A Review of Research on the Integrated Application and User Experience of Multimodal Human-Computer Interaction in Smart Home Systems for Older Adults

Zhen Zhou
Electronic Information Engineering, Putian University, Putian, China

Abstract: As the global population continues to age, smart home systems are playing an increasingly important role in ensuring the independent living of older adults and improving their quality of life . However, the common decline in vision, hearing, touch, and movement abilities of older adults makes it difficult for traditional single-modal interaction methods to meet their actual use needs . This article uses a literature review method to systematically retrieve and analyze research results on single modalities such as voice interaction, touch interaction, and graphical interfaces in the field of smart homes for older adults, as well as research on the integration practice and user experience evaluation of multimodal interaction. In order to enhance the focus of the research, this article limits the scope of analysis to three typical scenarios: environmental control, health management and social entertainment. On this basis, three multi-modal integration modes are summarized: complementary, redundant and adaptive.

Keywords: Multimodal Human-Computer Interaction; Older Adults; Smart Home Systems; User Experience; Age-Friendly Design

1. Introduction

The global population is aging. The United Nations population projections and the World Health Organization's research on healthy aging both point out that the size and proportion of the older adult population will continue to increase by the middle of this century, which will make independent home living and aging-appropriate technological support a more important social issue [1,2]. In this context, independent home living for older adults has become a key design challenge. Smart home systems can provide key guarantees for the safety and quality of life of older adults in terms of health monitoring, life assistance, and emergency response [3,4]. Therefore, improving the system usage experience of older adults through effective human-computer interaction design has become a core research direction in the field of human-computer interaction.

However, a widely ignored reality is that a large number of older adult users suffer from varying degrees of sensory decline - blurred vision, hearing loss, hand tremors or joint stiffness. These natural changes in body functions make it difficult for them to independently and effectively control smart home devices through traditional single-modal interaction methods. This problem is particularly prominent among older adults living alone, which directly affects their home safety and quality of life.

Judging from existing research, scholars have extensively explored various interaction methods in smart home systems, such as voice interaction, touch interaction, and graphical interface design. However, these single-modal interaction methods have significant limitations in the older adult population. Voice interaction is limited by the pronunciation intelligibility, dialect differences and environmental noise interference of older adults, and cannot effectively transmit spatial or orientation information; touch interaction relies on fine hand movement control, and the common hand tremors or joint stiffness of older adults make it difficult to complete precise clicking or sliding operations [5,6]; graphical interfaces require high visual acuity and cognitive processing capabilities, and hierarchical menus and abstract icons often cause older adult users to get lost or misoperate [7]. It is precisely because of these limitations that some research begins to focus on multimodal human-computer interaction, trying to enhance the flexibility and accessibility of the system by combining multiple interaction methods [8,9].

Although existing research has made some progress, there are still obvious deficiencies in the existing literature. First of all, the research perspectives on multimodal interaction are scattered and lack unified design standards and evaluation systems. In addition, older adult users have significant individual differences. A systematic and generalizable multimodal interaction integration framework has not yet been formed. Secondly, existing research generally focuses on technical implementation and pays insufficient attention to the experience, usability and understanding process of older adults in real use situations. Thirdly, many studies treat smart home as a general whole and fail to conduct differentiated analysis for the specific needs of different usage scenarios. In fact, there may be significant differences in the interaction needs and pain points of older adults in scenarios such as environmental control (such as adjusting lights, temperature control), health management (such as medication reminders, emergency calls), and social entertainment (such as video calls, media playback), and need to be examined separately.

In response to the above research gaps, this article uses a literature review method to conduct a systematic analysis of multimodal human-computer interaction in smart home systems for older adults. The research focuses on the following three key questions:

RQ1: How can the integration rules of multimodal interaction methods be summarized from existing research to form a classifiable analytical framework?

RQ2: How can the actual impact of different multimodal integration models on older adult users' experience be evaluated, especially regarding applicability across specific usage scenarios?

RQ3: How can the research findings be transformed into executable design guidelines for specific smart home scenarios?

In order to enhance the focus and practical guidance value of the research, this article limits the scope of analysis to the three types of scenarios that are most frequently used by older adults and have the most prominent requirements for interaction reliability: environmental control, health management, and social entertainment. By systematically reviewing the development status of interaction methods such as voice, touch, and graphical interfaces and their applicability among older adults, this article conducts an in-depth analysis of the user characteristics of older adults in the context of sensory decline, summarizes the integration models of multimodal interaction, and comprehensively evaluates the user experience from dimensions such as usability, cognitive load, and operational efficiency. Ultimately, this article aims to refine multimodal interaction design suggestions for smart home systems for older adults, and provide theoretical reference and practical guidance for the design and optimization of subsequent related systems.

2. Analysis of Sensory Decline Characteristics and Interaction Needs of Older Adults

The natural decline of older adults' sensory systems is the fundamental reason that affects their interaction with smart home systems. Studies have shown that the visual acuity, auditory discrimination, tactile sensitivity and fine motor control abilities of older adults will change with age, and these changes will reduce their ability to process low-contrast text, small icons, fast voice prompts and precise touch operations [7,10-11]. The degradation of hearing ability also limits the interactive experience. Older adults's ability to perceive high-frequency sounds is significantly reduced, and their ability to analyze fast speech is weakened, resulting in some older adult users being unable to hear the operating instructions clearly. As age increases, the number of tactile receptors on the finger skin decreases, causing the tactile threshold to increase.

At the cognitive level, older adults have reduced working memory capacity and reduced ability to process complex information. This means that menus that are too deep, multi-step operations, and command words that need to be memorized will significantly increase the cognitive load. In environmental control scenarios (such as adjusting air conditioning temperature, turning on and off lights), older adults need simple and direct interaction paths; in health management scenarios (such as medication reminders, emergency calls), they have extremely high requirements for the reliability and fault tolerance of information; in social entertainment scenarios (such as video calls, media playback), they hope that the interaction method can adapt to fluctuations in their sensory abilities. These characteristics together form the basis of user needs for multimodal interaction design.

3. Limitations of Single-Modal Interaction

3.1 Voice Interaction

Voice interaction is regarded as one of the most natural ways of interaction, but it faces multiple challenges among older adults. First of all, the pronunciation intelligibility of older adults is affected by physiological factors, and coupled with dialect differences, the error rate of the automatic speech recognition system may increase. Second, voice interaction requires users to remember specific command words or sentence patterns, and the memory decline of older adults makes this requirement a burden. Research on smart speakers and personal voice assistants shows that older adult users recognize their convenience and companionship value, but they are still affected by factors such as recognition accuracy, privacy concerns, command learning costs and contextual noise [12,13].

3.2 Touch Interaction

Touch interaction relies on fine hand movement control. However, hand tremors, joint stiffness, or reduced skin touch sensitivity common in older adults make it difficult to complete precise tapping, swiping, or long-pressing operations. Existing touch screen research shows that age-related differences in motion control will affect gesture input efficiency and error rates, and interface target size, input methods and feedback design will significantly change the performance of older adult users [5,6]. In addition, touch interaction requires visual gaze and cannot achieve "blind operation", which is an additional obstacle for older adults with poor vision. In an emergency situation (such as calling for help after a fall), interfaces that rely on precise touch operations may fail completely.

3.3 Graphical Interface

Graphical interfaces (GUIs) present information through icons, menus and visual feedback, but require high visual acuity and cognitive processing capabilities. Hierarchical menus and abstract icons often lead to disorientation or misunderstanding among older users. In addition, small fonts, low contrast, and complex navigation structures will significantly reduce the operating efficiency of older adults; interface design for older adult users usually requires higher contrast, clearer feedback, fewer levels, and a more stable navigation structure [7,11]. In health management scenarios, if the medication reminder interface contains too many layers, older adults are likely to miss key information.

The limitations of the above single modality indicate that no single interaction method can independently meet the needs of older adults in all scenarios. This provides sufficient basis for the necessity of the integration of multi-modal human-computer interaction.

4. Integrated Model of Multimodal Human-computer Interaction

Multimodal human-computer interaction provides older adults with a more flexible and tolerant operation method by using multiple channels such as voice, touch, gesture, vision, and tactile simultaneously or alternately. Multimodal interface research points out that the combination of different sensing and input channels can improve system robustness, expression efficiency and barrier-free adaptation capabilities, but only if the information distribution and feedback relationship between each channel are clear [8,9]. According to the integration logic of existing research, this article summarizes three typical integration models: complementary, redundant and adaptive.

4.1 Complementary Integration Model

The complementary mode means that different interaction methods each undertake different sub-tasks, complement each other, and collaborate to complete a complete operation. In smart home systems for older adults, the complementary mode is often used in environmental control scenarios. For example, the user first says "turn on the lights in the living room" with voice, and after the system responds, the brightness is accurately adjusted by touching the slider. Voice is responsible for fast, coarse-grained command input, and touch is responsible for fine, continuous control. The two complement each other. The key to the complementary model is the rationality of task decomposition - each mode should assume the function it is best at.

4.2 Redundant Integration Model

The redundant mode refers to multiple interaction methods transmitting the same information at the same time or one after another, forming information redundancy to improve the reliability of information and user acceptance rate. This model is of great value in health management scenarios. For example, medication reminders can be presented in four ways at the same time: voice broadcast, screen text display, pill box light flashing, and mobile phone push notifications. Research on multimodal reminders in assisted living technology shows that a combination of visual, auditory, and other reminder modalities can help improve reminder accessibility, but the number and timing of reminders need to be controlled to avoid interference or overload [14]. Even if older adults cannot hear speech due to hearing loss, they can still see the text on the screen; even if their vision is blurred and they cannot see the text clearly, they can still perceive the flickering of lights.

4.3 Adaptive Integration Model

Adaptive mode means that the system dynamically adjusts the most appropriate combination of interaction methods based on the user's status, situation, or preferences. This model has the highest technical requirements, but is also most consistent with the fluctuating abilities of older adults. Research on co-creation of smart home interfaces shows that older adult users are more likely to form trust in interaction methods that are understandable, controllable, and consistent with daily situations. The system should avoid automatically switching key interaction logic when the user cannot understand it [4]. For example, the system gives priority to touch interaction (accurate and efficient) when there is sufficient light during the day, and automatically switches to voice interaction at night or when the light is dim (to avoid operating in the dark); after the system detects that the user's recent operation speed has slowed down and the rate of false touches has increased, it actively simplifies the interface, enlarges the buttons, extends the response time, and asks whether to switch to "Elder Mode". The core challenge of adaptive mode is how to accurately perceive the user's status and context and make appropriate mode switching decisions.

4.4 Applicability of the Three Modes in Typical Scenarios

Table 1. Applicability of Three Multimodal Integration Modes in Typical Smart Home Scenarios

Scenario	Recommended integration model	Rationale
Environmental control (temperature adjustment, dimming)	complementary	Quick target selection by voice + precise adjustment by touch
Health management (medication reminders, emergency calls)	redundant	Multiple confirmations ensure that information is not missed
Social entertainment (video calls, media playback)	Adaptive	Dynamically adjust the interface and feedback methods based on sensory capabilities

As shown in Table 1, in actual systems, the three modes can be mixed. For example, a smart home system adopts a complementary mode in regular operation, superimposes a redundant mode during critical reminders, and realizes adaptive adjustment according to user changes during long-term use.

5. Dimensions and Methods of User Experience Evaluation

User experience evaluation of multimodal interactions needs to be conducted from multiple dimensions. Based on existing research, this article summarizes three core evaluation dimensions:

Usability: Including operation efficiency (the time required to complete a task), error rate (the number of incorrect operations) and learning cost (the number of attempts required from the first use to proficiency).

Cognitive load: refers to the attention and memory resources that users need to invest during use. Multimodal interaction can theoretically reduce cognitive load by decentralizing information presentation channels; however, multimodality does not mean that more information is better. If there is a lack of hierarchy, timing, and consistency between channels, it may lead to distraction and increased understanding costs [8,9].

Adaptability: For the adaptive mode, it is also necessary to evaluate the system's switching accuracy in different situations and the user's acceptance of automatic switching. At present, there are relatively few studies on this dimension, which is an important direction for future research.

6. Existing Research Deficiencies and Future Directions

Although the application of multimodal human-computer interaction in smart home systems for older adults has made some progress, existing research still has the following obvious shortcomings:

First, there is a lack of a unified multimodal integration framework. Most studies adopt an "improvisational" combination method and lack the abstraction and theorization of integration logic. The complementary, redundant, and adaptive classification proposed in this article is only a preliminary attempt, and more empirical research is needed to verify its explanatory power and guiding value.

Second, user experience evaluation focuses on short-term, laboratory environments. Existing studies mostly use one-time laboratory tests, which are difficult to reflect the learning effects, habit formation and fatigue changes of older adults during long-term use. Technical research on home scenarios has suggested that usability, acceptability and trust building in real home environments are often different from laboratory results, and home field studies lasting several weeks or even months will need to be carried out in the future [3,4].

Third, insufficient attention is paid to the differences within the older adult user group. There are huge differences in the degree of sensory decline, cognitive ability, and technology acceptance among older adults. Existing research often treats "older adults" as a homogeneous group and lacks segmentation analysis of users with different degrees of decline; technology design for older adults should consider age, experience, health status, and environmental support as common influencing factors [1,11].

Fourth, scenario differentiation analysis is weak. Although this article clearly distinguishes three types of scenarios: environmental control, health management, and social entertainment, there are few studies in the existing literature that compare the effects of integration models in different scenarios under the same framework. Future research needs to design and verify multi-modal integration solutions for different scenarios.

Fifth, the intelligent decision-making mechanism of the adaptive mode is immature. The adaptive integration mode requires the system to perceive the user status and situation in real time and make appropriate mode switching. At present, most research in this area remains at the proof-of-concept stage, and there is still a large gap between actual deployment.

7. Design Suggestions and Conclusions

Based on the above analysis, this article proposes multi-modal interaction design suggestions for smart home systems for older adults:

Compensating for sensory decline as a core goal: Design should prioritize compensating for vision, hearing, touch, and cognitive decline in older adults rather than pursuing technological novelty.

Scenario-driven selection of integration models: Complementary models are preferred in environmental control scenarios, redundant models are preferred in health management scenarios, and adaptive models are explored in social entertainment scenarios.

Provide autonomy in mode selection and switching: allow older adult users to choose or close certain interaction modes according to their own preferences and abilities, avoiding "technological determinism".

Simplify the presentation of redundant information: In the redundant mode, cognitive overload caused by simultaneous bombardment of multi-modal information should be avoided. Sequential presentation or clear priority can be used.

Carry out long-term, real-situation user evaluation: Future research should conduct longitudinal tracking in real home environments, collect interaction behavior data of older adult users in different seasons and different health conditions, and combine interviews, usability tasks and log data to evaluate the reliability, acceptability and burden changes of multi-modal systems in long-term use [3,4].

Conclusion: This article uses the literature review method to systematically sort out the integration model and user experience research of multi-modal human-computer interaction in smart home systems for older adults. By focusing on the core social issue of sensory decline and limiting the three typical scenarios of environmental control, health management, and social entertainment, three integration models are summarized: complementary, redundant, and adaptive. The analysis shows that multimodal interaction has significant potential in compensating for sensory decline in older adults and improving operational efficiency and subjective satisfaction. However, existing research still faces challenges such as lack of framework, biased evaluation, and insufficient scene differentiation. The design suggestions put forward in this article are intended to provide theoretical reference for subsequent research and system development. Future research should further explore the intelligent decision-making mechanism of the adaptive mode and verify the long-term effects of different integration modes in real home environments.

References

[1] World Health Organization. (2015). World report on ageing and health. World Health Organization. https://www.who.int/publications/i/item/9789241565042

[2] United Nations Department of Economic and Social Affairs, Population Division. (2022). World population prospects 2022: Summary of results. United Nations. https://www.un.org/development/desa/pd/content/World-Population-Prospects-2022

[3] Charness, N., Best, R., & Evans, J. (2016). Supportive home health care technology for older adults: Attitudes and implementation. Gerontechnology, 15(4), 233-242. https://doi.org/10.4017/gt.2016.15.4.006.00

[4] Ghorayeb, A., Comber, R., & Gooberman-Hill, R. (2023). Development of a smart home interface with older adults: Multi-method co-design study. JMIR Aging, 6, e44439. https://doi.org/10.2196/44439

[5] Findlater, L., Froehlich, J. E., Fattal, K., Wobbrock, J. O., & Dastyar, T. (2013). Age-related differences in performance with touchscreens compared to traditional mouse input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 343-346). ACM. https://doi.org/10.1145/2470654.2470703

[6] Gao, Q., & Sun, Q. (2015). Examining the usability of touch screen gestures for older and younger adults. Human Factors, 57(5), 835-863. https://doi.org/10.1177/0018720815581293

[7] Hawthorn, D. (2000). Possible implications of aging for interface designers. Interacting with Computers, 12(5), 507-528. https://doi.org/10.1016/S0953-5438(99)00021-1

[8] Larson, J. A., & Oviatt, S. (2004). Guidelines for multimodal user interface design. Communications of the ACM, 47(1), 39-43. https://doi.org/10.1145/962081.962106

[9] Oviatt, S., & Cohen, P. (2000). Perceptual user interfaces: Multimodal interfaces that process what comes naturally. Communications of the ACM, 43(3), 45-53. https://doi.org/10.1145/330534.330538

[10] Carmeli, E., Patish, H., & Coleman, R. (2003). The aging hand. The Journals of Gerontology: Series A, 58(2), M146-M152. https://doi.org/10.1093/gerona/58.2.M146

[11] Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009). Designing for older adults: Principles and creative human factors approaches (2nd ed.). CRC Press.

[12] Arnold, A., Kolody, S., Comeau, A., & Cruz, A. M. (2024). What does the literature say about the use of personal voice assistants in older adults? A scoping review. Disability and Rehabilitation: Assistive Technology, 19(1), 100-111. https://doi.org/10.1080/17483107.2022.2041246

[13] Kim, S. (2021). Exploring how older adults use a smart speaker-based voice assistant in their first interactions: Qualitative study. JMIR mHealth and uHealth, 9(1), e20427. https://doi.org/10.2196/20427

[14] Warnock, D., McGee-Lennon, M., & Brewster, S. (2012). Older users, multimodal reminders and assisted living technology. Health Informatics Journal, 18(3), 181-190. https://doi.org/10.1177/1460458212440979

