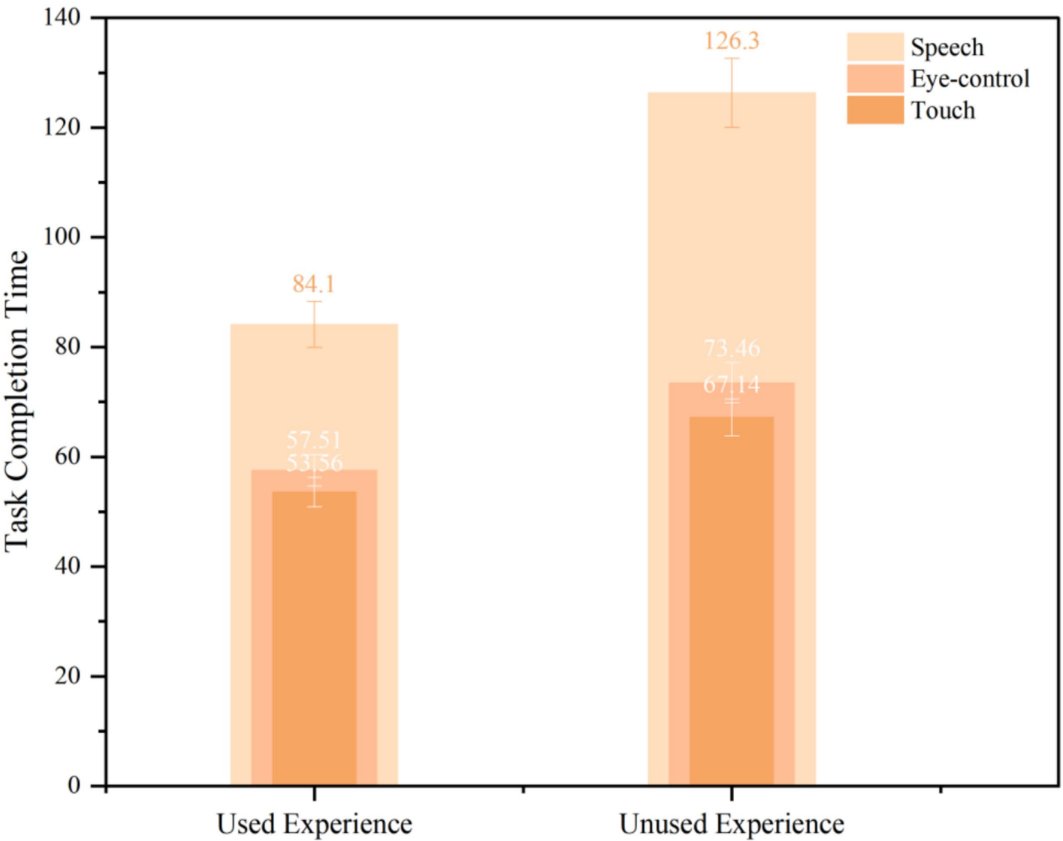


FIGURE 8 Multiple comparisons of task completion time according to users and input modalities.

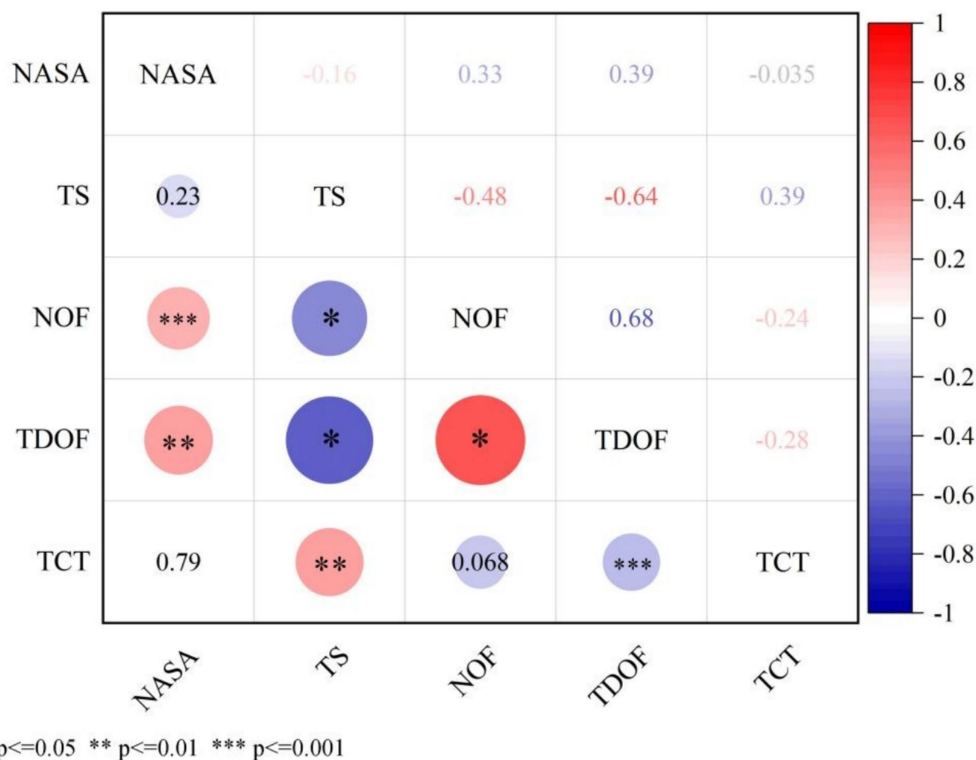


FIGURE 9

Pearson's correlation and significance for questionnaire data, eye tracking data, and task completion time data.

7.3 Eye track data

Experience of use ($F = 6.611$, $p = 0.03$, $\eta^2 = 0.424$) and input modality ($F = 47.478$, $p = 0.000$, $\eta^2 = 0.841$) had a differentially significant effect on total duration of fixation. There was no significant difference in the number of fixations for older adults' experience ($F = 3.227$, $p = 0.106$, $\eta^2 = 0.264$). Input modality significantly impacted the number of fixations ($F = 67.865$, $p = 0.000$, $\eta^2 = 0.883$) (Table 1). Older adults with usage experience had significantly shorter fixation times and fewer fixations for speech input ($M = 4.247$, $SD = 0.404$) ($M = 13.786$, $SD = 5.487$) compared to touch input ($M = 6.553$, $SD = 0.783$) ($M = 26.819$, $SD = 2.212$) and eye-control input ($M = 17.426$, $SD = 1.088$) ($M = 47.069$, $SD = 5.225$). Older adults with no usage experience had significantly shorter fixation times and fewer fixations for speech input ($M = 9.205$, $SD = 1.138$) ($M = 24$, $SD = 1.972$) and touch input ($M = 11.514$, $SD = 1.852$) ($M = 30.2$, $SD = 3.999$) compared to eye-control input ($M = 28.817$, $SD = 5.59$) ($M = 60.4$, $SD = 5.884$) (see Figure 7).

7.4 Task completion time

Older adults used experience ($F = 19.98$, $p = 0.002$, $\eta^2 = 0.689$) and input modality ($F = 55.489$, $p = 0.000$, $\eta^2 = 0.86$) to have a significant effect on task completion time (see Table 1). All older adults took significantly longer to complete the task using speech

input ($M = 84.088$, $SD = 3.279$) ($M = 126.302$, $SD = 7.933$) than touch ($M = 53.553$, $SD = 2.868$) ($M = 67.140$, $SD = 3.879$) and eye-control input ($M = 57.514$, $SD = 3.879$) ($M = 73.454$, $SD = 5.702$); there was no significant difference between touch and eye-control input in terms of task completion (see Figure 8).

7.5 Correlation analysis of questionnaire data, eyetracking data and task completion time data

As shown in Figure 9, a Pearson correlation analysis (Pearson, 1895) has been performed on five sets of data, with the first two sets of data being questionnaire data (subjective), the third and fourth sets of data being eye-tracking data (objective), and the fifth set of data being task completion time data (objective). From these three types of data, it can be concluded that NASA is positively correlated with NOF (0.33^{***} , $p < 0.001$) and TDOF (0.39^{**} , $p < 0.01$); TS is negatively correlated with NOF (-0.48^* , $p < 0.05$) and TDOF (-0.64^* , $p < 0.05$); NOF is positively correlated with TDOF (0.68^* , $p < 0.05$) was positively correlated; TDOF was negatively correlated with TCT (-0.28^{***} , $p < 0.001$). Correlation analyses further demonstrated the influence of older adults on the experience of trust with cognitive load. As the cognitive load increased, the eye movement data also increased. Higher eye-movement data indicated lower user trust.

8 Discussion

8.1 The difference in trust experience, cognitive load, and task performance

Significant differences were observed in trust perception, cognitive load, and task performance among older adults when using different input modalities (touch, speech, and eye control). Touch input elicited significantly higher user trust perception compared to speech and eye control inputs. This preference stems from touch interaction being a more prevalent and familiar input method for elderly users. As demonstrated by [Bong et al. \(2018\)](#), touch interfaces incorporate widely recognized graphical UI elements that enhance system accessibility and senior user acceptance. The long-term usage patterns have established a dependency effect, consequently elevating users' trust perception scores. Empirical evidence further confirms superior performance outcomes with touch input among elderly populations ([Sultana and Moffatt, 2019](#)).

Compared to touch and eye control, older adults have the lowest cognitive load when using speech. Older adults have the shortest total duration of fixation and the fewest number of fixations when using speech input; the duration of fixation and the number of fixations are the longest when using eye control input. As established by [Seaborn et al. \(2023\)](#), speech interaction leverages natural speech-based communication patterns, mirroring daily conversational flows. [Portet et al. \(2013\)](#) further confirmed speech input as a comfortable and low-stress modality, eliminating visual-motor coordination demands inherent to touch/eye control input. The latter modalities impose additional physical constraints—requiring precise finger movements or sustained visual attention—thereby reducing overall comfort and usability. However, existing literature notes that speech input necessitates continuous maintenance of command context (consuming 2.5 working memory chunks on average), whereas graphical interfaces provide visual persistence (e.g., button states) that offloads memory demands ([Baddeley, 2012](#)). The NASA-TLX scale shows that the cognitive load of speech input is 23% lower than that of touch input. For the elderly, with the decline of limb operation ability, the situations of accidental clicking and no feedback after clicking often occur, which increases the cognitive load of the elderly. Instead, voice can help the elderly better interact.

Among older adults, touch input yielded the shortest task completion time, while speech input resulted in the longest task completion time. This performance disparity stems from two fundamental factors: First, touch represents the most familiar interaction paradigm for older users, leveraging established mental models that enhance operational efficiency ([Bong et al., 2018](#)). Second, while speech enforces a linear dialog pattern (speak–listen–speak cycle) that creates sequential processing bottlenecks, touch permits parallel operations—enabling simultaneous finger gestures and visual scanning of interface elements ([Oviatt, 2022](#)).

8.2 Trust experience and cognitive load varied significantly among older adults with different levels of prior experience

Older adults with technological experience demonstrate greater trust in touch interfaces, cognitively mapping smartphone touch

interactions to conventional button operations ([Zhou et al., 2017](#)). Conversely, the inexperienced people exhibit strongest trust in speech input. This disparity stems from experience-mediated schema differentiation—seasoned users develop ingrained press-response mental models through prolonged use of tactile devices (e.g., feature phones, ATMs) ([Norman, 2013](#)), while novices adopt speech to bypass the cognitive demands of hierarchical UI navigation (typically 5–7 menu levels in touch systems), instead employing natural language expressions that mirror familiar conversational patterns.

The observed divergence further reflects generational asymmetries in technology adoption. Experienced users, having established digital self-efficacy through successful touch interactions, perceive speech as more cognitively demanding and question its reliability. Inexperienced users, while judging speech and touch as equally taxing cognitively, face a compound learning barrier with touch—simultaneously acquiring gesture vocabulary and interface logic—making speech comparatively more accessible despite equivalent perceived effort.

8.3 Trust experience, cognitive load, and task efficiency interacted significantly

The significant positive correlation among NASA-TLX scores, the duration of fixation, and the number of fixations demonstrates robust convergent validity in cognitive load measurement. This tripartite alignment confirms cross-modal consistency between subjective scales (NASA-TLX) and objective physiological metrics (eye-tracking data), satisfying the multitrait-multimethod matrix (MTMM) validation framework proposed by [Campbell and Fiske \(1959\)](#).

Further analysis reveals an inverse relationship between user trust perception and cognitive load—as cognitive demands increase, trust formation becomes progressively compromised. Neurocognitive evidence suggests that excessive task load consumption diminishes metacognitive monitoring capacity essential for trust establishment, indicating that trust development requires sufficient “cognitive slack.” Empirical thresholds show that when NASA-TLX scores exceed 60 points, trust assessment accuracy declines by approximately 42%, as demonstrated in controlled experiments (CHI Conference 2023).

8.4 From empirical findings to adaptive design strategies

To address the impracticality of direct user experience inquiries in real-world settings such as hospitals, future intelligent systems could infer user proficiency through continuous and low-intrusion analysis of micro-behavioral indicators. These metrics include interaction fluency (e.g., accuracy, hesitation time, task efficiency), error patterns, and exploration of advanced features. The system could initially operate in a default “guided mode” and automatically transition to an “advanced mode” upon detecting proficient behaviors. Furthermore, the system should be capable of continuous learning, dynamically adjusting interface complexity in response to evolving user proficiency. This approach directly accommodates the continuous nature of user experience,

moving beyond reliance on static binary classifications. To improve accessibility, public terminal interfaces (e.g., in hospital lobbies) should be preset with novice-friendly modes such as “voice-first” or “large-button touch,” while ensuring clear and readily available alternatives for mode switching.

We observed that trust levels among inexperienced users dropped sharply following speech recognition errors. This indicates that robust error tolerance is a more critical system requirement than achieving perfectly accurate user classification. Consequently, system design must prioritize mechanisms that effectively mitigate the negative consequences of interaction failures. An enhanced fault-tolerance approach should be implemented, including: providing clear command examples, offering contextual prompts after recognition failures, and ensuring seamless switching to alternative input modalities (e.g., one-touch fallback to a touch interface). Such a “safety net” design bridges experience gaps through inherent system qualities, ensuring interface resilience and robustness even without perfect user state awareness.

9 Conclusions and limitations

This study highlights the important role of input method and experience of use in influencing trust and cognitive load when using smart devices in older adults. It was found that experienced users preferred touch input because it fitted their existing mental models, while inexperienced users preferred voice input because of its natural interactive approach. Cognitive load plays a mediating role in this, and while voice input reduces initial learning costs, it also poses reliability issues. Design recommendations state that the touch input habits of experienced users should be preserved, whilst increasing voice input tolerance and system visibility for novices to increase trust and reduce usage anxiety.

The sample in this study may not be fully representative of all older populations, especially those with significant cognitive decline or limited access to technology. In addition, the controlled experimental environment may not reflect the complexities encountered when using speech recognition in the real world, such as the effects of environmental noise. Future research should explore longitudinal trust development and multimodal interaction design to better support diverse older users.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

References

- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422
- Basu, R. (2021). Age and Interface equipping older adults with technological tools. (doctoral dissertation). OCAD University.
- Bong, W. K., Chen, W., and Bergland, A. (2018). Tangible user interface for social interactions for the elderly: a review of literature. *Advances in Human-Computer Interaction* 2018:7249378. doi: 10.1155/2018/7249378
- Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 81–105. doi: 10.1037/h0046016
- Chêne, D., Pillot, V., and Bobillier Chaumon, M. É. (2016). Tactile interaction for novice user. In International conference on human aspects of IT for the aged population (pp. 412–423). Springer, Cham.
- Chi, O. H., Denton, G., and Gursoy, D. (2020). Artificially intelligent device use in service delivery: a systematic review, synthesis, and research agenda. *J. Hosp. Mark. Manag.* 29, 757–786. doi: 10.1080/19368623.2020.1721394
- Choi, J. K., and Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *Int. J. Hum. Comput. Interact.* 31, 692–702. doi: 10.1080/10447318.2015.1070549

Ethics statement

This study was reviewed and approved by the Inclusive User Experience Design Center of Ningbo University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

HH: Data curation, Formal analysis, Writing – original draft. GH: Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Claypoole, V. L., Schroeder, B. L., and Mishler, A. D. (2016). Keeping in touch: Tactile interface design for older users. *Ergonomics in Design* 24, 18–24. doi: 10.1177/1064804615611271
- De Raedemaeker, K., Foulon, I., Azzopardi, R. V., Lichtert, E., Buyl, R., Topsakal, V., et al. (2022). Audiometric findings in senior adults of 80 years and older. *Front. Psychol.* 13:861555. doi: 10.3389/fpsyg.2022.861555
- Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., and Bunge, S. A. (2017). Beyond eye gaze: what else can eye tracking reveal about cognition and cognitive development? *Dev. Cogn. Neurosci.* 25, 69–91. doi: 10.1016/j.dcn.2016.11.001
- Fisk, A. D., Czaja, S. J., Rogers, W. A., Charness, N., and Sharit, J. (2020). Designing for older adults: Principles and creative human factors approaches. CRC Press. doi: 10.1201/9781420080681
- Gefen, D., Karahanna, E., and Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Q.* 27, 51–90. doi: 10.2307/30036519
- Ghazizadeh, M., Lee, J. D., and Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cogn. Tech. Work* 14, 39–49. doi: 10.1007/s10111-011-0194-3
- Guo, S., Falaki, M. H., Oliver, E. A., Ur Rahman, S., Seth, A., Zaharia, M. A., et al. (2007). Very low-cost internet access using KioskNet. *ACM SIGCOMM Comput. Commun. Rev.* 37, 95–100. doi: 10.1145/1290168.1290181
- Hansen, J. P., Hansen, D. W., and Johansen, A. S. (2001). “Bringing gaze-based interaction back to basics” in HCI, 325–329.
- Hart, S. G., and Staveland, L. E. (1988). “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research” in Human mental workload. eds. P. A. Hancock and N. Meshkati, vol. 5 (Amsterdam: North-Holland), 139–183.
- Hengstler, M., Enkel, E., and Duelli, S. (2016). Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices. *Technol. Forecast. Soc. Change* 105, 105–120. doi: 10.1016/j.techfore.2015.12.014
- Hoff, K. A., and Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Hum. Factors* 57, 407–434. doi: 10.1177/0018720814547570
- Jian, J. Y., Bisantz, A. M., and Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* 4, 53–71. doi: 10.1207/S15327566IJCE0401_04
- Johnston, M., and Bangalore, S. (2004). MATCHKiosk: a multimodal interactive city guide. In Proceedings of the ACL interactive poster and demonstration sessions (pp. 222–225).
- Just, M. A., and Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cogn. Psychol.* 8, 441–480. doi: 10.1016/0010-0285(76)90015-3
- Kaufman, A. E., Bandopadhyay, A., and Shaviv, B. D. (1993). An eye tracking computer user interface. In Proceedings of 1993 IEEE research properties in virtual reality symposium (pp. 120–121). IEEE.
- Kaye, S. A., Lewis, I., and Freeman, J. (2018). Comparison of self-report and objective measures of driving behavior and road safety: a systematic review. *J. Saf. Res.* 65, 141–151. doi: 10.1016/j.jsr.2018.02.012
- Kim, J. C., Laine, T. H., and Åhlund, C. (2021). Multimodal interaction systems based on internet of things and augmented reality: a systematic literature review. *Appl. Sci.* 11:1738. doi: 10.3390/app11041738
- Kim, M., Lee, M. K., and Dabbish, L. (2015). Shop-i: gaze based interaction in the physical world for in-store social shopping experience. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1253–1258).
- Lee, Y., Jeon, H., Kim, H. K., and Park, S. (2020). Literature review on accessibility guidelines for self-service terminals. Proceedings of the ACHI.
- Lee, J. D., and See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Hum. Factors* 46, 50–80. doi: 10.1518/hfes.46.1.50.30392
- Leonardi, C., Albertini, A., Pianesi, F., and Zancanaro, M. (2010). An exploratory study of a touch-based gestural interface for elderly. In Proceedings of the 6th nordic conference on human-computer interaction: Extending boundaries (pp. 845–850).
- Likert, R. (2017). “The method of constructing an attitude scale” in Scaling (London and New York: Routledge), 233–242.
- Lowndes, B. R., Forsyth, K. L., Blocker, R. C., Dean, P. G., Truty, M. J., Heller, S. F., et al. (2020). NASA-TLX assessment of surgeon workload variation across specialties. *Ann. Surg.* 271, 686–692. doi: 10.1097/SLA.0000000000003058
- Ma, J., Zuo, Y., Du, H., Wang, Y., Tan, M., and Li, J. (2025). Interactive output modalities Design for Enhancement of user trust experience in highly autonomous driving. *Int. J. Hum. Comput. Interact.* 41, 6172–6190. doi: 10.1080/10447318.2024.2375697
- Manzke, J. M., Egan, D. H., Felix, D., and Krueger, H. (1998). What makes an automated teller machine usable by blind users? *Ergonomics* 41, 982–999. doi: 10.1080/001401398186540
- Mikulski, D. G. (2014). Trust-based controller for convoy string stability. In 2014 IEEE symposium on computational intelligence in vehicles and transportation systems (CIVTS) (pp. 69–75). IEEE.
- Niu, Y. F., Gao, Y., Zhang, Y. T., Xue, C. Q., and Yang, L. X. (2019). Improving eye-computer interaction interface design: Ergonomic investigations of the optimum target size and gaze-triggering dwell time. *J. Eye Mov. Res.* 12. doi: 10.16910/jemr.12.3.8
- Norman, D. (2013). The Design of Everyday Things: Revised and Expanded Edition. New York: Basic Books.
- Olsen, A. (2012). The Tobii I-VT fixation filter: algorithm description. *Tobii Technol.* 21:15.
- Oviatt, S. (2022). “Multimodal interaction, interfaces, and analytics” in Handbook of human computer interaction (Cham: Springer International Publishing), 1–29.
- Paradi, J. C., and Ghazarian-Rock, A. (1998). A framework to evaluate video banking kiosks. *Omega* 26, 523–539. doi: 10.1016/S0305-0483(97)00080-7
- Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. *Ergonomics* 43, 931–951. doi: 10.1080/001401300409125
- Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* 58, 240–242. doi: 10.1098/rsp.1895.0041
- Pfeuffer, K., Abdrabou, Y., Esteves, A., Rivu, R., Abdelrahman, Y., Meitner, S., et al. (2021). ARtention: a design space for gaze-adaptive user interfaces in augmented reality. *Comput. Graph.* 95, 1–12. doi: 10.1016/j.cag.2021.01.001
- Poole, A., and Ball, L. J. (2006). “Eye tracking in HCI and usability research” in Encyclopedia of human computer interaction (Hershey, PA: IGI Global Scientific Publishing), 211–219.
- Pop, V. L., Shrewsbury, A., and Durso, F. T. (2015). Individual differences in the calibration of trust in automation. *Hum. Factors* 57, 545–556. doi: 10.1177/0018720814564422
- Portet, F., Vacher, M., Golanski, C., Roux, C., and Meillon, B. (2013). Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Pers. Ubiquit. Comput.* 17, 127–144. doi: 10.1007/s00779-011-0470-5
- Roth, T. N., Hanebuth, D., and Probst, R. (2011). Prevalence of age-related hearing loss in Europe: a review. *Eur. Arch. Otorhinolaryngol.* 268, 1101–1107. doi: 10.1007/s00405-011-1597-8
- Rozado, D., Agustin, J. S., Rodriguez, F. B., and Varona, P. (2012). Gliding and saccadic gaze gesture recognition in real time. *ACM Transactions on Interactive Intelligent Systems (TiIS)* 1, 1–27. doi: 10.1145/2070719.2070723
- Sajji, M., Bundalo, D., Vidović, Ž., Bundalo, Z., and Lalic, D. (2021). Smart digital terminal devices with speech recognition and speech control. In 2021 10th Mediterranean conference on embedded computing (MECO) (pp. 1–5). IEEE.
- Sandnes, F. E., Tan, T. B., Johansen, A., Sulic, E., Vesterhus, E., and Iversen, E. R. (2012). Making touch-based kiosks accessible to blind users through simple gestures. *Universal Access Inf. Soc.* 11, 421–431. doi: 10.1007/s10209-011-0258-4
- Sato, H., Abe, K., Ohi, S., and Ohshima, M. (2018). A text input system based on information of voluntary blink and eye-gaze using an image analysis. *Electr. Commun. Japan* 101, 9–22. doi: 10.1002/ecj.12025
- Seaborn, K., Sekiguchi, T., Tokunaga, S., Miyake, N. P., and Otake-Matsuura, M. (2023). Voice over body? Older adults’ reactions to robot and voice assistant facilitators of group conversation. *Int. J. Soc. Robot.* 15, 143–163. doi: 10.1007/s12369-022-00925-7
- Shang, H., Shi, X., and Wang, C. (2020). Research on the design of medical self-service terminal for the elderly. In Advances in human factors and ergonomics in healthcare and medical devices: proceedings of the AHFE 2020 virtual conference on human factors and ergonomics in healthcare and medical devices, July 16–20, 2020, USA (pp. 335–349). Springer International Publishing.
- Slay, H., Wentworth, P., and Locke, J. (2006). BingBee, an information kiosk for social enablement in marginalized communities. In Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries (pp. 107–116).
- Sultana, A., and Moffatt, K. (2019). Effects of aging on small target selection with touch input. *ACM Transactions on Accessible Computing (TACCESS)* 12, 1–35.
- Vildjiounaite, E., Mäkelä, S. M., Lindholm, M., Riihimäki, R., Kyllönen, V., Mäntyjärvi, J., et al. (2006). Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices. In International Conference on Pervasive Computing (pp. 187–201). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ya-feng, N., Jin, L., Jia-qi, C., Wen-jun, Y., Hong-rui, Z., Jia-xin, H., et al. (2022). Research on visual representation of icon colour in eye-controlled systems. *Adv. Eng. Inform.* 52:101570. doi: 10.1016/j.aei.2022.101570
- Zhang, T., Liu, X., Zeng, W., Tao, D., Li, G., and Qu, X. (2023). Input modality matters: a comparison of touch, speech, and gesture based in-vehicle interaction. *Appl. Ergon.* 108:103958. doi: 10.1016/j.apergo.2022.103958
- Zhou, J., Chourasia, A., and Vanderheiden, G. (2017). Interface adaptation to novice older adults’ mental models through concrete metaphors. *Int. J. Hum. Comput. Interact.* 33, 592–606. doi: 10.1080/10447318.2016.1265827