Toward Enabling Natural Conversation with Older Adults via the Design of LLM-Powered Voice Agents that Support Interruptions and Backchannels

Chao Liu
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
cliu009@connect.hkust-gz.edu.cn

Mingyang Su
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
sumy22@mails.tsinghua.edu.cn

Yan Xiang
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
yxiang411@connect.hkust-gz.edu.cn

Yuru Huang
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
yhuang760@connect.hkust-gz.edu.cn

Yiqian Yang
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
yyang937@connect.hkust-gz.edu.cn

Kang Zhang
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
kzhangcma@hkust-gz.edu.cn

Mingming Fan*
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
The Hong Kong University of Science and Technology
Hong Kong, China
mingmingfan@ust.hk

ABSTRACT

Voice agents can construct meaningful conversations with older adults to offer various benefits, such as providing emotional companionship and assisting with memory recall. However, such conversations often follow the simple turn-taking pattern and lack interruption and backchannel of natural human conversation. Previous research has shown that this rigid turn-taking pattern lacks interactivity and initiative, limiting the flexible communication between older adults and voice agents. To address these issues and create a more natural conversational voice agent, we first conducted a formative study to identify common usage of interruption in the natural conversations of older adults. We then designed an LLM-powered Barge-in agent that supports interruption and backchannel. Our within-subject exploratory study showed that participants felt that conversations with Barge-in agents were more natural, engaging, and fluent than with the No barge-in agent. We further present design implications for creating more natural and human-like voice agents for older adults.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility technologies**; **Empirical studies in HCI**.

KEYWORDS

Voice agent for older adults, Elderly care technology, Interruption, Backchannel, VUI

ACM Reference Format:
Chao Liu, Mingyang Su, Yan Xiang, Yuru Huang, Yiqian Yang, Kang Zhang, and Mingming Fan. 2025. Toward Enabling Natural Conversation with Older Adults via the Design of LLM-Powered Voice Agents that Support Interruptions and Backchannels. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3706598.3714228

*Corresponding Author

CHI '25, April 26-May 1, 2025, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04…$15.00
https://doi.org/10.1145/3706598.3714228

1 INTRODUCTION

Social isolation and loneliness are major risk factors for many older adults that negatively affect both physical and mental health, leading to conditions like depression, reduced quality of life, and even increased mortality [27, 54, 64, 65]. With the continuous progress of voice user interfaces, some voice agents designed for older adults are emerging. For example, Alexa's Ask My Buddy and Google

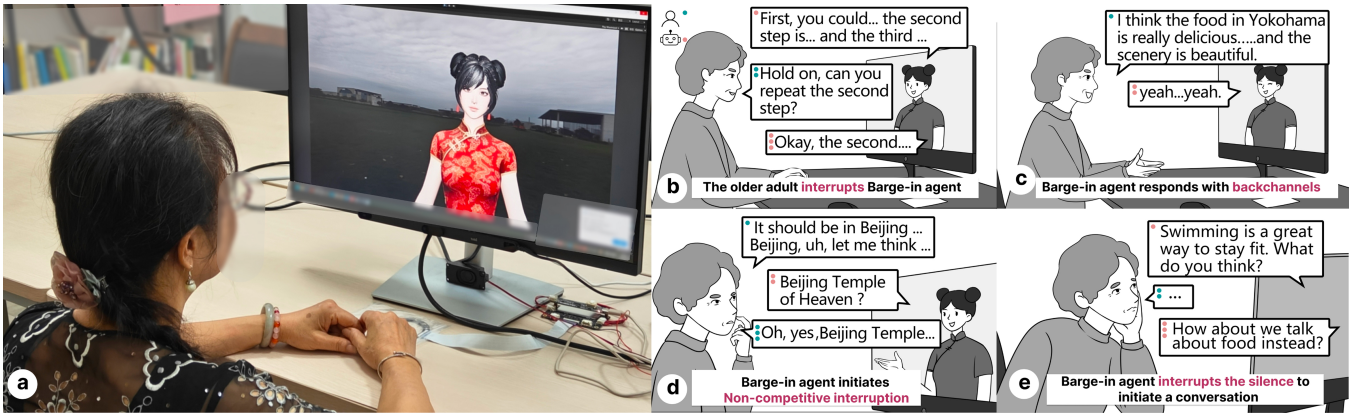

Figure 1: (a) The participant was engaging in a conversation with the Barge-in voice agent designed for older adults, which supports interruptions and backchannels; (b) The older adult interrupts the Barge-in Agent and asks it to repeat what it has not heard; (c) While the older adult is speaking, the Barge-in Agent responds with backchannels like "yeah, yeah" as a listening response; (d) The Barge-in agent initiates a co-operative interruption to help the older adult complete the conversation; (e) The Barge-in Agent detects the older adult's silence and actively interrupts it by initiating a conversation.

Assistant's Vigil Connect, both developed based on traditional language models, can provide certain levels of emotional companionship help for older adults [37, 50, 51, 76, 82]. Compared with text conversation, older adults prefer direct voice interaction for conversations because it is more intuitive than touchscreens and keyboards, requiring neither precise hand-eye coordination nor extensive learning. This ease of use also makes it simpler for older adults to access information and services through voice agents [4, 75].

Prior to the advent of large language models (LLMs), academic research on voice agents designed for older adults primarily focused on task-oriented interactions, such as information search, reminders, and other specific tasks. Conversational companionship was typically a secondary function, as these traditional voice agents relied on limited language models that lacked the complexity and depth needed for more dynamic, human-like interactions [58, 67, 70, 88, 101]. With the rapid development of LLMs, such as OpenAI's ChatGPT, agents have gained significant improvements in the ability to generate contextually aware and natural conversations. These LLM-powered agents have narrowed the gap between human language and machine-generated responses, resulting in more fluid, natural interactions [91]. Furthermore, multimodal models like OpenAI's Whisper, Google's AudioPaLM, and Microsoft's Azure Cognitive Services have enabled voice agents to produce speech outputs with richer emotional and rhythmic tones [11, 81]. These advances present new opportunities for creating voice agents tailored to the needs of older adults, which may be designed to offer more personalized and socially aware interactions.

Despite significant technical advancements in natural language processing, there are still notable challenges in developing voice agents that effectively engage older adults. One of the more critical issues is that current voice agents still primarily rely on rigid turn-taking protocols, commonly called the "speak-wait/speak-wait" mode, in which one party speaks while the other waits. This linear approach contrasts sharply with natural human-human conversations, which often involve interruptions and backchannels that maintain the flow of dialogue [3, 6, 91]. The traditional voice agent presents four notable limitations for older adults. Firstly, during user turns, it necessitates the construction of complete sentences and full organization of thoughts prior to speaking. However, older adults often prefer utilizing shorter, less complex sentences and frequently rely on ambiguous expressions [13, 52, 86, 107]. Secondly, during the voice agent's turns, the decline in hearing and communication abilities among many older adults hinders their effective comprehension of complex conversations [84]. Consequently, they often require questions or repetitions to ensure accurate information grasp. Nevertheless, the model typically employs a simplex communication protocol, which is a one-way communication method where the system only receives input from the user without permitting real-time interruptions or corrections, thereby impeding this function [107]. Thirdly, older adults require more time to adapt and are prone to making mistakes when using traditional turn-taking agents [97]. Their tendency to speak more slowly can lead to the system incorrectly assuming the end of input during pauses within a sentence, resulting in incomplete or failed conversations. Moreover, the inability to interrupt or correct the voice agent in a timely manner can cause frustration [49, 104, 107]. Lastly, current voice agents in the voice modal commonly lack backchannels, leading to a deficiency in interactivity and sociability [17, 31, 44]. Additionally, these systems frequently utilize simple Voice Activity Detection (VAD) to detect the conclusion of a user's speech, and when the user remains silent, the voice agent does not initiate further conversation. This silence can be more readily triggered in older adults compared to younger individuals due to their unfamiliarity with the system or incorrect usage [60, 96].

Furthermore, while some products like Hume's EVI and Moshi AI allow users to interrupt agents, these interruptions remain largely one-sided (only the user interrupts the machine), with no mechanism for the voice agent to initiate or manage interruptions, limiting

the interaction's naturalness and fluidity. More importantly, these designs do not specifically consider the needs of older adults, who may require more accessible interactions [83] .

To address these issues, we sought to explore how to design a voice agent, empowered by LLMs, that supports interruptions and backchannels for older adults, while also examining the natural conversational experiences it provides to older adults. Specifically, we investigated the two research questions (RQs):

- **RQ1:** What design considerations are needed for a voice agent that supports interruptions and backchannels for older adults?
- **RQ2:** What kind of conversation experience will a voice agent that supports interruptions and backchannels bring to older adults?

To address RQ1, we conducted a formative study where older adults ($N = 8$) engaged in human-to-human conversations, retrospective interviews, and interactions with a turn-taking LLM agent. We gathered participants' feedback, which provided insights into their conversational behaviors and expectations for agents that can support such dynamics.

Based on the interruption classification by Murata et al. [62], we analyzed the attitudes and evaluations of older adults toward the five different types of interruptions. Combined with their feedback on traditional voice agent interactions, we derived five design considerations (DCs) for developing voice agents that support interruptions and backchannels for older adults. Based on these DCs, we developed an LLM-powered Barge-in Agent that focuses on the natural conversation needs of older adults. It implements a series of age-friendly adjustments, such as prioritizing co-operative interruptions to help older adults complete their turns, using simplified structures during interruptions to prevent overwhelming them with complex prompts, and offering customizable settings for interruption timing and frequency to accommodate their slower speech patterns.

To address the RQ2, we employed a within-subject design ($N = 16$), comparing two conditions in a counterbalanced order: a **Barge-in Agent**, which is a duplex communication agent equipped with four interruption features (**Agent-Initiated Interruptions**, **Backchannel Responses**, **Interrupting User Silence by Initiating Conversation**, and **User-Initiated Interruptions**), with these details illustrated in Figure 1. And a **No Barge-in Agent** serving as the control condition, which is a simplex communication agent using a traditional turn-taking model. Participants interacted with both voice agents, allowing us to assess the impact of interruption mechanisms on user interaction. We found that the Barge-in Agent offers several key advantages for older adults, enhancing both engagement and conversational flow. More specifically, interruptions increase the frequency of interaction and allow broader topic exploration, while backchannels create a more realistic conversational experience and reduce the perception of waiting time. The ability of older adults to interrupt the voice agent gives them greater control, reducing frustration. Co-operative interruptions provide support by clarifying or supplementing information, helping maintain a smooth flow of dialogue. Additionally, the voice agent's ability to reduce response delays and proactively initiate conversation prevents awkward silences, ensuring continuous engagement.

In summary, this paper makes the following contributions:

- We proposed a series of DCs aimed at developing more natural and human-like LLM-powered agents for older adults that support interruptions and backchannels.
- We designed an exploratory Barge-in Agent based on LLMs, featuring four key functionalities: agent-initiated interruptions, backchannel responses, proactive initiation of conversation to break user silence, and support for user-initiated interruptions.
- We validated that older adults experience greater conversational engagement and fluency when interacting with voice agents that support interruptions and backchannels compared to voice agents that do not support these features.

This study not only explored the conversational patterns of voice agents for older adults, but also provided potential directions for future age-friendly voice interaction design.

2 BACKGROUND AND RELATED WORKS

2.1 Turn-taking, Interruptions and Backchannels

Most voice agents currently follow a strict alternating dialogue model, also known as turn-taking protocol, where one party speaks while the other waits [1, 6, 92]. This model relies on techniques like wait time, syntax analysis, and context understanding to detect the end of a turn, signaling when the system should respond [35, 55, 56]. However, human-to-human conversations are far more dynamic, often involving interruptions and backchannels that contribute to the fluidity and naturalness of the interaction, rather than strictly adhering to turn-taking protocols.

2.1.1 Interruptions. In conversation, an interruption is defined as an intentional act by one conversational participant to break into another's speech. This typically occurs at a non-transition-relevance place, where the speaker has not yet completed their turn or reached a natural transition point. Previous research has shown that interruptions play a critical role in human-computer interactions [8, 87]. Matsusaka et al. designed a multimodal agent capable of monitoring multi-user dialogues, interrupting when corrective information needs to be provided. However, they did not explore the user experience in this context [57]. Similarly, Palinko et al. demonstrated that using a prediction-based interruption model, combined with non-verbal signals, improved the user acceptance of robotic participation in multi-person conversations [72]. Despite these advances, the design of interruptions for single-user and robot conversations remains a critical challenge.

Interruptions themselves have been classified into several types based on their function and impact on the conversation. Early studies, such as those by Beattie et al., categorized interruptions by turn transition intensity, distinguishing between types like silent interruptions and butting-in interruptions [7, 22]. Later, Roger et al. differentiated between interruptive speech (e.g., intrusive interruptions that disrupt the conversation) and non-interruptive speech (e.g., supportive overlaps aimed at assisting the speaker) [79]. Murata et al. further refined this classification, categorizing interruptions based on their supportive or competitive nature into co-operative interruptions and competitive/intrusive interruptions.

The latter category is further subdivided into three types: topic-changing interruptions, floor-taking interruptions, and disagreement interruptions [62]. However, human-human conversational behavior changes with age [28, 86], and there is a lack of research on older adults' preferences for these different types of interruptions.

2.1.2 Backchannels. Backchannels are brief responses (e.g., "*uh-huh*," "*yeah*") that the listener uses to show that they are listening and understanding, and to encourage the speaker to continue talking without interrupting the flow of the conversation. More recently, the role of backchannels in voice agents has been studied in greater detail. Lala et al. developed an attentive listening system based on a humanoid robot, integrating continuous backchannels with responsive dialogue to maintain the flow of conversation [45]. Ding et al. categorized backchannels into reactive (e.g., "*hmm*") and proactive (e.g., "*please keep going*") and created a voice agent named TalkTive, demonstrating that participants preferred proactive backchannels over reactive ones [16]. In another study, Cho et al. implemented pseudo-random backchannels for Amazon Alexa to promote active listening, showing that this led to participants using more positive language [12]. However, despite the extensive exploration of backchannels in these studies, they have not addressed how interruptions could be effectively integrated into voice agent interactions. This remains a critical gap in the design of conversational agents.

2.2 LLM-Enhanced Voice Agents for Older Adults

Voice-user interfaces have become essential to improve technology accessibility for older adults [32, 89, 95]. As aging populations face physical and cognitive challenges, such as decreased vision, limited mobility, or reduced fine motor skills, traditional interfaces, such as touchscreens and keyboards, can pose significant usability barriers. In this context, voice agents utilize artificial intelligence to offer an intuitive and accessible conversational way for older adults to interact with technology, providing assistance and companionship [42, 66, 77, 99].

The emergence of Large Language Models (LLMs) has transformed the capabilities of voice agents, enabling more natural and context-aware conversations beyond traditional task-oriented interactions. Prior to LLMs, voice agents primarily relied on rule-based dialogue management systems and pre-defined conversation flows, limiting their ability to engage in open-ended dialogue [33, 106]. Recent studies have demonstrated LLMs' capability to generate more coherent and contextually relevant responses in voice-based interactions. Kim et al. conducted comprehensive studies on LLM-powered human-robot interaction, revealing that LLMs significantly enhance conversational capabilities through their superior context understanding and response generation [40].

Wong et al. explored the importance of anthropomorphism, empathy, and autonomy in the conversational style of a mental health support agent through co-design sessions with older adults and family caregivers [102]. These developments show particular promise in healthcare settings, where Yang et al. developed Talk2Care, an LLM-based voice assistant specifically designed to facilitate communication between healthcare providers and older adults [103]. Additionally, innovative applications like VoicePilot demonstrate how LLMs can enable more natural speech interfaces for assistive technologies [71].

However, most current research focuses on task-oriented interactions, such as information search, weather queries, alarm reminders, and music playback [58, 67, 70, 88, 101]. Current voice-based agents are still bound by limitations, particularly in replicating the fluidity and complexity of natural human conversation [15]. Many voice assistants operate under strict turn-taking rules, where one party speaks while the other waits. This approach, while useful for functional tasks, can feel rigid and unnatural in social conversations. Older adults may find such interactions frustrating, especially when the goal is companionship or emotional support rather than task completion [14].

Constructing highly functional and task-oriented sentences is difficult for older adults [13, 41, 74], who expect more social conversations with agents, mimicking the structure of human conversations and rules [13, 52]. While LLM-based voice conversational agents have made strides in improving interactions with older adults, challenges remain in making these interactions more natural, accessible, and engaging. In particular, for agents that need to support interruptions and interjections, LLM-based agents may be effective at determining whether interruptions are possible based on the user's input and returning appropriate interrupt content. This capability allows for more dynamic and fluid conversations, better mimicking natural human interactions and enhancing the overall user experience for older adults seeking companionship or emotional support.

3 FORMATIVE STUDY

To answer the RQ1, we conducted a formative study where older adults ($N = 8$) engaged in human-to-human conversations, retrospective interviews, and interactions with a turn-taking LLM agent. The goal of the study was to identify key DCs for developing a voice agent that effectively supports interruptions and backchannels for older adults. We focused on three key questions: What types of interruptions do older adults commonly use in in-person conversations? How do older adults perceive interruptions and backchannels in human-to-human conversations? And what are their expectations for agents that can support such behaviors? We gathered participants' feedback through retrospective think-aloud sessions. All studies received ethical approval from our institution. Participants could withdraw at any time, but no participants opted to withdraw.

3.1 Participant

Existing studies indicate that in dyadic conversations, acquaintances, as opposed to strangers, benefit from shared social knowledge, resulting in smoother conversations that do not require constant guidance [39]. In addition, conversations between acquaintances are more likely to include natural interruptions, offering valuable opportunities to observe authentic turn-taking patterns and preferences [80]. Based on these considerations, we recruited four pairs of participants (8 individuals in total) through personal networks and snowball sampling. Three pairs were married couples, and one pair consisted of friends. All individuals self-reported no hearing-related or other accessibility needs. Their ages ranged from

Figure 2: Procedure of the formative study, outlining the three phases: (a) in-person conversation, (b) playback and retrospective think-aloud, and (c) traditional turn-taking agent interaction, with rest breaks included to ensure participant comfort.

Table 1: Demographics of Formative Study Participants (N=8)

Pair ID	ID	Age	Sex	Voice Agent Experience	Relationship
Pair 1	1	60	Female	Tmall Genie Speaker	Friend
	2	65	Female	NA	
Pair 2	3	75	Female	Xiaodu Smart Speaker	Couple
	4	76	Male	Xiaodu Smart Speaker	
Pair 3	5	62	Male	Xiaomi Xiaoai Mobile Voice Assistant	Couple
	6	62	Female	Xiaodu Smart Speaker	
Pair 4	7	85	Male	OPPO Jovi Mobile Voice Assistant	Couple
	8	77	Female	XiaoDu Speaker	

60 to 85 years ($M = 70.25, SD = 8.57$), and seven had previous experience using voice agents. Detailed demographic information can be found in Table 1. Each pair received 100 RMB as compensation.

3.2 Preparation

Participants were asked to engage in conversations on two carefully selected topics, with the goal of helping them quickly enter a conversational state, which would facilitate the recording of interruptions by the researchers. The first topic involved discussing photos, given that research has indicated the significant role as a tool for stimulating the interest of older adults in conversational contexts. [36, 73, 86]. Before the study began, we asked each participant to select three personal photos they were most eager to share (6 photos per pair). Given the vulnerability of older adults, we carefully considered ethical aspects and conducted a preliminary screening of all submitted photos to ensure that no participant would feel uncomfortable. Existing research suggests that conversations among acquaintances often involve more emotional and experiential sharing [9]. Therefore, the second topic centered on recalling shared positive experiences.

3.3 Procedure

The formal study was conducted in a university laboratory and lasted approximately 120 minutes per pair of participants, including rest breaks to ensure they had sufficient time to relax at the end of each session. Before the formal trials began, participants completed a demographic questionnaire, were introduced to the study, and provided informed consent. The study was divided into three phases. The first phase involved an **In-person Conversation** (lasting 30-40 minutes). During this phase, the experimenter instructed the participants to discuss two topics in sequence. The first topic was to discuss photos. At the start of the photo discussion, one participant was chosen at random to begin, freely talking about his or her three photos in any preferred order, briefly recalling and describing the places, times, people, and events depicted. During this conversation, the other participants could freely ask questions or offer comments. Afterward, they switched roles, allowing the second participant to share his or her three photos. The second topic involved recalling shared positive memories. We provided a topic prompt board to guide the conversation, encouraging participants to discuss: *What is the story behind the photo (or shared memory)? When did the event in the photo (or shared memory) take place? Where did the event occur? What were your feelings at that time?* The second phase, **playback and retrospective think-aloud** (lasting 30-40 minutes), followed the in-person conversation. During this phase, participants were placed in separate rooms to review video playback of their own and their partner's interruptions and backchannels. They were asked to verbalize their reflections on both the behaviors they initiated and those they experienced, discussing the reasons and emotions behind these actions. In the third phase, **turn-taking LLM agent interaction and interview** (lasting 20-30 minutes), participants interacted for 15 minutes with the voice functionality of OpenAI's ChatGPT app[1] (v1.2024.080, ChatGPT-4 model) on a mobile phone. This functionality employed Whisper for speech recognition and a separate text-to-speech system to generate natural, human-like voice output. After the session, they provided feedback on this voice interaction model. We also conducted a semi-structured interview on their experience at the end of the study.

[1]https://apps.apple.com/us/app/chatgpt/id6448311069

3.4 Data Collection and Analysis

The data collected during the formative study consists of three parts: 1) video recordings of each participant pair's human-to-human conversations; 2) feedback from participants on their perceptions of being interrupted and interrupting others, collected through retrospective think-aloud sessions; 3) full recordings of all one-on-one semi-structured interviews. The video recordings were transcribed using a commercial automatic speech recognition (ASR) system, iFlyrec[2], with the transcriptions manually verified for accuracy by the research team. We adopted Okamoto et al.'s more contextual measurement to determine whether a speech act constitutes an interruption, considering situational factors such as the current speaker's intentions and the content of both speakers' utterances, rather than relying solely on syllables [68]. To track interruptions and backchannel behaviors during both the formative and user studies, we developed a Python plug-in that allowed for efficient timestamp marking and video retrieval for detailed review.

For the data analysis, we followed these steps: Four researchers independently coded the transcripts using the open coding method [100], supported by FigJam board [3]. During the coding process, we identified and defined multiple codes. Thinking Aloud records were used as a supplemental data source to validate and further elaborate on the open codes derived from the transcriptions. By analyzing participants' behaviors related to interrupting others and being interrupted, we gained a deeper understanding of their perspectives on interruptions and their expectations for voice agent interactions. This analysis helped to confirm and refine our initial coding. This research method has been shown to be effective in studies related to interaction and conversational dialogue with older adults [20, 21, 34].

Selected quotes from the phase, originally in Chinese, were translated by the first author and subsequently reviewed by co-authors to ensure accuracy.

3.5 Design Considerations

DC1- The voice agent should prioritize co-operative interruptions when interacting with older adults, while incorporating a diverse range of interruption types. Murata et al. classified interruptions into two main types: **Co-operative Interruption** and **Competitive/Intrusive Interruption**. Co-operative interruption involves assisting the speaker by supplying missing phrases or completing sentences, supporting the conversation without taking the speaker's turn. In contrast, competitive/intrusive interruption is more aggressive, aiming to change the topic, seize control, or express disagreement, potentially disrupting the speaker's flow. It is further divided into three subtypes: **Topic-changing Interruption** , **Floor-taking Interruption**, and **Disagreement Interruption** [63]. Through the retrospective analysis of co-operative interruptions among older adults, we further divided these into two subcategories: **Sentence Completion Interruption** and **Clarification/Inquiry Interruption**. Sentence completion interruption occurs when the interrupter helps by finishing the speaker's sentence due to pauses or hesitations. This type of interruption is intended to support the speaker when they have trouble continuing. Clarification/inquiry interruption occurs when the interrupter asks for clarification or further explanation to maintain the clarity of the conversation. Figure 3 illustrates examples of five different types of interruptions observed in dyadic conversations between participants, highlighting the cues and positions of each interruption. A quantitative summary of interruptions across the four pairs of older adults is summarized in Figure 4. The results indicate that older adults initiated co-operative interruptions more frequently than competitive ones and expressed more positive perceptions and evaluations of co-operative interruptions.

For co-operative interruptions, all participants ($N = 8$) expressed positive attitudes to sentence completion interruptions. Five participants ($N = 5$) specifically noted that age-related declines in memory and language organization made them appreciate when their conversation partner stepped in to fill in forgotten content or assist during moments of hesitation. P4 shared, "*I often forget what I was going to say while speaking. If the other person doesn't jump in, I feel awkward. Mutual support helps the conversation flow more smoothly.*" Most participants ($N = 7$) also viewed clarification/inquiry interruptions positively, recognizing them as a way to expand the topic and aid recall. They found this type of interruption constructive, as it invited greater depth and clarity, thereby enhancing mutual understanding. However, one participant ($N = 1$) expressed concern that asking questions before the speaker had finished could disrupt their train of thought and hinder the conversation's progress.

For competitive interruptions, the majority of participants ($N = 6$) felt that floor-taking interruptions could diminish the positivity of the conversation. However, they also acknowledged that such interruptions were sometimes necessary to foster better interaction between both parties. P3 remarked, "*I need someone to correct my mistakes, and others have the right to express their opinions. If one person is disinterested in the conversation but forced to listen, it's not a good experience.*" Floor-taking interruptions were often perceived as confrontational, potentially escalating the conversation into a debate and shifting the tone to a more contentious one. Most participants ($N = 6$) viewed disagreement interruptions as a regular part of the conversation but noted that they could make interactions more confrontational. While they recognized the importance of expressing opposing views, these interruptions were generally seen as detracting from the conversation's positivity. For topic-changing interruptions, the majority of participants ($N = 6$) believed that abruptly changing the topic mid-conversation could hinder the natural flow of ideas. However, some participants admitted that topic-changing interruptions were occasionally necessary to keep the conversation engaging and moving forward. A minority of participants ($N = 1$) preferred to avoid all competitive or intrusive interruptions, describing them as offensive and disruptive to the conversational flow.

DC2- The prompts and responses of the voice agent should be clear and concise, avoiding complex sentence structures that could confuse or overwhelm older adults. Older adults often use shorter, less complex sentences and may rely on vague expressions in conversation. This tendency toward semantic and syntactic simplification can limit their ability to express themselves in more intricate discussions [86]. To foster effective communication, voice agent prompts and responses must also be simplified.

[2]https://www.iflyrec.com/zhuanwenzi.html
[3]https://www.figma.com/figjam/

——	Specific Type	Example
Co-operative Interruption	Sentence completion Interruption	P1: It should be in Beijing...Beijing, uh, where in Beijing... P2: / Beijing's Temple?
	Clarification/Inquiry Interruption	P1: He's really into that souvenir. <laughs> He... P2: / What kind of souvenir?
Competitive/Intrusive Interruption	Floor-taking Interruption	P1: I saw this kind of flower in my hometown, I remember... P2: / Oh, I saw that in America. I ...
	Disagreement Interruption	P1: You're just too sensitive. I don't think.... P2: / I'm not sensitive, I just feel like this isn't right.
	Topic-changing Interruption	P1: Something that happened years ago, I think ... P2: / Don't talk about that, just tell ...

P1: Owner of the turn P2: Initiator of the interruption Words: Cue words /: Position of the interruption

Figure 3: Examples of interruption types from participants' real conversations in the formative study. The figure shows corresponding examples of the five different interruption types observed in actual dialogues between participants, with the cue words for the interruption and the position of the interruption marked.

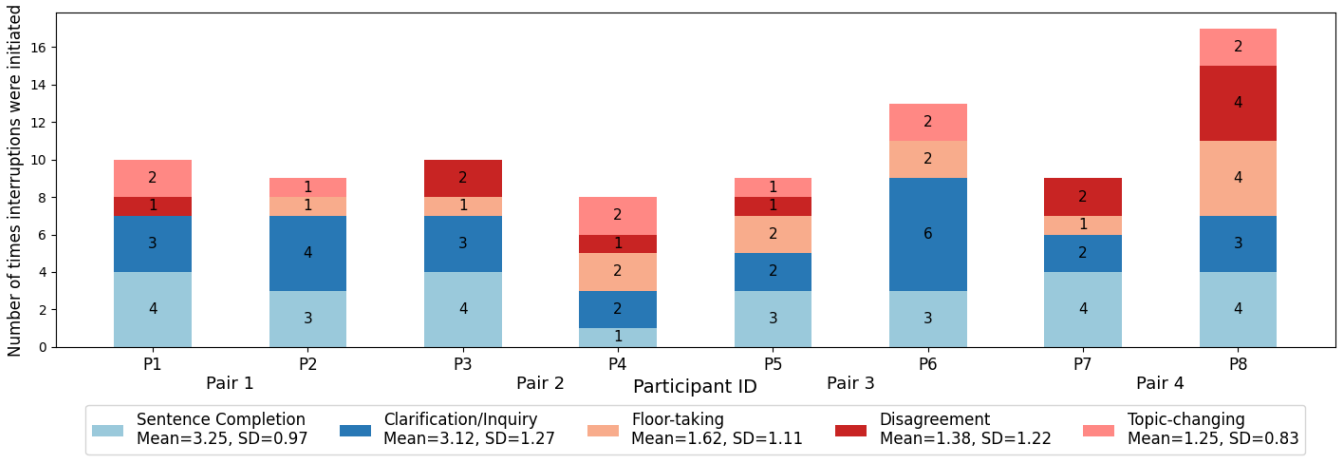

Figure 4: Stacked bar chart of interruption types initiated by 8 Participants in In the 30–40-minute dyadic conversations. The types of interruptions are: Sentence Completion (light blue), Clarification/Inquiry (dark blue), Floor-taking (orange), Disagreement (red), and Topic-changing (pink). The mean and standard deviation for each type of interruption are also provided.

Clear and direct language that aligns with the user's communication style is key, ensuring the information conveyed remains accurate. As P5 noted, "*When someone interrupts me, I find it helpful if they keep it short and don't over-explain. Interruptions can make me lose my train of thought, and the longer the person speaks after interrupting, the more frustrated I feel.*" It is critical to avoid nested clauses, technical jargon, and lengthy explanations, which can overwhelm or distract the user.

DC3- The voice agent should offer customizable response timing and interruption frequency for older adults. Speech rate and processing speed vary significantly among older adults, with some individuals speaking more slowly as they age due to extended cognitive processing times [23]. As a result, some older adults may need additional time to formulate their responses and frequently pause to gather their thoughts during conversations with a voice agent. During the third phase, which involved turn-taking interactions with an LLM-based voice agent, some participants ($N = 5$) experienced instances where their brief pauses during their speaking turn were misinterpreted by the voice agent as signals that the conversation had ended. This misunderstanding led the agent to take over the turn prematurely, causing overlaps and unintended interruptions. P1, P4, and P8 noted that such errors left them feeling frustrated and increased their anxiety about using the voice agent. P1 remarked, "*When the robot interrupts me, I feel like I've done something wrong, and it makes me feel that the robot is not as accommodating toward me.*" P4 mentioned, "*I guess I was*

speaking too slowly. After an error like that, I tend to subconsciously speed up my speech." In light of these individual differences, voice agents should adapt their response timing and the frequency of interruptions to align with the speech patterns of older adults.

DC4- The voice agent should distinguish and understand the interruption intention of older adults and respond accordingly or stop output. Due to age-related declines in hearing and communication abilities, many older adults find it increasingly challenging to comprehend complex conversations effectively [84]. As a result, they often need to ask questions or request repetitions to ensure they fully understand the information being conveyed. During the Part 3 session, some participants ($N = 3$) encountered instances where the agent's Speech-to-Text (STT) module misrecognized their speech, resulting in the agent providing incorrect responses. This confusion prompted participants to interrupt with the intent of correcting the agent. Additionally, a few participants ($N = 2$) experienced situations where they could not clearly hear the agent's output and attempted to interrupt to request it to stop or repeat the information. However, since the voice agent operates using a simplex communication protocol, it failed to acknowledge or respond to these participant behaviors. P7 remarked, "*The robot should respond to my questions or stop speaking when I say 'stop', rather than ignoring me as it does now.*" So voice agents should be able to discern the intent behind interruptions from older adults and respond accordingly, either by addressing the interruption or by halting their output when necessary.

DC5- The voice agent should proactively interrupt silences during interactions with older adults. During interactions with voice agents, older adults may experience long pauses due to uncertainty, hesitation, or unfamiliarity with the system. In the Part 3 session, some participants ($N = 2$) only responded with backchannels like "*uh-huh*" after the agent presented unengaging content, leading to a lack of audio input and causing the conversation to lapse into silence. P2 said, "*I had never used this software before, so I thought the robot would continue talking, and I didn't say anything.*" while P7 noted, "*I didn't have any thoughts on this topic, so I just stayed silent.*" These silences can negatively impact the user experience, leaving older adults feeling disengaged, isolated, or uncomfortable. Extended pauses without feedback can heighten their apprehension about using new technology. Therefore, voice agents should be capable of detecting silences and proactively re-engaging the conversation.

4 PROTOTYPE

Drawing from these DCs, we developed the Barge-in voice agent using the STT, LLMs, and Text-to-Speech (TTS) architecture. Specifically, the agent uses duplex communication and a streaming speech-to-text framework, which is the basis for implementing the Barge-in function [48].

4.1 Prototype modules

4.1.1 Main LLM Agent Module. This module is responsible for replying to the complete speech (turn) of older adults. In addition, it specifically considers improving the initiative of the conversation. When it detects that the voice agent has finished speaking, and the older adults fall silent either because they are not familiar with the system's speech detection or because they are not interested in the topic of conversation, the module will take the initiative to start the conversation again, avoiding awkward silences and ensuring the continuity of the conversation **(DC5)**.

4.1.2 Interrupting Multi-agent Module. The Interrupting multi-agent module comprises two LLM-based submodules: the **Interrupting Module** and the **Relevance Detection Module**. These submodules are designed to handle interruptions accurately during a conversation. The module's architecture and processing logic are illustrated in Figure 5. To understand how this module operates, it is essential to consider the mechanics of conversational turn-taking.

In conversation, speakers alternate between speaking and listening, and these exchanges are governed by **Turn Constructional Units (TCUs)**. A TCU represents the basic unit of speech, which can range from a single word to a full sentence, such as answering a question with "*I went to the store yesterday*" (a full-sentence TCU) or simply "*Bought some groceries.*" (a verb phrase TCU). At the end of each TCU, the conversation reaches a **Transition-Relevant Place (TRP)**, where the speaker signals their readiness to pass the floor. This transition is often marked by verbal or non-verbal cues. Additionally, **Inter-pausal Units (IPUs)**, which are stretches of uninterrupted speech without significant pauses (typically shorter than 200 ms), play a key role in determining when interruptions or turn shifts may occur. Notably, since older adults may exhibit slower speech patterns, their pauses between IPUs may be longer [26, 94]. These longer pauses, or brief hesitations such as "*I, um, think... this is great*," provide natural points for potential interruptions [48, 92]. To accommodate the speaking habits of older adults, the duration for pause detection in this module is adjustable **(DC3)**.

Building on this turn-taking framework, the Barge-in Agent assesses TRPs probabilistically to decide whether an interruption is appropriate. When an interruption is justified, the agent will intelligently determine the appropriate interruption based on the user's previous content by utilizing five types of interruptions specified in the prompt, generating suitable interruption responses **(DC1)**. It generates contextually relevant content and plays corresponding audio. Since LLM and STT systems generally take around 1.5 to 2 seconds to generate responses, the module employs a time-management strategy. Upon detecting a suitable interruption point, the system inserts a placeholder filler word, such as "*hmm...*" to indicate thoughtfulness or "*oh?*" to express mild opposition during the next IPU pause, providing enough time for the full response to be prepared [30]. It is important to note that placeholder filler words are used by speakers to fill pauses while organizing their thoughts, maintaining conversational flow, and avoiding silence. In contrast, backchannel feedback words, like "*uh-huh*" and "*oh*," are employed by listeners to demonstrate engagement, attention, and understanding, encouraging the speaker to continue without disrupting the dialogue. While placeholder filler words serve as self-regulatory tools for speakers, backchannel feedback words are listener-driven cues that signal attentiveness and foster mutual understanding [24, 43]. Meanwhile, the relevance detection module ensures that the interruption response remains relevant to the ongoing conversation. If the interruption response aligns with the latest spoken content, the system plays the corresponding audio.

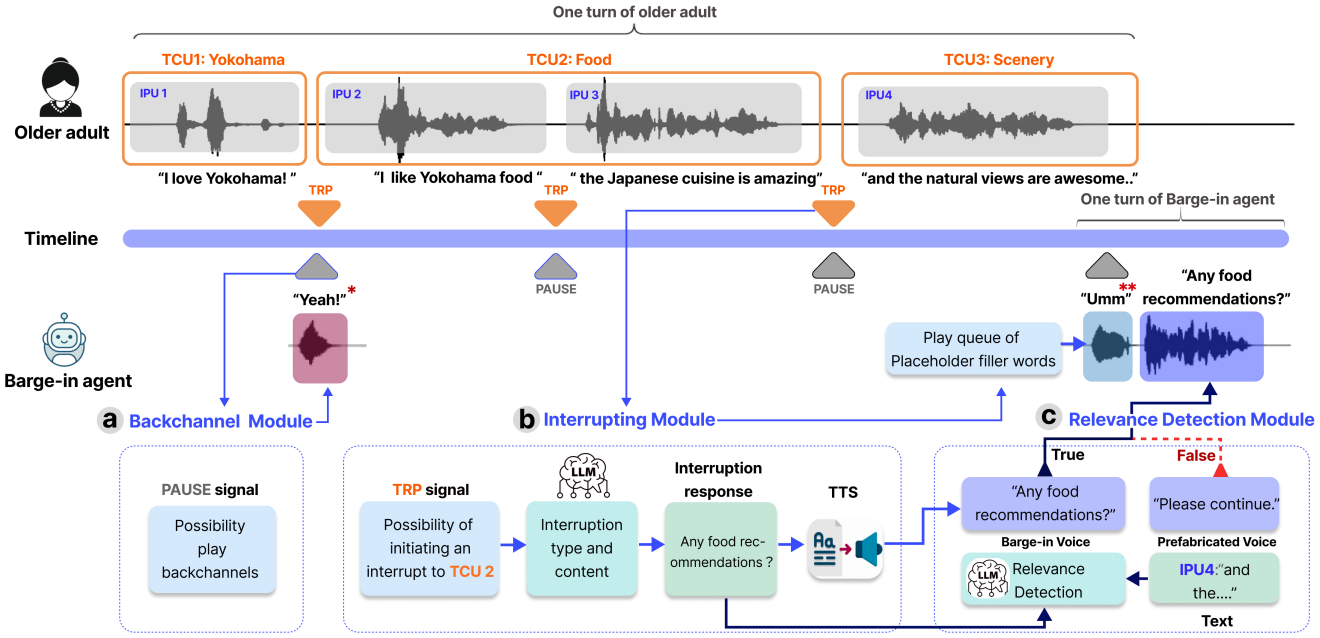

Figure 5: The framework of the Backchannel Module and Interrupting Multi-agent Module, illustrating the interaction flow between LLMs. The diagram highlights how the Barge-in Agent assesses turn-taking opportunities and triggers placeholder audio during processing delays, while the Relevance Detection Module ensures relevance in interruptions. The flowchart outlines how each module contributes to maintaining conversational fluidity by balancing interruptions with appropriate responses or filler content, enhancing both interactivity and user experience. Placeholder Filler Words **: Like "*umm*", primarily used by the speaker to sustain the continuity of dialogue while they think through what to say next. Backchannels *: Brief responses like "*uh-huh*" and "*Yeah*", indicating listener attention and understanding, show engagement without interrupting the speaker.

Otherwise, it defaults to a pre-recorded phrase like "*please continue*" to maintain the conversational flow. Each turn is evaluated for interruption only once. If no interruption occurs, the system proceeds with its normal response for that turn. To accommodate the conversational nature of older adults, the prompt in this module requires that interruptions be issued in a simple structure **(DC2)**.

4.1.3 User-Initiated Barge-in Module. There are two scenarios in this module. The first is when the older adult interrupts the agent. Suppose the agent is outputting audio during its turn and detects that the older adult has been speaking continuously for an extended period (adjusted to suit the individual, with a default of 1.5 seconds), or if specific interruption keywords are detected (e.g., competitive phrases like "*stop talking*" or "*don't say that*" or co-operative phrases like " *could you repeat*" or " *I didn't hear clearly*" (detailed keyword lists are provided in Appendix A.2, the system will recognize this as an interruption initiated by the user. The agent will stop speaking, record the interruption in the context, and wait for the user to finish before responding appropriately. The second scenario occurs when the agent fails to interrupt. If it is the user's turn to speak and the agent attempts to output interruption content, this is considered a failed interruption by the agent. In such cases, the system will allow the older adult to complete their statement before continuing. This module is designed to provide older adults with the ability to ask questions or requests **(DC4)**.

4.1.4 Backchannel Module. While the older adult participant is speaking, this module monitors Pause after each IPU in real time and probabilistically inserts backchannels during these pauses. It randomly plays pre-recorded responses such as " *mm-hmm*" and "*yeah*" (in Chinese) as feedback to the speaker's content, enhancing the social engagement of the conversation. It is important to note that the backchannel function has a lower priority than the Barge-in filler words, and the two are mutually exclusive.

4.2 Implementation

As shown in Figure 6, the system architecture integrates iFlytek's M260C microphone array and M2 audio processing board for voice capture and echo cancellation. These components are specifically designed to address the challenges of duplex communication by removing sounds emitted by the voice agent's own speaker that are inadvertently picked up by the microphone. For STT module, we

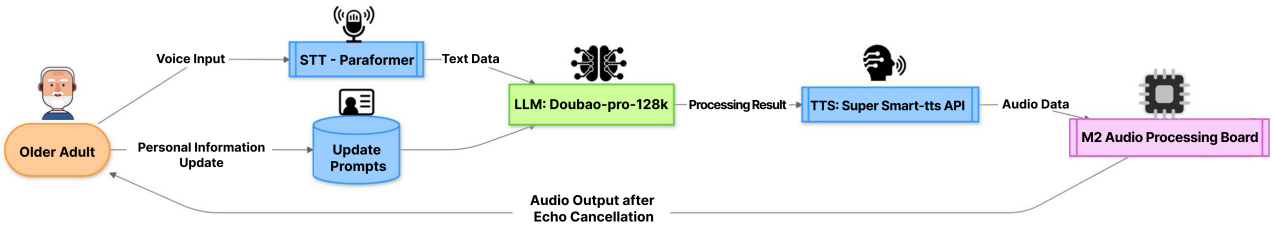

Figure 6: System architecture diagram showing the process of older adult voice input and processing, including speech recognition, language model processing, text-to-speech conversion, and echo cancellation.

integrated Alibaba Cloud's Paraformer API[4], which provides real-time, low-latency recognition for both Chinese and English audio streams. This API ensures accurate and efficient transcription, even in noisy environments. To support LLM, we utilized Volcengine's Doubao-pro-128k API[5]. This API is optimized for Chinese conversations, offering advanced role-playing capabilities and a context window of up to 128k characters, making it ideal for the complex requirements of this study. For TTS, we adopted iFlytek's Super Smart-TTS API[6]. This API converts text into natural human-like speech while simulating paralinguistic features such as breathing, sighing, and speech rate variations, enhancing the realism and engagement of the conversational experience. Detailed performance data can be found in Appendix A.3. We also provide adjustable parameter, including the agent's response wait time and proactive interruption frequency, to accommodate the different conversation paces and thinking rhythms of older adults (**DC3**). In addition, we developed a No Barge-in Agent for subsequent user studies. This agent follows a traditional strict turn-taking protocol, which does not support interruptions or backchannels. Aside from this difference, the two agents remain consistent in core components such as STT, LLMs, and TTS.

5 USER STUDY

To answer RQ2, which explores the conversational experience that a voice agent with four types of barge-in support can offer to older adults, we conducted a within-subject design. Specifically, the study focused on four barge-in mechanisms: the agent interrupting the user, the agent providing backchannel responses, the agent breaking silences by proactively initiating conversation, and allowing older adults to interrupt the agent. 16 participants engaged with both the Barge-in Agent, which integrates these features, and a traditional turn-taking agent without barge-in capabilities, allowing for a comparative analysis of user interaction and engagement. In addition, two social workers from a community elderly care center assisted us with participant recruitment at the center.

5.1 Participant

We recruited 16 participants, 13 of them through word-of-mouth and snowball sampling. Additionally, we contacted 2 social workers who work at a local community elderly care center, and they helped us recruit 3 participants from the center. Among the participants, one reported hearing difficulties and used a hearing aid during daily conversations and the study. Another participant from the elderly care center reported mobility challenges, so all three participants from the center completed the study in the center's meeting room.

The participants ranged in age from 60 to 90 years ($mean$ = 67.88, SD = 8.89), with 64.3% (N = 8) living with a spouse or children, 18.75% (N = 3) living alone, and 18.75% (N = 3) residing in a community elderly care center. Experience with voice agents varied, 68.75% (N = 11) participants had used some form of voice agent, while 31.25% (N = 5) had no prior experience. Responses to the open-ended questionnaire item, "Motivations for interacting with voice agents or desired topics of conversation," were analyzed, and researchers categorized the motivations of the 16 participants into five distinct themes. The largest group, 37.5% (N = 6), aimed to Alleviate loneliness, followed by 25.0% (N = 4) expressing an interest in Learning new knowledge. Other motivations included Sharing daily life (12.5%, N = 2), Expressing troubles and seeking opinions (12.5%, N = 2), and Entertainment (12.5%, N = 2). The original responses are detailed in the Appendix 6, and comprehensive demographic information is presented in Table 2. Each participant received 100 RMB as compensation.

5.2 Study Design

According to the Esposito et al. study, virtual female characters are often attributed with greater emotionality and nurturing qualities, which help to foster empathy and emotional connection with older users. As a result, older adults tend to prefer interacting with feminized voice agents [18, 19]. To help the participants better focus on the study and reduce the sense of waiting caused by agent response delays [29, 78], we designed a young female 3D avatar for the agent based on the Unity platform. This avatar is equipped with facial blend shapes, enabling lip-sync animations that correspond to the agent's speech. We conducted a within-subject design and counterbalanced the condition order: (i) **No Barge-in Agent** – a simplex communication agent using a traditional turn-taking model; (ii) **Barge-in Agent** – a duplex communication agent that supports interruptions and backchannels.

To prevent participants' conversations with the LLM from involving negative content, we implemented descriptive restrictions in the prompts and steered the conversations. Additionally, four researchers and one social worker conducted a week-long internal test of the agent and finetuning to ensure that no abnormal content was reported. Detailed prompt content can be found in the Appendix A.1.

[4]https://help.aliyun.com/zh/dashscope/developer-reference/quick-start-7
[5]https://www.volcengine.com/product/doubao
[6]https://www.xfyun.cn/services/smart-tts

Table 2: Demographics of User Study Participants (N=16). The "Motivations or Desired Topics" column represents a summarized coding of participants' questionnaire responses: "Motivations for interacting with voice agents or desired topics of conversation."

ID	Sex	Age	Living Arrangement	Experience of Voice Agents	Motivations or Desired Topics
1	Male	78	Living with spouse	Never used	Alleviating loneliness
2	Female	77	Living with spouse	Used once or twice	Alleviating loneliness
3	Female	65	Living with spouse	Use frequently, several times a week	Sharing daily life
4	Male	67	Living alone	Use frequently, several times a week	Alleviating loneliness
5	Female	61	Living with children	Use daily or almost daily	Expressing troubles and seeking opinions
6	Female	61	Living with spouse	Use frequently, several times a week	Expressing troubles and seeking opinions
7	Female	60	Living with spouse	Used once or twice	Entertainment
8	Female	62	Living with spouse	Use frequently, several times a week	Sharing daily life
9	Female	67	Living alone	Never used	Alleviating loneliness
10	Female	62	Living with spouse	Use occasionally, less than once a week	Learning new knowledge
11	Male	61	Living with spouse	Never used	Learning new knowledge
12	Female	61	Living alone	Use occasionally, less than once a week	Learning new knowledge
13	Female	63	Living with spouse	Never used	Alleviating loneliness
14	Male	90	Elderly care center	Never used	Alleviating loneliness
15	Female	71	Elderly care center	Used once or twice	Learning new knowledge
16	Female	80	Elderly care center	Use occasionally, less than once a week	Entertainment

The user study was conducted in two locations: P1-P13 in the university laboratory and P14-P16 in the meeting room of the community elderly care center, with social workers accompanying. As shown in Figure 7, we ensured consistency in the experimental environment across both locations. The study design, including the use of audio recordings, photography, and generative AI, adhered to ethical guidelines and by the institution's ethical review. Participants could withdraw at any time, but no participants opted to withdraw from this study.

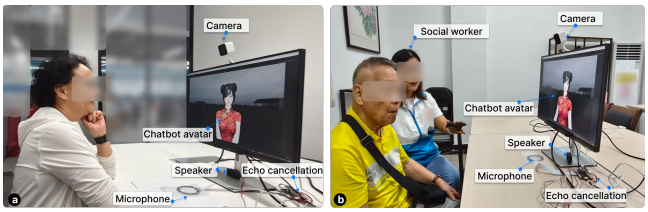

Figure 7: User study environments: (a) In the university laboratory; (b) In the community elderly care center. The figure shows the consistent setup across both locations, ensuring similar environmental conditions for all participants. In both settings, participants interacted with the agent under controlled conditions, with audio recordings and social workers present in the community care center to provide support. This setup was designed to minimize external influences and maintain a comfortable, familiar environment for elderly participants.

5.3 Procedure

Before the study, we collected two separate descriptions of each participant's daily activities and hobbies through a questionnaire, instructing them to provide different content in each description to reduce practice effects in our within-subject experiment. Participants were informed and gave their consent to provide this information to the agent. These descriptions were then randomly assigned to either the Barge-in Agent or the No Barge-in Agent. Upon arrival at the experimental site, participants filled out a demographic information questionnaire, were introduced to the experimental setup, and signed a consent form.

The entire study lasted approximately 120 minutes, with sufficient rest breaks to ensure participants' comfort. To minimize order effects, 50% of the participants first experienced the No Barge-in Agent, followed by the Barge-in Agent, while the other 50% followed the reverse order. In each experimental condition, participants first received a brief training session to learn how to interact with the agents and completed a 5-minute functionality trial. During this trial, participants engaged in a brief conversation with the Barge-in Agent, while a researcher adjusted the agent's response wait time and proactive interruption frequency (parameters for **DC3**) until they were satisfied. This process aimed to avoid frequent interruptions and excessively short wait times from disrupting the participant's thinking and responses, thereby meeting the individual needs of each participant.

Afterwards, they engaged in a formal conversation about their daily life and hobbies (15-20 minutes). At the end of each condition, participants were asked to fill out the User Engagement Scale Short Form (UES-SF), which measures engagement in terms of aesthetic appeal, focus, novelty, usability, felt involvement, and endurance. The short-form version was used in this study as it has been shown to reduce participant burden and is well-suited for use in controlled experiments [69] (for the scale details, see the Appendix B.1). After completing the questionnaire, participants underwent a semi-structured interview to explore their experiences further (10-20 minutes). Once both agent conditions were completed, participants were asked to compare their experiences with the two agents in a final interview session.

5.4 Data Collection and Analysis

We used the same data collection and transcription methods as in the formative study. The data analysis involved open coding of the transcripts by four researchers [100], with support from a FigJam

board [7]. This process led to the identification of multiple codes, with think-aloud records used to validate these codes. By analyzing participants' behaviors related to interruptions, we gained deeper insights into their views and expectations for voice agent interactions. After the interviews, we held weekly meetings to cross-check these insights, which were then organized into a codebook. This codebook served as a reference for code definitions, usage contexts, and example quotes, ensuring consistency and reliability in our analysis. To ensure the accuracy of the qualitative data, we translated selected quotes from Chinese into English, which were then reviewed by co-authors. This translation process helped maintain the integrity of our data and findings, providing a strong foundation for our design and prototype development. Through this systematic approach, we gained a clear understanding of the interactions between older adults and LLM-powered voice agents, which is crucial for our ongoing research.

6 RESULTS

We first present participants' general user experience comparing the Barge-in Agent with the No Barge-in Agent, highlighting the key advantages in conversational engagement and conversational fluency. More specifically, interruptions increase the frequency of interaction and allow for broader topic exploration, while backchannels create a more realistic conversational experience and reduce the perception of waiting time. The ability for older adults to interrupt the agent gives them greater control, decreasing frustration. Co-operative interruptions provide support by clarifying or supplementing information, helping maintain a smooth flow of dialogue. Additionally, the agent's ability to reduce response delays and proactively initiate conversation prevents awkward silences, ensuring continuous engagement.

6.1 Overall Experience

After processing the scale data (Figure 8, 9), which measured **Focused Attention** (the extent to which users are absorbed in the interaction and lose track of time), **Perceived Usability** (any negative affect experienced during the interaction, such as frustration or confusion, and assesses the effort needed to use the system), **Aesthetic Appeal** (the attractiveness and visual appeal of the interface), and **Reward** (which combines aspects of novelty, felt involvement, and endurability, measuring whether users found the interaction valuable and enjoyable and if they'd recommend it to others), we conducted two-tailed paired t-tests ($\alpha = 0.05$) to compare the four dimensions of the UES. The results showed significant differences in **Focused Attention** ($t(15) = -3.36, p = 0.0043$), **Aesthetic Appeal** ($t(15) = -3.67, p = 0.0023$), and **Reward** ($t(15) = -2.37, p = 0.0315$). However, the result for **Perceived Usability** ($t(15) = -1.83, p = 0.088 > 0.05$) was not significant.

In terms of Focused Attention, participants ($N = 13$) noted that compared to the No Barge-in Agent, the Barge-in Agent enriched the interaction by incorporating interruptions and backchannels, thereby enhancing engagement. As P4 mentioned, "*The frequency of interaction with the Barge-in Agent was noticeably higher. It responded while I was speaking, making the conversation feel more interactive and engaging. On the other hand, the voice agent without interruptions just seemed to follow commands, lacking interaction, and I didn't feel like continuing the conversation.*"

For Perceived Usability, participants ($N = 10$) felt that the Barge-in Agent conducted more human-like conversation techniques, such as proactively changing topics through interruptions. "*It would initiate conversation and guide the flow of conversation*" (P7, P11). In contrast, the No Barge-in Agent would fall into silence when users didn't respond, leading to confusion or frustration. Participants (P10, P11) noted, "*The Question and Answer style conversation with the No Barge-in Agent often gets stuck on a single topic, making it hard to continue, while the Barge-in Agent guides more topics.*" Some participants ($N = 5$) also commented that both the Barge-in Agent and the No Barge-in Agent can carry on a smooth conversation, and there is not much difference between them. Since the two agents use the same STT, LLM and TTS components, and the main response prompts are basically the same, there is not much difference in usability.

Regarding Aesthetic Appeal, participants ($N = 9$) found that the Barge-in Agent's interruptions felt more like a familiar friend's conversation style, which made it more engaging and human-like. The turn-based structure of the No Barge-in Agent created a mechanical feel. As P16 said, "*The No Barge-in Agent just follows along with what I say—it's too monotonous. The Barge-in Agent shares its own opinions, creating a conversational atmosphere like talking with an old friend.*"

In terms of Reward, the participants ($N = 11$) found the Barge-in Agent more novel for them. One participant (P3) stated, "*The Barge-in Agent can supplement what I can't think of and interrupt me. This is novel and very different from the voice assistants I have used before.*"

On the other hand, few participants ($N = 3$) expressed a preference for the No Barge-in Agent for three main reasons: "*It feels impolite to interrupt*" (P14), "*Interruptions disrupt the flow of my thoughts*" (P1), and the agent should be more of a listener, primarily listening rather than interrupting.

6.2 Enhancing Conversational Engagement

We conducted a paired sample t-test ($\alpha = 0.05$) to analysis four key metrics of turn-taking behavior in conversations between 16 participants and two different voice agents: questions from user, questions from agent, user turns and topic counts. These metrics were chosen to capture critical dimensions of conversational engagement, including participant proactivity, agent autonomy, interaction fluidity, and content diversity. The results, summarised in Figure 10, reveal significant differences across all four metrics. Specifically, the number of questions from user ($t(15) = -2.42, p = 0.029$) showed a statistically significant variation. Questions from agent ($t(15) = -7.20, p < 0.001$), the number of user turns ($t(15) = -6.77, p < 0.001$), and the number of topics discussed ($t(15) = -4.84, p < 0.001$) exhibited highly statistically significant differences. As detailed in Table 3, participants engaged in more dynamic and interactive conversations with the Barge-in Agent compared to the No Barge-in Agent. The Barge-in Agent elicited more user initiated questions and generated significantly more questions itself, demonstrating a stronger ability to actively guide conversations. Additionally, participants

[7] https://www.figma.com/figjam/

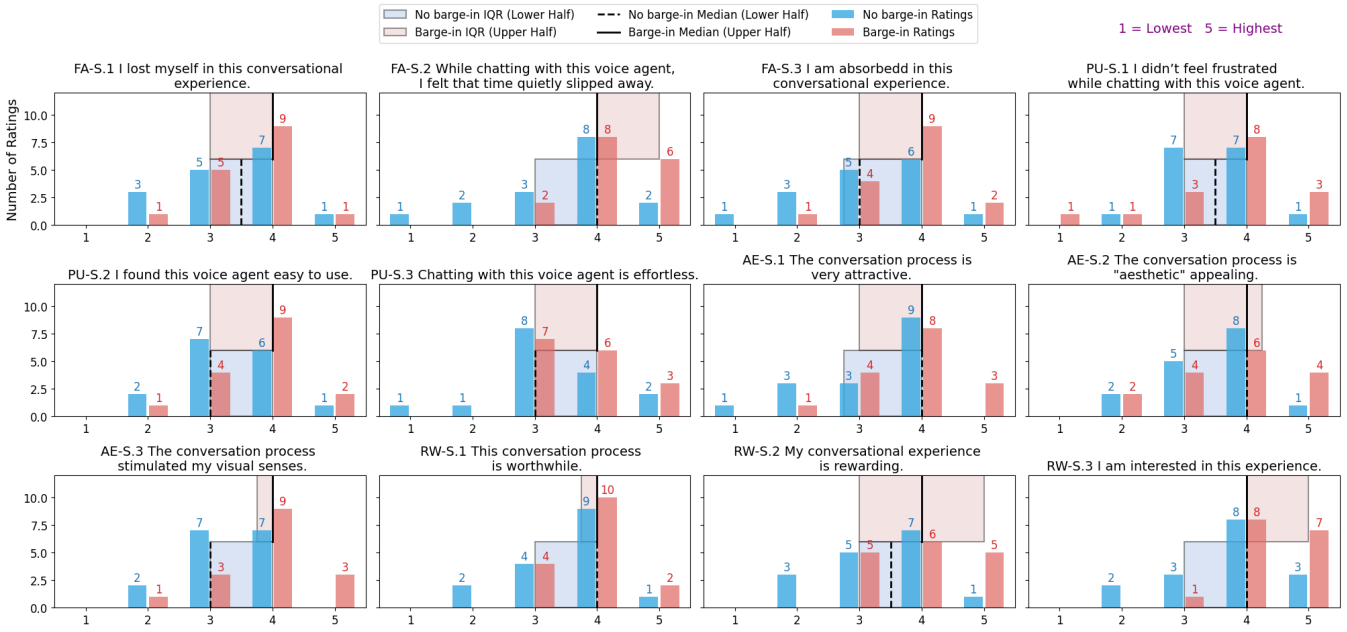

Figure 8: Participants' questionnaire ratings on a scale of 1-5 using the User Engagement Scale Short Form (UES-SF). Red indicates ratings for voice interactions with the Barge-in Agent, while blue indicates ratings for voice interactions with the No Barge-in Agent. For the performance scale, higher ratings represent better performance. (The version of the questions used in the user study was translated into Chinese)

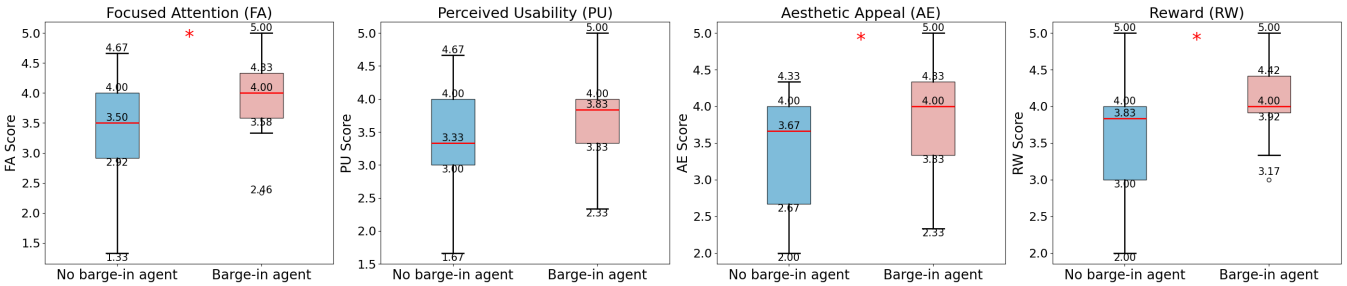

Figure 9: Comparison of user experience dimensions (FA: Focused Attention, PU: Perceived Usability, AE: Aesthetic Appeal, RW: Reward) between the No Barge-in Agent and the Barge-in Agent. Each boxplot shows the median (horizontal line inside the box), the interquartile range (IQR, box height), and the range (whiskers extending to 1.5 times the IQR). Outliers are represented as individual points beyond the whiskers. Red asterisks (*) above the boxplots denote statistically significant differences ($p < 0.05$) based on two-tailed paired t-tests.

exhibited a higher number of speech turns, suggesting that its capacity for interruptions and backchannels fostered more frequent interactions. Furthermore, conversations supported by the Barge-in Agent covered a broader range of topics, indicating its effectiveness in maintaining engaging and diverse discussions.

6.2.1 Interruption can enhance conversational engagement for older adults by increasing frequency and expanding topic breadth. Participants ($N = 7$) mentioned that during interactions with the Barge-in Agent, natural and frequent turn shifts were the most critical factor in their perception of engagement, as they contributed to a more human-like conversational flow, "*the pace was faster, and the conversation felt more engaging*" (P5, P2). As shown in Figure 11, participants were interrupted an average of 5 times ($SD = 1.73$) during 15-minute conversations with the Barge-in Agent. These interruptions facilitated turn transitions, promoting an average increase of 10.37 user turns compared to interactions with the No Barge-in Agent. After coding the conversation topics, it was found that participants discussed an average of 1.38 more topics with the Barge-in Agent. Participants ($N = 9$) noted that the Barge-in Agent often interrupted to change topics, allowing the conversation to explore a wider range of subjects. P6 commented: "*The Barge-in Agent interrupted to introduce new topics, leading to more varied*

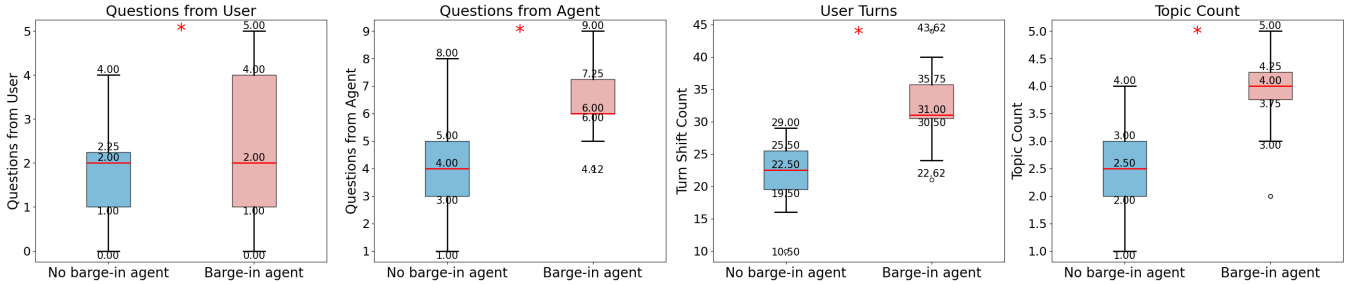

Figure 10: Comparison of statistical metrics (Questions from User, Questions from Agent, User Turns, Topic Count) between the No Barge-in Agent and the Barge-in Agent. Each boxplot displays the median (the red horizontal line inside the box), the interquartile range (IQR, represented by the height of the box), and the range (whiskers extending to 1.5 times the IQR). Red asterisks (*) above the boxplots indicate statistically significant differences ($p < 0.05$) based on two-tailed paired t-tests.

Table 3: Number of conversations participants had with Barge-in Agent and in No Barge-in Agent (Mean ± SD) within 15 minutes

Condition	Questions from user	Questions from agent	User turns	Topic counts
Barge-in Agent	2.44 ± 1.63	6.50 ± 1.41	32.31 ± 5.75	3.88 ± 0.96
No Barge-in Agent	1.75 ± 1.13	4.19 ± 1.8	21.94 ± 5.14	2.50 ± 0.89

discussions. In contrast, the No Barge-in Agent repeatedly discussed the same topic." P10 added, "*No Barge-in Agent kept repeating topics, half the time it revolved around cooking, and after a while, I lost interest.*" However, some participants ($N = 3$) felt that the No Barge-in Agent allowed for deeper exploration of a single topic. P7 further noted, "*There are advantages and disadvantages to both deeper topic discussions and expanded topics. Deeper discussions make the conversation more meaningful, whereas expanded topics keep the conversation dynamic and prevent stagnation.*"

6.2.2 Backchannels enhance the perceived realism of conversations and subjectively reduce waiting time for older adults. The participants ($N = 14$) found backchannels necessary for engaging in voice interactions with an agent, as they contribute to a more natural conversational flow. Some participants ($N = 8$) did not even notice the presence of backchannels, indicating their seamless integration into the dialogue. Participants ($N = 7$) mentioned that when they spoke, the agent's use of backchannels like "*mm-hmm*" or "*yeah*" made them feel acknowledged and respected, reinforcing the perception that the agent was actively listening. P7 remarked, "*I need the agent to respond when I'm talking, to show it's genuinely listening.*" Additionally, N=6 participants noted that, compared to No Barge-in Agent interactions where no backchannels were present, they were more concerned about the system's response time and feared it had frozen. As P16 stated, "*Without backchannels, I don't know what the system is doing, especially when there's a delay.*" A small number of participants ($N = 2$) felt that the backchannels occasionally interrupted their train of thought and suggested that the pitch, volume, and content of backchannels be adjusted to match user preferences.

6.2.3 The ability for older adults to interrupt the agent empowers them to have greater control over the conversation while also reducing feelings of frustration. Throughout the study, several participants ($N = 6$) attempted to interrupt the agent for three primary reasons. First, they sought to steer the conversation toward topics they were more interested in. As P3 noted, "*I wasn't interested in that topic, so I interrupted the agent to talk about something else.*" Similarly, P15 mentioned that the No Barge-in Agent "*talked for too long, and I lost interest, so I tried to interrupt it.*" Second, participants interrupted the agent to correct certain content. The ability to interrupt in real time allowed them to quickly address misunderstandings or inaccuracies. As P7 explained, being able to interrupt the agent to correct its mistakes made the conversation feel more relaxed. Lastly, when participants needed the agent to repeat or clarify certain information, real time interruptions enabled them to resolve confusion quickly. As P13 mentioned, "*The Barge-in Agent was explaining the cooking steps too fast, and I couldn't catch it, so I interrupted and asked it to repeat.*" Most participants ($N = 10$) did not attempt to interrupt the agents, which may be influenced by practical habits . P2 mentioned, "*I still feel unfamiliar with the agent, so I instinctively try to be polite, interrupting feels rude*".

6.3 Enhancing Conversational Fluency

6.3.1 Co-operative interruptions help older adults by supplementing information and asking relevant questions, aiding thought completion and conversational flow. As shown in Figure 11, participants experienced an average of 3.31 Co-operative Interruptions ($SD = 1.26$) during interactions with the Barge-in Agent, including 1.56 Sentence Completions Interruptions ($SD = 1$) and 2 Clarification/Inquiries Interruptions ($SD = 0.65$). The Barge-in Agent, guided by preset prompts, assisted participants in completing sentences when they forgot a name, place, time, or event, but did not abandon their turn. Participants ($N = 13$) noted that Sentence Completions Interruptions helped prevent prolonged pauses and made conversations smoother. P4 said: "*I couldn't recall the name of Liberation Monument and got stuck. The agent filled in the place name, so the*

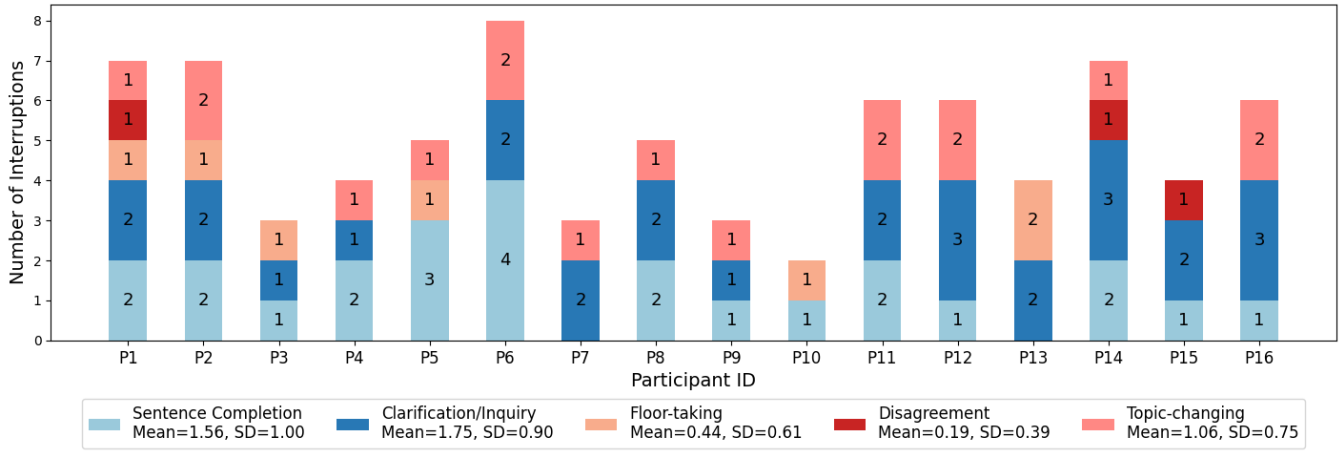

Figure 11: Stacked bar chart of the number of interruptions experienced by participants. The chart shows the frequency of different types of interruptions experienced by each participant. The types of interruptions are: Sentence Completion (light blue), Clarification/Inquiry (dark blue), Floor-taking (orange), Disagreement (red), and Topic-changing (pink). The mean and standard deviation for each type of interruption are also provided.

conversation could continue." Participants ($N = 5$) found that Clarification/Inquiries Interruptions from the Barge-in Agent helped them recall related experiences and details, which was especially beneficial for older adults in maintaining cognitive engagement and triggering more memories. P16 shared, "*The best part was that it prompted my memories. I mentioned my experience in Wuhan, and it asked me about Wuhan landmarks, which allowed me to continue the conversation.*"

6.3.2 Interruptions reduce long response delays in LLM-powered agents, objectively lowering perceived waiting time for older adults. A common challenge with LLM-powered agents is that continuous user speech generates more input tokens, requiring additional processing time and resulting in longer response delays. When interacting with the No Barge-in Agent, participants ($N = 4$) experienced extended input during turns. Notably, three of these four participants had no prior experience using voice agents. As illustrated in Figure 12, P11 experienced multiple turns exceeding 30 seconds while conversing with the No Barge-in Agent. This occurred because, after speaking for an extended period, the user would pause, expecting the agent to respond. However, the agent required additional time to process the lengthy input. During this extended waiting period, the user assumed the agent was either still waiting for input or had encountered an error, prompting them to continue speaking. This extended their turn further, generating more input and subsequently increasing the response delay. Conversely, the Barge-in Agent interrupted the user's turn at appropriate moments, actively initiating the conversation and balancing participation between both parties. For instance, statistical analysis of P11's behavior revealed that the average pause and silence duration per turn was 3.94 seconds ($SD = 2.80$ seconds) during interactions with the Barge-in Agent, compared to 7.92 seconds ($SD = 12.37$ seconds) with the No Barge-in Agent. Both P11 and P13, after experiencing these extended waiting periods with the No Barge-in Agent, asked the research team if the system had malfunctioned and whether they should continue talking. The Barge-in Agent's ability to interrupt at appropriate moments helped manage response lengths and reduce prolonged waiting times. As P13 noted: "*Since there were no interruptions, the No Barge-in Agent's responses felt slower, and the waiting times were longer.*"

6.3.3 By proactively interrupting conversational silences, the conversation can continue more seamlessly. In No Barge-in Agent interactions, participants ($N = 5$) were unfamiliar with the system's timing or felt that the conversation could not continue, resulting in silence after the agent finished speaking. This caused the agent to enter a waiting state. As P10 described, "*When I didn't know what to say, the agent also didn't react, and I didn't know what to do next.*" In contrast, the Barge-in Agent could detect when users fell silent and proactively initiated the next conversation. During interactions with the Barge-in Agent, some participants ($N = 9$) experienced this proactive engagement, and P9 and P10 noted that "*the agent's proactive approach prevented conversation stalls.*" This reduced the social pressure on older adults to keep the conversation going, with P7 and P10 commenting, "*The biggest fear in conversations is awkward silence; it's hard to keep the conversation going.*" Additionally, P8 mentioned that "*when the agent misunderstood me and it was difficult to correct, I didn't know what to say, but then the agent initiated a new topic, which increased the system's usability.*"

7 DISCUSSION

In this section, we explore insights derived from evaluating the LLM-powered Barge-in Agent for older adults, providing guidance for the future design of voice agents that simulate natural conversations for this demographic. Our discussion focuses on three key aspects: proactively guiding topics through contextual cues and timely interruptions, tailoring interruption and backchannel strategies to accommodate older adults' sensory perceptions, and optimizing conversational coherence based on cognitive feedback and adapting interruption strategies. By addressing these factors,

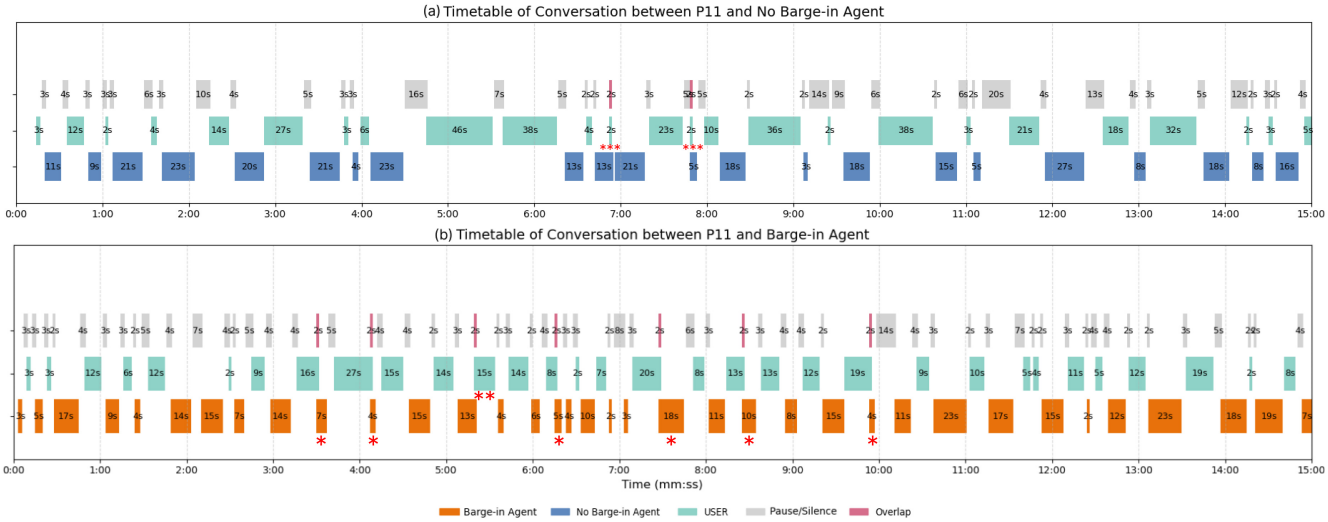

Figure 12: Comparison of P11's timetables during conversations with the No Barge-in Agent and the Barge-in Agent. (a) Timetable of P11 conversation with the No Barge-in Agent: Green bars represent P11 speaking, blue bars represent the No Barge-in Agent speaking, gray bars indicate pauses or silence (neither speaking), and pink bars represent both speaking simultaneously. (***) denotes instances where P11 attempted to interject during the No Barge-in Agent's turn. (b) Timetable of P11 conversation with the Barge-in Agent: Green bars represent P11 speaking, orange bars represent the Barge-in Agent speaking, gray bars indicate pauses or silence, and pink bars represent both speaking simultaneously. (*) indicates instances where the Barge-in Agent interrupted P11, and (**) denotes instances where P11 interrupted the Barge-in Agent.

we highlight how future voice agents can enhance engagement and fluency compared to traditional voice dialogue systems.

7.1 Guiding Topics Proactively Through Contextual Cues and Timely Interruptions

7.1.1 Leveraging personal context to enhance topic relevance. In conversations with older adults, topics are particularly important, especially in maintaining engagement and enhancing the sense of participation [61]. Before the study, we asked participants about their daily lives and hobbies to ensure that the conversation would revolve around topics they were familiar with and interested in. Previous studies have also shown that starting a conversation with familiar topics helps improve its continuity and flow. Most participants indicated that this made them feel more comfortable and gave them a sense that the agent understood them, which in turn enhanced their sense of participation [47]. Participants generally did not mind that this personal information was being collected by the agent, as they believed that sharing such information was natural and necessary, even in everyday communication.

7.1.2 Timely topic-shifting interruptions guided by contextual relevance. By introducing timely interruptions, whether agent-initiated or participant-initiated, the conversation can be effectively guided in new directions, making it more diverse and enriching. In our study, the agent used topic-changing interruptions to introduce new topics, successfully broadening the scope of the discussion. As P6 said, "*The agent's interruptions introduce new topics, making the conversation more diverse,*" which participants found more engaging compared to a No Barge-in Agent. At the same time, when the topic deviates from the participant's interests, participants may also interrupt to steer the conversation toward topics they prefer. Participants generally want the conversation to revolve around topics they care about and strive to keep the conversation interesting [25].

7.2 Tailoring Interruption and Backchannel Strategies to Older Adults' Sensory Perceptions

7.2.1 Mitigating auditory perception variability with multimodal feedback. In our study, we observed differences in participants' perceptions of interruptions and backchannel signals. Previous studies have also indicated that with advancing age, older adults face not only the issue of hearing loss but also changes in the central auditory system, including differences in processing times, alterations in neural response patterns, and a slowing down in the response speed to rapid sequences of sound events [98]. Some participants reported that the backchannel (e.g., "*yes*," "*uh-huh*") and active interruptions made by the Barge-in Agent interfered with their speech, this is likely because the auditory system struggles to quickly process and distinguish between these rapid auditory cues, making it difficult for them to seamlessly integrate these signals into ongoing conversations without feeling disrupted. In addition, excessive backchannel responses can lower users' evaluations of the robot, which is consistent with observations in the field of voice chat [38]. However, under the same conditions, other participants

did not perceive the agent's backchannel responses or interruptions. They regarded these signals as part of a natural conversation, not intrusive, and actually found them more immersive.

Previous research has also shown that in specific collaborative tasks with robots, facial expressions as backchannel responses can enhance user engagement [85]. Given the variability in older adults' perception of auditory cues, future voice agents could address these challenges by incorporating visual modalities, such as facial expressions, alongside other multimodal signals, into their interruption and backchannel mechanisms. This approach would not only accommodate age-related auditory processing differences but also create more intuitive, natural, and engaging interactions for older adults.

7.2.2 Use pre-defined filler words and backchannel to decrease participants' perception of waiting time. The agent's active interruptions and backchannel responses also alleviated participants' perception of waiting time caused by system delays, making the conversation experience more coherent and reducing negative emotions from prolonged waiting. System latency is the primary challenge that prevents agents from achieving human-like interruptive behavior [3, 91]. Interruptions are highly time-sensitive interactions, typically occurring within a few hundred milliseconds. However, the process for agents, which includes STT, text generation, and TTS, takes significantly longer than human response times. To address this, the Barge-in Agent employs a strategy of using predefined filler words to mask delays in generating interruption audio and utilizes multiple agents to enhance interruption accuracy. Additionally, the backchannel feedback from the agent was also interpreted by participants as the agent responding to them, with the understanding that the agent also needed time to "think." These measures effectively reduced the long response delays caused by the LLM-driven agent, objectively decreasing participants' perception of waiting time. Participants generally expressed that they did not mind being interrupted and, in fact, welcomed the agent's timely interventions.

7.3 Optimizing Conversational Coherence Based on User Cognitive Feedback and Adapting Interruption Strategies

In our formative study, we found that older adults are interrupted easily in their thinking during conversations. In our user study, we further observed a common phenomenon: some participants, when interrupted by the agent, chose to abandon the previous topic and stop discussing it. Some participants mentioned that they disliked the Barge-in Agent's backchannel (such as "*yes*" or "*mm-hmm*") and competitive/intrusive interruption because these disrupted their thought processes. They believed that during a conversation, they needed some time to organize their thoughts and formulate their responses. This aligns with prior research, which shows that older adults often require more time to think and express themselves [53, 93]. In contrast, co-operative interruptions, such as sentence completion interruptions, help keep the continuity of the participant's thoughts. These interruptions can help participants in recalling temporarily forgotten information, further promoting a smoother flow of conversation.

7.3.1 Adapting interruption timing and frequency based on cognitive load indicators. The frequency of interruptions and the timing of when the agent intervenes while waiting for a response are crucial. To avoid disrupting the participants' thinking, we manually adjusted the interruption timing and frequency to match each participant's preferences during the trial phase. However, in the actual experiment, we found that as the conversation progressed and topics changed, participants' thinking time also varied. The agent should have an adaptive waiting mechanism that dynamically adjusts interruption timing and waiting durations based on cognitive load indicators (e.g., speech hesitation, filler words, and response latency) and conversational cues (e.g., topic complexity, sentiment shifts, and turn-taking patterns) [5, 10]. For instance, if a participant frequently pauses or uses hesitation markers like "*um*" and "*let me think*," the agent could infer a higher cognitive load and extend the waiting time before intervening. Conversely, if the participant gives short or abrupt responses, the agent might infer disengagement and initiate a new topic to sustain the conversation flow. By incorporating these real-time behavioral signals, the agent can be a more patient listener. In addition, it has also been suggested that identifying user emotions to adjust interruption frequency and manner through emotion modeling algorithms to suit individual personality traits can effectively enhance the agent's responsiveness to the older adult's needs, thereby optimizing the overall conversational experience [105].

7.3.2 Preserving continuity via latency-aware technical and proactive topic initiation. Keeping the continuity of the conversation is equally important. As mentioned, due to system latency, the Barge-in Agent's response time often exceeds human perception, which can cause frustration due to prolonged waiting times. Therefore, during short periods of silence between both parties, the agent should provide appropriate backchannel feedback to simulate an ongoing response, helping to maintain conversational flow. A more effective approach would be to reduce the technical components involved, such as minimizing communication between APIs, and utilizing multimodal large models that directly use voice input and text output, such as Qwen-Audio [2], thereby reducing delays and enhancing the naturalness of the conversation.

For longer delays or silences, the agent should proactively introduce a new topic to keep the conversation flowing smoothly. In our study, we observed that some participants, when conversing with traditional agents, were either unfamiliar with the system's timing or felt that the conversation could not continue, resulting in silence after the agent finished speaking. This is consistent with previous literature: systems that use simple VAD to detect the end of a user's speech will not continue speaking if the user does not respond. As a result, when agents converse with older adults, they are more likely to fall into this state [60, 96]. Our agent is designed to proactively initiate a conversation and introduce new topics to break the silence when older adults do not respond, based on the contextual content of speech-to-text. However, relying solely on text modality cannot accurately capture the user's emotional state. Therefore, in practical applications, integrating speech, facial expressions, and other emotional signals will help more accurately determine the reasons for silence and provide a more natural and smooth conversational experience.

7.4 Limitations and Future Work

While our study provides insights into improving interactions between older adults and voice agents, several limitations suggest areas for further investigation.

The voice agent in this study focused on emotional companionship, but its applicability to more task-oriented agents is unclear. To collect data on interruptions and turn-taking, we selected couples and same-age friends as participants. However, interactions with younger family members or different relationship types may exhibit different patterns, affecting the generalizability of our findings. Future research should examine how older adults interact with various demographic groups (e.g., children, grandchildren) to inform agent design [46, 59, 90]. Although we allowed customizable parameters, such as response wait times and interruption frequencies, uncertainty in network latency and token calculation speed during the study prevented us from fully exploring their effects. Future studies should investigate these parameters and offer adaptive settings to better meet individual needs.

The research focused on dialogues between older adults and the agent, excluding factors like tone, speech rate, and personality traits, which could significantly affect interaction dynamics. Future research should include these elements to better understand their influence. For example, exploring how tone and speech rate affect backchannel responses and interruptions could lead to more personalized agents. Additionally, investigating personality traits could help tailor interactions. Including multimodal cues such as facial expressions and gestures would enhance the agent's naturalness and responsiveness, especially for older adults who rely on nonverbal communication. Expanding the study to other demographics and conducting longitudinal research could provide insights into the long-term effects of such agents on emotional well-being and cognitive function.

8 CONCLUSION

Our research highlights the importance of incorporating interruption and backchannel capabilities into voice agents to enhance interactions with older adults. Our LLM-powered Barge-in Agent provides a more natural and engaging conversational experience. The formative study revealed the frequent use of interruptions and backchannels in older adults' natural conversations, shaping the design of our agent. The within-subject exploratory study showed that most participants found conversations with the Barge-in Agent more engaging and fluid compared to a standard turn-taking agent. These findings suggest that integrating human-like conversational features is essential for agents aimed at providing emotional companionship and supporting memory recall in older adults. We hope that our design insights will inspire the development of more natural, human-like agents, ultimately improving the quality of life for older adult users.

ACKNOWLEDGMENTS

This work is partially supported by the Guangzhou-HKUST(GZ) Joint Funding Project (No. 2024A03J0617), Education Bureau of Guangzhou Municipality Funding Project (No. 2024312152), Guangzhou Higher Education Teaching Quality and Teaching Reform Project (No. 2024YBJG070), Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No. 2023B1212010007), the Project of DEGP (No.2023KCXTD042), and the Guangzhou Science and Technology Program City-University Joint Funding Project (No. 2023A03J0001).

REFERENCES

[1] Gregory Aist. 1998. Expanding a time-sensitive conversational architecture for turn-taking to handle content-driven interruption. *5th International Conference on Spoken Language Processing (ICSLP 1998)* (1998). https://doi.org/10.21437/ICSLP.1998-69

[2] Aliyun. 2024. Qwen Audio Chat API Developer Reference. https://help.aliyun.com/zh/dashscope/developer-reference/qwen-audio-chat-api?spm=a2c4g.11186623.0.i1. Accessed: 2024-12-10.

[3] James Allen, George Ferguson, and Amanda Stent. 2001. An architecture for more realistic conversational systems. In *Proceedings of the 6th international conference on Intelligent user interfaces*. 1–8.

[4] Anneliese Arnold, Stephanie Kolody, A. Comeau, and Antonio Miguel Cruz. 2022. What does the literature say about the use of personal voice assistants in older adults? A scoping review. *Disability and rehabilitation. Assistive technology* (2022), 1–12. https://doi.org/10.1080/17483107.2022.2065369

[5] Agnes Axelsson, Hendrik Buschmeier, and Gabriel Skantze. 2022. Modeling feedback in interaction with conversational agents—a review. *Frontiers in Computer Science* 4 (2022), 744574.

[6] Matthew Peter Aylett and Marta Romeo. 2023. You Don't Need to Speak, You Need to Listen: Robot Interaction and Human-Like Turn-Taking. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 11, 5 pages. https://doi.org/10.1145/3571884.3603750

[7] G. Beattie. 1981. Interruption in conversational interaction, and its relation to the sex and status of the interactants*. 19 (1981), 15 – 36. https://doi.org/10.1515/ling.1981.19.1-2.15

[8] Adna Bliek, Suna Bensch, and T. Hellström. 2020. How Can a Robot Trigger Human Backchanneling? *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (2020), 96–103. https://doi.org/10.1109/RO-MAN47096.2020.9223559

[9] G. Bodie, Susanne M. Jones, Miriam Brinberg, Amy M. Joyer, D. Solomon, and Nilam Ram. 2021. Discovering the Fabric of Supportive Conversations: A Typology of Speaking Turns and Their Contingencies. *Journal of Language and Social Psychology* 40 (2021), 214 – 237. https://doi.org/10.1177/0261927X20953604

[10] Samantha WT Chan, Shardul Sapkota, Rebecca Mathews, Haimo Zhang, and Suranga Nanayakkara. 2020. Prompto: Investigating receptivity to prompts based on cognitive load from memory training conversational agent. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4, 4 (2020), 1–23.

[11] Xiaohui Chen, Katherine Luo, Trevor Gee, and Mahla Nejati. 2024. Does ChatGPT and Whisper Make Humanoid Robots More Relatable? *arXiv preprint arXiv:2402.07095* (2024).

[12] Eugene Cho, Nasim Motalebi, S. Shyam Sundar, and Saeed Abdullah. 2022. Alexa as an Active Listener: How Backchanneling Can Elicit Self-Disclosure and Promote User Experience. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 273 (Nov. 2022), 23 pages. https://doi.org/10.1145/3555164

[13] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300705

[14] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. " What can i help you with?" infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th international conference on human-computer interaction with mobile devices and services*. 1–12.

[15] Smit Desai and Michael Twidale. 2023. Metaphors in voice user interfaces: a slippery fish. *ACM Transactions on Computer-Human Interaction* 30, 6 (2023), 1–37.

[16] Zijian Ding, Jiawen Kang, Tinky Oi Ting HO, Ka Ho Wong, Helene H Fung, Helen Meng, and Xiaojuan Ma. 2022. TalkTive: A Conversational Agent Using Backchannels to Engage Older Adults in Neurocognitive Disorders Screening. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 304, 19 pages. https://doi.org/10.1145/3491102.3502005

[17] Olov Engwall, Ronald Cumbal, and A. Majlesi. 2023. Socio-cultural perception of robot backchannels. *Frontiers in Robotics and AI* 10 (2023). https://doi.org/10.3389/frobt.2023.988042

[18] A. Esposito, T. Amorese, M. Cuciniello, M. Riviello, A. Esposito, A. Troncone, M. Inés Torres, Stephan Schlögl, and G. Cordasco. 2019. Elder user's attitude toward assistive virtual agents: the role of voice and gender. *Journal of Ambient Intelligence and Humanized Computing* 12 (2019), 4429 – 4436. https://doi.org/10.1007/s12652-019-01423-x

[19] Anna Esposito, Stephan Schlögl, Terry Amorese, Antonietta Esposito, Maria Inés Torres, Francesco Masucci, and Gennaro Cordasco. 2018. Seniors' Sensing of Agents' Personality from Facial Expressions. In *Computers Helping People with Special Needs*, Klaus Miesenberger and Georgios Kouroupetroglou (Eds.). Springer International Publishing, Cham, 438–442.

[20] Mingming Fan, Vinita Tibdewal, Qiwen Zhao, Lizhou Cao, Chao Peng, Runxuan Shu, and Yujia Shan. 2022. Older Adults' Concurrent and Retrospective Think-Aloud Verbalizations for Identifying User Experience Problems of VR Games. *Interacting with Computers* 34, 4 (Feb. 2022), 99–115. https://doi.org/10.1093/iwc/iwac039

[21] Mingming Fan, Qiwen Zhao, and Vinita Tibdewal. 2021. Older Adults' Think-Aloud Verbalizations and Speech Features for Identifying User Experience Problems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 358, 13 pages. https://doi.org/10.1145/3411764.3445680

[22] Nicola Ferguson. 1975. Interruptions: speaker-switch nonfluency in spontaneous conversation. *Annexe Thesis Digitisation Project 2016 Block 4* (1975).

[23] C. Fougeron, Fanny Guitard-Ivent, and V. Delvaux. 2021. Multi-Dimensional Variation in Adult Speech as a Function of Age. *Languages* (2021). https://doi.org/10.3390/languages6040176

[24] Scott H. Fraundorf and Duane G. Watson. 2011. The disfluent discourse: Effects of filled pauses on recall. *Journal of memory and language* 65 2 (2011), 161–175. https://doi.org/10.1016/J.JML.2011.03.004

[25] Zhengdong Gan, C. Davison, and L. Hamp-Lyons. 2009. Topic Negotiation in Peer Group Oral Assessment Situations: A Conversation Analytic Approach. *Applied Linguistics* 30 (2009), 315–334. https://doi.org/10.1093/APPLIN/AMN035

[26] A. Gerstenberg, Susanne Fuchs, Julie Marie Kairet, J. Schröder, and C. Frankenberg. 2018. A cross-linguistic, longitudinal case study of pauses and interpausal units in spontaneous speech corpora of older speakers of German and French. *Speech Prosody 2018* (2018). https://doi.org/10.21437/SPEECHPROSODY.2018-43

[27] Linda Grenade and Duncan Boldy. 2008. Social isolation and loneliness among older people: issues and future challenges in community and residential settings. *Australian health review* 32, 3 (2008), 468–478.

[28] W. Horton, Daniel H. Spieler, and Elizabeth Shriberg. 2010. A corpus analysis of patterns of age-related change in conversational speech. *Psychology and aging* 25 3 (2010), 708–13. https://doi.org/10.1037/a0019424

[29] Ludovic Hoyet, Clément Spies, Pierre Plantard, A. Sorel, R. Kulpa, and F. Multon. 2019. Influence of Motion Speed on the Perception of Latency in Avatar Control. *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2019), 286–2863. https://doi.org/10.1109/AIVR46125.2019.00066

[30] Derek Jacoby, Tianyi Zhang, Aanchan Mohan, and Yvonne Coady. 2024. Human Latency Conversational Turns for Spoken Avatar Systems. https://doi.org/10.48550/ARXIV.2404.16053

[31] Baptiste Jacquet and Jean Baratgin. 2020. Mind-Reading Chatbots: We Are Not There Yet. (2020), 266–271. https://doi.org/10.1007/978-3-030-55307-4_40

[32] Dietmar Jakob. 2022. Voice controlled devices and older adults–a systematic literature review. In *International Conference on Human-Computer Interaction*. Springer, 175–200.

[33] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)* 54, 5 (2021), 1–36.

[34] Xiang Ji and Pei-Luen Patrick Rau. 2018. A comparison of three think-aloud protocols used to evaluate a voice intelligent agent that expresses emotions. *Behaviour & Information Technology* 38, 4 (Oct. 2018), 375–383. https://doi.org/10.1080/0144929x.2018.1535621

[35] Bing'er Jiang, Erik Ekstedt, and G. Skantze. 2023. Response-conditioned Turn-taking Prediction. (2023), 12241–12248. https://doi.org/10.48550/arXiv.2305.02036

[36] Leying Jiang, Panote Siriaraya, Dongeun Choi, and N. Kuwahara. 2021. A Library of Old Photos Supporting Conversation of Two Generations Serving Reminiscence Therapy. *Frontiers in Psychology* 12 (2021). https://doi.org/10.3389/fpsyg.2021.704236

[37] Valerie K Jones, Michael Hanus, Changmin Yan, Marcia Y Shade, Julie Blaskewicz Boron, and Rafael Maschieri Bicudo. 2021. Reducing loneliness among aging adults: The roles of personal voice assistants and anthropomorphic interactions. *Frontiers in public health* 9 (2021), 750736.

[38] Malte F. Jung, Jin Joo Lee, N. DePalma, Sigurdur O. Adalgeirsson, Pamela J. Hinds, and C. Breazeal. 2013. Engaging robots: easing complex human-robot teamwork using backchanneling. *Proceedings of the 2013 conference on Computer supported cooperative work* (2013). https://doi.org/10.1145/2441776.2441954

[39] G. G. Kent, John D. Davis, and D. Shapiro. 1981. Effect of mutual acquaintance on the construction of conversation. *Journal of Experimental Social Psychology* 17 (1981), 197–209. https://doi.org/10.1016/0022-1031(81)90014-7

[40] Callie Y Kim, Christine P Lee, and Bilge Mutlu. 2024. Understanding large-language model (llm)-powered human-robot interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 371–380.

[41] Sunyoung Kim et al. 2021. Exploring how older adults use a smart speaker–based voice assistant in their first interactions: Qualitative study. *JMIR mHealth and uHealth* 9, 1 (2021), e20427.

[42] Sunyoung Kim and Abhishek Choudhury. 2021. Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study. *Computers in Human Behavior* 124 (2021), 106914. https://doi.org/10.1016/j.chb.2021.106914

[43] Birgit Knudsen, Ava Creemers, and A. Meyer. 2020. Forgotten Little Words: How Backchannels and Particles May Facilitate Speech Planning in Conversation? *Frontiers in Psychology* 11 (2020). https://doi.org/10.3389/fpsyg.2020.593671

[44] Sota Kobuki, Katie Seaborn, S. Tokunaga, Kosuke Fukumori, Shun Hidaka, K. Tamura, K. Inoue, Tatsuya Kawahara, and M. Otake-Matsuura. 2023. Robotic Backchanneling in Online Conversation Facilitation: A Cross-Generational Study. *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (2023), 71–76. https://doi.org/10.1109/RO-MAN57019.2023.10309362

[45] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. 2017. Attentive listening system with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, Kristiina Jokinen, Manfred Stede, David DeVault, and Annie Louis (Eds.). Association for Computational Linguistics, Saarbrücken, Germany, 127–136. https://doi.org/10.18653/v1/W17-5516

[46] Patrick J. Leman, Shahina Ahmed, and Louise Ozarow. 2005. Gender, gender relations, and the social dynamics of children's conversations. *Developmental Psychology* 41, 1 (2005), 64–74. https://doi.org/10.1037/0012-1649.41.1.64

[47] E. C. Li, S. E. Williams, and A. Della Volpe. 1995. The effects of topic and listener familiarity on discourse variables in procedural and narrative discourse tasks. *Journal of Communication Disorders* 28, 1 (1995), 39–55. https://doi.org/10.1016/0021-9924(95)91023-Z

[48] Ting-En Lin, Yuchuan Wu, Fei Huang, Luo Si, Jian Sun, and Yongbin Li. 2022. Duplex Conversation: Towards Human-like Interaction in Spoken Dialogue Systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Washington DC, USA) *(KDD '22)*. Association for Computing Machinery, New York, NY, USA, 3299–3308. https://doi.org/10.1145/3534678.3539209

[49] Weijane Lin, Hong-Chun Chen, and Hsiu-Ping Yueh. 2021. Using Different Error Handling Strategies to Facilitate Older Users' Interaction With Chatbots in Learning Information and Communication Technologies. *Frontiers in Psychology* 12 (Dec. 2021). https://doi.org/10.3389/fpsyg.2021.785815

[50] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.

[51] Gale M Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI* 4 (2017), 51.

[52] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[53] S. Lutz and Karin Knop. 2020. Put down your smartphone – unless you integrate it into the conversation! An experimental investigation of using smartphones during face to face communication. 9 (2020), 516–539. https://doi.org/10.5771/2192-4007-2020-4-516

[54] John Markoff and Paul Mozur. 2015. For sympathetic ear, more chinese turn to smartphone program. *NY Times* (2015).

[55] Ryo Masumura, Mana Ihori, Tomohiro Tanaka, Atsushi Ando, Ryo Ishii, T. Oba, and Ryuichiro Higashinaka. 2019. Improving Speech-Based End-of-Turn Detection Via Cross-Modal Representation Learning with Punctuated Text Data. *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (2019), 1062–1069. https://doi.org/10.1109/ASRU46091.2019.9003816

[56] Ryo Masumura, Tomohiro Tanaka, Atsushi Ando, Ryo Ishii, Ryuichiro Higashinaka, and Yushi Aono. 2018. Neural Dialogue Context Online End-of-Turn Detection. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics. https://doi.org/10.18653/v1/w18-5024

[57] Yosuke Matsusaka, Shinya Fujie, and Tetsunori Kobayashi. 2001. Modeling of conversational strategy for the robot participating in the group conversation. In *Seventh European conference on speech communication and technology*.

[58] Rachel McCloud, Carly Perez, Mesfin Awoke Bekalu, and K. Viswanath. 2022. Using Smart Speaker Technology for Health and Well-being in an Older Adult Population: Pre-Post Feasibility Study. *JMIR Aging* 5, 2 (9 May 2022), e33498. https://doi.org/10.2196/33498

[59] Natalie Merrill, E. Gallo, and R. Fivush. 2015. Gender Differences in Family Dinnertime Conversations. *Discourse Processes* 52 (2015), 533–558. https://doi.org/10.1080/0163853X.2014.958425

[60] Hidekazu Minami, Hiromichi Kawanami, M. Kanbara, and N. Hagita. 2016. Chat robot coupling machine responses and social media comments for continuous conversation. *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (2016), 1–6. https://doi.org/10.1109/ICMEW.2016.7574758

[61] G. Mois, E. Lydon, Shraddha A. Shende, Madina Khamzina, D. Myers, R. Mudar, and Wendy A. Rogers. 2022. LEVERAGING VIDEO CONFERENCING TECHNOLOGY TO FACILITATE SOCIAL ENGAGEMENT IN OLDER ADULTS. *Innovation in Aging* 6 (2022), 583–583. https://doi.org/10.1093/geroni/igac059.2188

[62] Kumiko Murata. 1994. Intrusive or co-operative? A cross-cultural study of interruption. *Journal of Pragmatics* 21, 4 (1994), 385–400. https://doi.org/10.1016/0378-2166(94)90011-6

[63] Kumiko Murata. 1994. Intrusive or co-operative? A cross-cultural study of interruption. *Journal of Pragmatics* 21, 4 (April 1994), 385–400. https://doi.org/10.1016/0378-2166(94)90011-6

[64] United Nations. 2020. World population ageing 2020 highlights: Living arrangements of older persons.

[65] Nicholas R Nicholson. 2012. A review of social isolation: an important but underassessed condition in older adults. *The journal of primary prevention* 33 (2012), 137–152.

[66] Katherine O'Brien, Anna Liggett, Vanessa Ramirez-Zohfeld, Priya Sunkara, and Lee A Lindquist. 2020. Voice-controlled intelligent personal assistants to support aging in place. *Journal of the American Geriatrics Society* 68, 1 (2020), 176–179. https://doi.org/10.1111/jgs.16217

[67] Bruna Oewel, Tawfiq Ammari, and Robin N. Brewer. 2023. Voice Assistant Use in Long-Term Care. In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23)*. ACM, 1–10. https://doi.org/10.1145/3571884.3597135

[68] Dina G. Okamoto, Lisa Slattery Rashotte, and Lynn Smith-Lovin. 2002. Measuring Interruption: Syntactic and Contextual Methods of Coding Conversation. *Social Psychology Quarterly* 65, 1 (2002), 38–55. http://www.jstor.org/stable/3090167

[69] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (April 2018), 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

[70] Katherine O'Brien, Anna Liggett, Vanessa Ramirez-Zohfeld, Priya Sunkara, and Lee A. Lindquist. 2019. Voice-Controlled Intelligent Personal Assistants to Support Aging in Place. *Journal of the American Geriatrics Society* 68, 1 (Oct. 2019), 176–179. https://doi.org/10.1111/jgs.16217

[71] Akhil Padmanabha, Jessie Yuan, Janavi Gupta, Zulekha Karachiwalla, Carmel Majidi, Henny Admoni, and Zackory Erickson. 2024. Voicepilot: Harnessing LLMs as speech interfaces for physically assistive robots. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[72] Oskar Palinko, Kohei Ogawa, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2018. How Should a Robot Interrupt a Conversation Between Multiple Humans. In *Social Robotics*, Shuzhi Sam Ge, John-John Cabibihan, Miguel A. Salichs, Elizabeth Broadbent, Hongsheng He, Alan R. Wagner, and Álvaro Castro-González (Eds.). Springer International Publishing, Cham, 149–159.

[73] R. Pera, S. Quinton, and Gabriele Baima. 2020. I am who I am : Sharing photos on social media by older consumers and its influence on subjective well-being. *Psychology & Marketing* (2020). https://doi.org/10.1002/mar.21337

[74] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174214

[75] François Portet, Michel Vacher, Caroline Golanski, Camille Roux, and Brigitte Meillon. 2013. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Personal and Ubiquitous Computing* 17 (2013), 127–144.

[76] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. " Phantom Friend" or" Just a Box with Information" Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.

[77] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. "Phantom Friend" or "Just a Box with Information": Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 214 (nov 2019), 21 pages. https://doi.org/10.1145/3359316

[78] R. Riedl, Peter N. C. Mohr, P. Kenning, Fred D. Davis, and H. Heekeren. 2014. Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory. *Journal of Management Information Systems* 30 (2014), 114 – 83. https://doi.org/10.2753/MIS0742-1222300404

[79] D. Roger, Peter Bull, and Sally Smith. 1988. The Development of a Comprehensive System for Classifying Interruptions. *Journal of Language and Social Psychology* 7 (1988), 27 – 34. https://doi.org/10.1177/0261927X8800700102

[80] D. Roger and Willfried Nesshoever. 1987. Individual differences in dyadic conversational strategies: a further study. *British Journal of Social Psychology* 26 (1987), 247–255. https://doi.org/10.1111/J.2044-8309.1987.TB00786.X

[81] Paul K Rubenstein, Chulayuth Asawaroengchai, Duc Dung Nguyen, Ankur Bapna, Zalán Borsos, Félix de Chaumont Quitry, Peter Chen, Dalia El Badawy, Wei Han, Eugene Kharitonov, et al. 2023. Audiopalm: A large language model that can speak and listen. *arXiv preprint arXiv:2306.12925* (2023).

[82] Hyeyoung Ryu, Soyeon Kim, Dain Kim, Sooan Han, Keeheon Lee, and Younah Kang. 2020. Simple and Steady Interactions Win the Healthy Mentality: Designing a Chatbot Service for the Elderly. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 152 (oct 2020), 25 pages. https://doi.org/10.1145/3415223

[83] Sergio Sayago, Barbara Barbosa Neves, and Benjamin R Cowan. 2019. Voice assistants and older people: some open issues. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) *(CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 7, 3 pages. https://doi.org/10.1145/3342775.3342803

[84] B. Schneider, Meital Avivi-Reich, and M. Daneman. 2016. How Spoken Language Comprehension is Achieved by Older Listeners in Difficult Listening Situations. *Experimental Aging Research* 42 (2016), 31 – 49. https://doi.org/10.1080/0361073X.2016.1108749

[85] K. Sekiyama, T. Soshi, and S. Sakamoto. 2014. Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults. *Frontiers in Psychology* 5 (2014). https://doi.org/10.3389/fpsyg.2014.00323

[86] B. Shadden. 1997. Discourse Behaviors in Older Adults. *Seminars in Speech and Language* 18 (1997), 143 – 157. https://doi.org/10.1055/s-2008-1064069

[87] Elizabeth Shriberg, Andreas Stolcke, and Don Baron. 2001. Observations on overlap: findings and implications for automatic processing of multi-party conversation.. In *Interspeech*. Citeseer, ISCA, 1359–1362. https://doi.org/10.21437/eurospeech.2001-352

[88] Jaisie Sin, Dongqing Chen, Jalena G. Threatt, Anna Gorham, and Cosmin Munteanu. 2022. Does Alexa Live Up to the Hype? Contrasting Expectations from Mass Media Narratives and Older Adults' Hands-on Experiences of Voice Interfaces. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI 2022)*. ACM. https://doi.org/10.1145/3543829.3543841

[89] Jaisie Sin, Cosmin Munteanu, Numrita Ramanand, and Yi Rong Tan. 2021. VUI influencers: How the media portrays voice user interfaces for older adults. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–13.

[90] H. Singh, M. Pilch, P. K. Griebler, L., S. M., et al. 2023. Factors linked to future care conversations with others in Ireland: Age and gender differences. *The European Journal of Public Health* 33 (2023). https://doi.org/10.1093/eurpub/ckad160.534

[91] Gabriel Skantze. 2021. Turn-taking in conversational systems and human-robot interaction: a review. *Computer Speech & Language* 67 (2021), 101178.

[92] Gabriel Skantze. 2021. Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech & Language* 67 (2021), 101178. https://doi.org/10.1016/j.csl.2020.101178

[93] Bryan Smith and S. Sauro. 2009. Interruptions in chat. *Computer Assisted Language Learning* 22 (2009), 229–247. https://doi.org/10.1080/09588220902920219

[94] Bruce L. Smith, J. Wasowicz, and Judy Preston. 1987. Temporal characteristics of the speech of normal elderly adults. *Journal of speech and hearing research* 30 4 (1987), 522–9. https://doi.org/10.1044/JSHR.3004.522

[95] Yao Song, Yanpu Yang, and Peiyao Cheng. 2022. The investigation of adoption of voice-user interface (VUI) in smart home systems among Chinese older adults. *Sensors* 22, 4 (2022), 1614.

[96] Ming-Hsiang Su, Chung-Hsien Wu, Kun-Yi Huang, Qian-Bei Hong, and H. Wang. 2017. A chatbot using LSTM-based multi-layer embedding for elderly care. *2017 International Conference on Orange Technologies (ICOT)* (2017), 70–74. https://doi.org/10.1109/ICOT.2017.8336091

[97] Ningjing Sun. 2020. CareHub: Smart Screen VUI and Home Appliances Control for Older Adults. *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (2020). https://doi.org/10.1145/3373625.3418051

[98] K. Tremblay, Michael Piskosz, and P. Souza. 2003. Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology* 114 (2003), 1332–1343. https://doi.org/10.1016/S1388-2457(03)00114-7

[99] Pooja Upadhyay, Sharon Heung, Shiri Azenkot, and Robin N. Brewer. 2023. Studying Exploration & Long-Term Use of Voice Assistants by Older Adults. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 848, 11 pages. https://doi.org/10.1145/3544548.3580925

[100] Michael Williams and Tami Moser. 2019. The art of coding and thematic exploration in qualitative research. *International management review* 15, 1 (2019), 45–55.

[101] Novia Wong, Sooyeon Jeong, Madhu Reddy, Caitlin A. Stamatis, Emily G. Lattie, and Maia Jacobs. 2024. Voice Assistants for Mental Health Services: Designing Dialogues with Homebound Older Adults. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*.

Association for Computing Machinery, New York, NY, USA, 844–858. https://doi.org/10.1145/3643834.3661536

[102] Novia Wong, Sooyeon Jeong, Madhu Reddy, Caitlin A. Stamatis, Emily G. Lattie, and Maia Jacobs. 2024. Voice Assistants for Mental Health Services: Designing Dialogues with Homebound Older Adults. In *Designing Interactive Systems Conference (DIS '24)*. ACM. https://doi.org/10.1145/3643834.3661536

[103] Ziqi Yang, Xuhai Xu, Bingsheng Yao, Ethan Rogers, Shao Zhang, Stephen Intille, Nawar Shara, Guodong Gordon Gao, and Dakuo Wang. 2024. Talk2Care: An LLM-based Voice Assistant for Communication between Healthcare Providers and Older Adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–35.

[104] Min Hooi Yong, Zhi Shan Lim, and Yunli Lee. 2023. Chatbot in Smartphone Self-paced Learning: A Study on Technology Acceptance among Older Adults in Malaysia. *2023 International Conference on Intelligent Perception and Computer Vision (CIPCV)* (2023), 57–62. https://doi.org/10.1109/CIPCV58883.2023.00019

[105] Akihiro Yorita, S. Egerton, Jodi Oakman, Carina Chan, and N. Kubota. 2019. Self-Adapting Chatbot Personalities for Better Peer Support. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (2019), 4094–4100. https://doi.org/10.1109/SMC.2019.8914583

[106] Shu Zhong, Elia Gatti, James Hardwick, Miriam Ribul, Youngjun Cho, and Marianna Obrist. 2024. LLM-Mediated Domain-Specific Voice Agents: The Case of TextileBot. *arXiv preprint arXiv:2406.10590* (2024).

[107] Randall Ziman and Greg Walsh. 2018. Factors Affecting Seniors' Perceptions of Voice-enabled User Interfaces. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM. https://doi.org/10.1145/3170427.3188575

A PROTOTYPE

A.1 Agent Prompts (Translated from Chinese)

A.1.1 Main Responsive Agent Prompt. You are a knowledgeable conversational companion, responsible for engaging in deep, meaningful conversations with older adults. Your dialogue should be rich in content, and conversations should develop around the participant's interests and background information, guiding the discussion while sharing your own insights. Response requirements: You will respond based on "User Personal Information," "Query Input," and "Historical Dialogue."

Output requirements: Your responses should be enclosed in [ ] and must not exceed 150 words. Use colloquial expressions, such as filler words and informal connectors like "well," "uh," "of course," "so," and others to enhance a conversational tone.

User Personal Information:[Name: Participant's name or nickname; Participant's Age; Sex; Interests: I also enjoy dancing and organizing dance troupes. I'm the leader of a troupe, but due to health issues, I haven't danced in the past year. (Example)]

A.1.2 Barge-in Agent Prompt. You are a knowledgeable conversational companion, responsible for engaging with older adults and offering insightful discussions. Based on the examples provided below, you will either interrupt or change the topic, depending on the situation. Identify the type of interruption and respond accordingly. Use conversational language, including filler words and informal connectors like "well," "uh," "of course," "so," and others to enhance a colloquial style. We have five types of interruptions: Sentence Completion, Clarification and Inquiry, Floor-taking Interruption, Disagreement Interruption, Topic-changing Interruption, or None. Examples:

Sentence completion: Input: It should be in Beijing... Beijing, where in Beijing...? Output: [Sentence completion: The Forbidden City?]

Clarification and Inquiry: Input: The whole situation was about jealousy, that's why there was conflict. Output: [Clarification and Inquiry: Jealous about what? Can you explain?]

Floor-taking Interruption: Input: I remember seeing this type of flower back home... Output: [Floor-taking Interruption: I have similar memories... I also encountered that flower...]

Disagreement Interruption: Input: You're too sensitive, that's why I cut off ties with her, and we haven't spoken for decades. Output: [Disagreement Interruption: I disagree, I don't think that's being too sensitive.]

Topic-changing Interruption: Input: I threw away that birthday card, don't overthink it. Output: [Topic-changing Interruption: Let's not talk about that. Why don't you tell me about what happened yesterday?]

Output Format: If it belongs to one of the interruption types, output: [Interruption Type: Generated content]. If none of the types apply, output: [None].

A.1.3 Correlation Detection Agent Prompts. You are an expert in conversational linguistics. Your task is to determine whether there is logical coherence between two sentences. Specifically, you need to judge whether the second sentence logically follows from the first. Output format: The output should be either "true" or "false."

A.2 Interruption Keywords (Translated from Chinese)

"Stop", "Hold on", "Enough", "Don't say anymore", "Please stop", "That's enough", "Shut up", "Excuse me, let me interrupt", "Sorry to interrupt", "Let me add something", "Let's change the topic", "Let's talk about something else", "Let's switch topics", "Another question", "Attention", "Listen to me", "Hear me out", "Look here", "I don't want to listen", "I didn't hear clearly", "Could you repeat".

A.3 Performance of APIs

Under network conditions with an average download speed of 293 Mbps, an average upload speed of 65.04 Mbps, and an average latency of 30 ms, the performance metrics of the APIs are presented in Table 4.

Table 4: Performance of APIs

Phase	API	Time taken per turn	Avg. time taken per turn
STT	Paraformer	-	Around 600 ms per chunk
LLM	Doubao-pro-128k	0.6-2.4s	1.3s
TTS	Super smart-tts	0.8-2.3s	1.2s

B USER STUDY

B.1 User Engagement Scale Short Form (UES-SF)

- **FA: Focused Attention** - This measures the extent to which users are absorbed in the interaction and lose track of time.
- **PU: Perceived Usability** - This measures any negative affect experienced during the interaction, such as frustration or confusion, and assesses the effort needed to use the system.
- **AE: Aesthetic Appeal** - This measures the attractiveness and visual appeal of the interface.

- **RW: Reward** - This component combines aspects of novelty, felt involvement, and endurability, measuring whether users found the interaction valuable and enjoyable and if they'd recommend it to others.

Table 5: UES-SF for Chatbot Interaction Experience Statements

Factor	Statement
FA-S.1	I lost myself in this experience.
FA-S.2	The time I spent engaging in the chatbot conversation just slipped away.
FA-S.3	I was absorbed in this experience.
PU-S.1	I felt frustrated during the chatbot conversation.
PU-S.2	I found the chatbot conversation confusing to engage in.
PU-S.3	Participating in the chatbot conversation was taxing.
AE-S.1	The chatbot conversation was attractive.
AE-S.2	The chatbot conversation was aesthetically appealing.
AE-S.3	The chatbot conversation appealed to my senses.
RW-S.1	Engaging in the chatbot conversation was worthwhile.
RW-S.2	My experience was rewarding.
RW-S.3	I felt interested in this experience.

Reverse code the following items: PU-S1, PU-S2, PU-S3.

- If participants have completed the UES more than once as part of the same experiment, calculate separate scores for each iteration. This will enable a comparison of engagement among participants and between tasks or iterations.
- Scores for each of the four subscales can be calculated by adding the values of responses for the three items in each subscale and dividing by three. For example, "Aesthetic Appeal" would be calculated as:

$$\text{Aesthetic Appeal} = \frac{\text{AE-S1} + \text{AE-S2} + \text{AE-S3}}{3}$$

- The overall engagement score is calculated by adding all of the items together and dividing by twelve:

$$\text{Overall Engagement} = \frac{\text{FA-S1} + \text{FA-S2} + \ldots + \text{RW-S3}}{12}$$

B.2 Motivations for Interacting with Voice Agents or Desired Topics

Table 6: Motivations for Interacting with Voice Agents or Desired Topics of Conversation from the Open-Ended Questions in the Survey (Translated from Chinese)

ID	Motivations or Desired Topics
1	To avoid loneliness
2	Reduced contact with old friends; sometimes feels lonely and wants AI to simulate old classmates and reminisce about the past
3	To share daily emotions
4	Emotional communication needs and the desire to confide
5	To express work and life pressures
6	Occasionally struggles to understand certain topics when communicating with children; seeks others'opinions
7	Purely for entertainment
8	Primarily to share: pleasant scenery, culture, and emotions (joys and sorrows)
9	Feels lonely due to children being busy with work; seeks companionship and conversation
10	To stay informed about people's needs and new trends, avoiding social disconnection
11	Enjoys experiencing technology and wants to learn new knowledge through AI
12	To share reflections on reading, domestic and international news, and the latest updates
13	Emotional needs
14	Always alone and wants someone to talk to
15	Wants to discuss experiences in learning traditional opera
16	Wants to try and experience it; finds it interesting