



Speech Interaction Enhances the Perceived Ease of Use of Conversational AI Software: The Moderating Role of Task Type

Yixuan Wang

University of International Business and Economics, Beijing, China

930521473@qq.com

Abstract. Conversational artificial intelligence (AI) is gradually becoming an essential tool in people's lives. This study uses experimental methods to explore the impact of speech interaction and text interaction on users' perceived ease of use of conversational AI, the mediating role of anthropomorphism, and the moderating role of task type on the impact of interaction methods. The results show that interaction methods significantly affect users' willingness to use, with speech interaction enhancing user satisfaction in entertainment-oriented tasks and text interaction having an advantage in information-oriented tasks. Anthropomorphism plays a partial mediating role between interaction methods and users' willingness to use, especially in entertainment-oriented tasks. Task type significantly moderates the impact of interaction methods, with significant differences in users' preferences for interaction methods under different task types. This study provides valuable theoretical and practical guidance for the design and optimization of conversational AI systems, fills the gap in existing research on the relationship between interaction methods and user, and the unresolved mystery of the anthropomorphism, offering a new perspective for understanding users' behavior and psychological mechanisms in different interaction scenarios.

Keywords: Interaction Type, Perceived Anthropomorphism, Perceived Ease of Use, Task Type

1 Introduction

In the digital age, conversational AI has become a daily necessity, with applications ranging from customer service to smart home control and mobile assistants ^[1]. These systems provide information and execute tasks via speech or text interaction ^[2]. However, user satisfaction is influenced by factors like interaction methods, task type, and system design ^[3]. Optimizing interaction methods to boost satisfaction is crucial ^[4].

Speech interaction, noted for its naturality and efficiency, suits hands-occupied scenarios like driving but isn't universally superior. Users prefer text interaction for complex or precision tasks due to its accuracy ^[5] ^[6]. Anthropomorphism, or adding

© The Author(s) 2025

G. M. Lee et al. (eds.), *Proceedings of the 2025 4th International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2025)*, Advances in Intelligent Systems Research 195,

https://doi.org/10.2991/978-94-6463-742-7_41

human traits to AI, enhances user trust and satisfaction but varies in effectiveness based on task type and interaction method^[7]. For example, Siri optimizes the efficiency of driving scenarios through voice interaction, while ChatGPT supports complex coding tasks through text interaction, which confirms the design principle that the interaction mode needs to adapt to the type of tasks.

An experiment was designed to explore these impacts, setting entertainment and information-oriented tasks and having users interact via speech and text. Results showed speech interaction excels in entertainment tasks, while text interaction is better for information tasks. Anthropomorphism significantly mediates perceived ease of use in entertainment tasks but less so in information tasks. The study provides design optimization suggestions for conversational AI systems, filling research gaps on interaction methods and anthropomorphism's role, guiding developers to better meet user needs and improve experience.

2 Literature Review

2.1 Speech Interaction and Text Interaction

The development of speech interaction technology can be traced back to simple voice command recognition, and it has gradually evolved into complex natural language processing systems^[8]. In recent years, the introduction of deep learning technology has significantly improved the performance of speech recognition and synthesis, showing unique advantages in interaction naturality and convenience^[9]. Speech interaction does not require manual input and is especially suitable for scenarios where hands are occupied or quick information is needed, such as driving and cooking. However, speech interaction also has limitations. First, the accuracy of speech recognition may drop significantly in complex noise environments. Second, the multi-turn dialogue ability and context understanding ability of speech interaction systems still need to be improved, and the security and privacy protection issues of speech interaction systems are also increasingly concerned.

Text interaction, as a traditional interaction method, has the advantages of accuracy and traceability. In scenarios requiring precise input or complex task execution, users tend to use text interaction, such as filling out forms, writing code, or conducting complex queries. Text interaction can provide higher accuracy and flexibility and has a natural advantage in privacy protection because user input can be encrypted and anonymized. However, the limitation of text interaction is that it is not natural, requiring manual input, and may be restricted in multi-task operations, with relatively slower interaction speed, especially in scenarios where quick information is needed^[10].

In recent years, many studies have compared speech interaction and text interaction. It has been shown that in simple tasks or when quick information is needed, the efficiency of speech interaction is higher. For example, in querying weather or playing music, speech interaction can quickly respond to user needs. However, in complex tasks or when precise input is needed, the advantage of text interaction is more evident. For example, in filling out complex online forms or debugging code, text inter-

action can provide higher accuracy and flexibility^[11]. In addition, the user's choice of interaction method is also affected by task type and usage scenario. In entertainment-oriented tasks, speech interaction can provide a more natural and interesting interaction experience; while in information-oriented tasks, users may be more concerned with system efficiency and accuracy. Therefore, speech interaction and text interaction each have their advantages, and the choice should be optimized according to specific tasks and usage scenarios.

2.2 Anthropomorphism

Anthropomorphism refers to endowing non-human entities with human characteristics, emotions, or behaviors, making them appear more "human" in the eyes of users. In conversational AI, anthropomorphic design enhances users' trust and affinity for the system through speech intonation, language style, emotional expression, and other means^[12]. The mechanism of anthropomorphism mainly reflects in the following aspects: First, anthropomorphism can enhance users' trust and affinity for the system^[13]. For example, through speech intonation and emotional expression, the system can better communicate emotionally with users. Second, anthropomorphism can improve users' perception of the system's capabilities, making it easier for them to understand and accept the system's behavior^[14]. Finally, anthropomorphism can enhance users' sense of participation and immersion^[15], especially in entertainment-oriented tasks^[16].

The application of anthropomorphism in conversational AI covers various fields, including smart homes, intelligent customer service, and mobile device assistants^[17]. For example, smart home devices enhance the interactivity between users and devices through anthropomorphic design; intelligent customer service systems improve user satisfaction with the system through anthropomorphic design; mobile device assistants enhance users' emotional resonance through speech intonation and emotional expression.

However, anthropomorphism may negatively impact satisfaction if overused (e.g., inflating user expectations)^[18]. Its effectiveness also varies across cultures and scenarios. Designers should tailor anthropomorphic features to task types: prioritize emotional resonance for entertainment tasks, and balance efficiency for information-oriented tasks.

As shown in Table 1, this study adapted a mature scale, translating the items to help participants understand. Since this study focuses on the conversational tool attributes of AI, some emotional items of AI intelligence were deleted.

Table 1. The scale of perceived anthropomorphism.

Items	Original Scale Statement
PA1	The personal intelligent agent is able to speak like a human.
PA2	The personal intelligent agent can be happy.
PA3	The personal intelligent agent can feel love.
PA4	The personal intelligent agent can get upset at times.

PA5	The personal intelligent agent can get frustrated at times.
PA6	The personal intelligent agent can be friendly.
PA7	The personal intelligent agent can be respectful.
PA8	The personal intelligent agent can be funny.
PA9	The personal intelligent agent can be caring

2.3 Technology Acceptance Model and Task Characteristic Theory

In the field of human-computer interaction, users' willingness to use a tool is affected by various factors, including interaction methods, task types, and the overall system design. According to the Technology Acceptance Model (TAM), users' acceptance of technology is mainly influenced by perceived usefulness and perceived ease of use. Perceived usefulness refers to the user's belief that using a certain technology can improve their work or task efficiency, while perceived ease of use refers to the user's belief that using a certain technology is not difficult. In conversational AI tools, perceived usefulness mainly depends on the model's capabilities and the quality of the generated information itself, while interaction methods (speech interaction vs. text interaction) and task types (information-oriented vs. entertainment-oriented) are key factors affecting users' perceived ease of use. This study refers to a mature scale of perceived ease of use to comprehensively reflect users' experience. Given that speech interaction and text interaction have different costs for tool use, items 1 and 5 related to personal learning and tool operation were removed.

In addition, Task Characteristic Theory points out that the nature and complexity of tasks affect users' choice of interaction methods. Information-oriented tasks usually require quick and efficient information acquisition and processing, while entertainment-oriented tasks focus more on emotional resonance. Under different task types, users' preferences for interaction methods may differ, thereby affecting their willingness to use the tool. The scale of perceived ease of use can be seen in Table 2.

Table 2. The scale of perceived ease of use.

Items	Original Scale Statement
PEU1	Learning to operate the AI would be easy for me.
PEU2	I would find it easy to get the AI to do what I want it to do.
PEU3	My interaction with the AI would be clear and understandable.
PEU4	I would find the AI to be easy to use.
PEU5	I would find it easy to become skillful at using the AI .
PEU6	I would find the AI to be flexible to interact with.

3 Hypothesis Proposal

Speech interaction, with its naturalness, convenience, and efficiency, shows significant advantages in scenarios where hands are occupied or quick information is needed, such as driving, cooking, or performing other multi-task operations, significantly

enhancing user experience. However, text interaction has an advantage in scenarios requiring precise input or complex task execution, such as filling out forms, writing code, or conducting complex queries, with higher accuracy and flexibility. Despite this, text interaction requires manual input, which is relatively cumbersome and less efficient in multi-task operations. Therefore, this study proposes that input mode (speech interaction vs. text interaction) may significantly affect users' willingness to use AI tools.

H1: Input mode has a significant impact on users' willingness to use AI tools.

Anthropomorphism refers to endowing non-human entities with human characteristics, emotions, or behaviors, making them appear more "human" for users. In conversational AI tools, speech interaction can simulate the communication method between people, making it easier for users to perceive the anthropomorphic characteristics of the AI tool. Speech interaction has stronger naturalness and emotional expressiveness, which can enhance users' perception of anthropomorphism through speech intonation, rhythm and others. For example, a voice assistant can respond to users with a warm intonation and humorous language, making it appear more "human," and this characteristic gives it an advantage in anthropomorphic design, significantly enhancing users' trust and affinity for the system.

In contrast, while text interaction can enhance users' perception of anthropomorphism through humorous, witty, or caring language styles, this perception may not be as intuitive and strong as that of speech interaction. Anthropomorphic design can enhance the emotional resonance between users and the system, making users more likely to have a good feeling and trust for the system, thereby improving users' perceived ease of use and making them more willing to use the tool. In addition, anthropomorphic design, by adding "human" characteristics to the system, makes users consider the system more reliable and easier to understand, thereby significantly improving users' acceptance of the tool. Therefore, this study proposes that anthropomorphism may mediate the relationship between input mode and users' willingness to use AI tools.

H2: Perceived anthropomorphism mediates the relationship between input mode and users' willingness to use AI tools.

According to Task Characteristic Theory, the nature and complexity of tasks affect users' choice of interaction methods. Information-oriented tasks usually require quick and efficient information acquisition and processing, while entertainment-oriented tasks focus more on emotional resonance and fun. This study believes that under different task types, users' preferences for interaction methods may differ, thereby affecting their willingness to use the tool.

H3: Task type moderates the impact of input mode on users' willingness to use AI tools.

In information-oriented tasks, users' main goal is to efficiently complete the process and obtain accurate information. Text interaction can quickly respond to user needs and provide accurate text output to users, thereby improving task completion efficiency. According to the Technology Acceptance Model (TAM), perceived ease of use is a key factor affecting users' willingness to use. Speech interaction may not be able to provide visual results in task-oriented scenarios and may have lower effi-

ciency in multi-task operations. Therefore, this study proposes that text interaction may have a greater impact on users' willingness to use in task-oriented AI tools than speech interaction.

H3a: For information-oriented tasks, text interaction has a greater impact on users' willingness to use than speech interaction.

In entertainment-oriented tasks, users' goals are to obtain emotional resonance, social experience, or entertainment effects. Speech interaction can enhance users' immersion through emotional voice and intonation, while text interaction can attract users through humorous and witty responses. Since entertainment-oriented tasks are more inclined to provide real-time feedback on users' feelings, speech interaction may have an advantage in the process of information-oriented tasks.

H3b: For entertainment-oriented tasks, speech interaction has a greater impact on users' willingness to use than text interaction.

The research model can be seen as the figure 1.

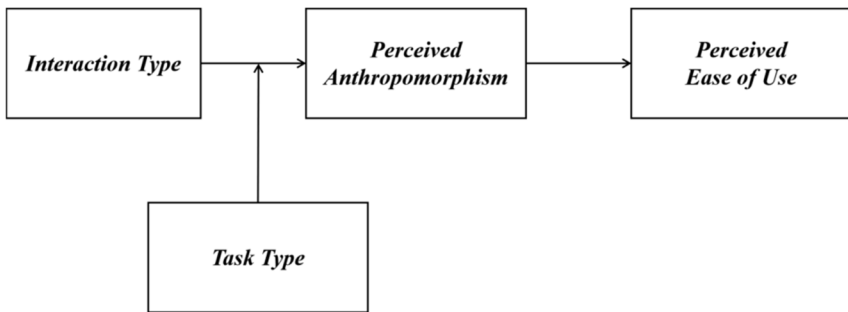


Fig. 1. Research Model.

4 Research Design

This study uses a scenario experimental method to verify the above hypotheses and explore the impact of interaction methods, task types, and anthropomorphism on users' willingness to use AI. The experiment adopts a 2 (interaction type: speech interaction vs. text interaction) \times 2 (task type: information-oriented vs. entertainment-oriented) between-subjects experimental design, with 4 groups. After authenticity review, the study retained data from 243 participants (60.1% female; Mage = 26.54, SD = 6.087), and each participant received a 2-yuan red packet transfer as compensation. The samples of this study mainly focus on young Chinese groups, and future research can be extended to different age groups and cross-cultural users to verify the universality of the conclusions.

4.1 Experimental Materials

In terms of experimental tools, after careful comparison and analysis, this study ultimately chose the Kimi as the experimental test tool. The Kimi model performs out-

standingly in speech recognition and synthesis, natural language processing, and response speed, providing users with a smooth and natural interaction experience. Researchers provided an embedded link to the Kimi applet to ensure that participants could conveniently use this AI tool during the experiment.

Regarding task instructions, this study was divided into four experimental groups, corresponding to two types of experimental tasks: information-oriented tasks and entertainment-oriented tasks. Information-oriented tasks (e.g., travel planning) require precision and logical coherence, while entertainment-oriented tasks (e.g., emotional communication) focus on emotional resonance. The classification is based on the core objectives of tasks (efficiency vs. emotional needs). For participants in the information-oriented task group, researchers clearly prompted them with "You will use the Kimi intelligent assistant to customize a travel guide for your trip to a certain place," aiming to guide participants to focus on obtaining accurate and efficient information. For participants in the entertainment-oriented task group, they were informed with "You will use the Kimi intelligent assistant to inquire about the feelings of traveling to a certain place," which is more inclined to guide participants to engage in emotional communication and interaction with the intelligent assistant. Through the task instructions, effective reduction of experimental endogeneity errors can be achieved.

4.2 Experimental Process

In terms of pre-test preparation, to ensure that the experimental design is close to the real-life usage scenarios of AI tools and to avoid interference from irrelevant variables, this study strictly required participants to have experience using AI tools when selecting the experimental sample. Before the experiment, researchers used carefully designed questionnaires to preliminarily screen the usage experience of participants to ensure that they had a certain foundation. At the same time, researchers provided detailed introductions to the specific operation methods and precautions of the two interaction methods (speech interaction and text interaction) involved in the experiment through clear text and picture teaching instructions, ensuring that participants could correctly and skillfully use the corresponding interaction methods. Before the official start of the test, the experimenter also asked participants questions to further confirm that they had fully understood the actions they needed to take.

In the test phase, after completing the pre-test preparation, participants randomly drew an interaction type and received the corresponding questionnaire. Subsequently, participants entered the usage interface of the Kimi AI tool through the embedded applet. Participants in each group were randomly assigned to a task, and after clearly understanding the instructions, they began the test. During the test, participants needed to complete the assigned tasks and interact with the Kimi intelligent assistant within a specified time (each test lasted 3 minutes). Researchers recorded the interaction process and related data of participants through the background for subsequent in-depth analysis.

After the test, participants were required to fill out the subsequent questionnaire content. First, researchers asked each participant about the chosen travel destination to ensure that participants had indeed carried out the operations according to the task

instructions during the experiment. Next, participants were asked to fill out the anthropomorphism scale and the AI tool usage intention scale. Based on the pre-survey of the satisfaction scale for conversational AI agents, this study carefully adjusted the scale items and finally adopted two formal items for anthropomorphism scale and AI tool usage intention scale, using a Likert five-point scale for measurement, with "1 (strongly agree)" to "5 (strongly disagree)" to accurately reflect participants' subjective feelings about the degree of anthropomorphism and usage intention. Finally, participants completed the manipulation check, gender, age, and income, and other demographic information. At the end of the experiment, researchers promptly collected the experimental results and provided corresponding compensation and thanks to the participants.

5 Research Results and Discussion

This study used the mediation model (Model 7) proposed by Hayes (2009) and the Bootstrap method (5000 samples) to examine the moderating effect of task type (W) on the impact path of interaction type (X) on perceived ease of use (Y) through perceived anthropomorphism (M). The results showed that after controlling for task type, the main effect of interaction type on perceived anthropomorphism was not significant ($\beta = 0.0789$, $p = 0.705$), and the interaction effect of interaction type and task type (Int_1) also did not reach a statistically significant level ($\beta = 0.3727$, $p = 0.200$). However, simple effect analysis indicated that when the task type was information-oriented ($W = 0$), the mean difference in perceived anthropomorphism between text interaction ($X = 0$) and speech interaction ($X = 1$) was 0.0789 (5.1852 vs. 5.1063), and when the task type was entertainment-oriented ($W = 1$), the difference expanded to 0.4516 (5.4516 vs. 5.0000).

In the second-stage model, perceived anthropomorphism had a significant positive impact on perceived ease of use ($\beta = 0.3926$, $p < 0.001$), while the direct effect of interaction type was not significant ($\beta = -0.0452$, $p = 0.606$). Conditional indirect effect analysis found that in entertainment-oriented tasks, the indirect effect of interaction type on perceived ease of use through perceived anthropomorphism was significant (Effect = 0.1773, BootLLCI = 0.0196), while in information-oriented tasks, this path was not significant (Effect = 0.0310, BootLLCI = -0.1296). The mediation effect index showed that the difference in indirect effects between the two types of tasks did not reach a statistically significant level (Index = 0.1463, BootLLCI = -0.0760).

The results show that compared with text interaction, speech interaction is more likely to enhance users' perception of ease of use by improving perceived anthropomorphism, but this mechanism only holds in entertainment-oriented tasks. Although the interaction item did not reach a significant level, the difference in conditional indirect effects under different task types suggests that the task context may play a situational role in the process of anthropomorphism perception in technical interactions.

This study has potential limitations, such as user subjective evaluation may be affected by experimental scenarios, and bias can be reduced through longitudinal experiments or physiological indicators (such as eye tracking) in the future. In addition, task types are only divided into two categories, and interaction preferences for complex tasks, such as educational consultation, need to be further explored. Based on this research, developers can optimize the design: (1) entertainment tasks prioritize voice interaction and strengthen anthropomorphic tone; (2) The information task adopts text interaction and weakens emotion expression to improve efficiency; (3) Provide hybrid interaction mode for users to switch on demand.

6 Conclusion

This study used experiments to examine how speech and text interaction affect users' willingness to use conversational AI tools, the mediating role of anthropomorphism, and task type's moderating effect. The findings offer guidance for AI system design. First, interaction methods significantly impact usage willingness. Speech interaction excels in entertainment tasks, enhancing anthropomorphism perception, trust, and affinity, thus boosting usage willingness. Text interaction is superior in information tasks, providing accurate information quickly and improving efficiency, which increases usage willingness.

Second, anthropomorphism partially mediates the relationship between interaction methods and usage willingness. Speech interaction enhances trust and affinity by increasing anthropomorphism perception, improving user experience. However, this mediating effect is stronger in entertainment tasks and weaker in information tasks, where users prioritize efficiency and accuracy.

Finally, task type significantly moderates the impact of interaction methods. In information tasks, speech and text interaction have similar effects on perceived ease of use. In entertainment tasks, speech interaction's impact on usage willingness is comparable to text interaction. This highlights the importance of task type in choosing interaction methods. When designing conversational AI systems, it's crucial to select appropriate interaction methods based on task types and user needs to enhance satisfaction and usage willingness.

References

1. Cai, W., & Chen, L. Predicting user intents and satisfaction with dialogue-based conversational recommendations. *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 1–10(2020).
2. Jiang, H., Cheng, Y., Yang, J., & Gao, S. AI-powered chatbot communication with customers: Dialogic interactions, satisfaction, engagement, and customer behavior. *Computers in Human Behavior*, 134-139(2022).

3. Palma, P., Whitford, V., & Titone, D. Cross-language activation and executive control modulate within-language ambiguity resolution: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 507–528(2020).
4. Ansara, C. AI and prompt architecture – A literature review. *International Journal of Computer Applications*, 39–45(2023)
5. Dan, W., & Jin, Q.. A study on users' collaborative information seeking behavior and system evaluation: A perspective of tasks and collaborative abilities. *Shuju Fenxi Yu Zhishi Faxian*, 62–68(2012).
6. Riefler, L., & Benz, C. User preferences for interaction modalities: The influence of task, context, and user characteristics when interacting with conversational agents. *Proceedings of the 57th Hawaii International Conference on System Sciences*, 3960–3969(2024).
7. Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340(1989).
8. Kou, P. The exploration of human-computer interaction technology based on intelligent voice technology. *Science and Technology of Engineering, Chemistry and Environmental Protection*, 10-15(2024).
9. Tokala, V., Grinstein, E., Brookes, M., Doclo, S., Jensen, J., & Naylor, P. A. Binaural speech enhancement using deep complex convolutional transformer networks. *Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing*, 10-18(2024).
10. Cheng, M.-Y., Gupta, A. Coevolutionary multitasking for concurrent global optimization: With case studies in complex engineering design. *Engineering Applications of Artificial Intelligence*, 13–24(2017).
11. Alkathiri, M. S. Artificial intelligence assisted improved human-computer interactions for computer systems. *Computers and Electrical Engineering*, 101-105(2022)
12. Blut, M., Wang, C., Wunderlich, N. V., & Brock, C.. Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, 632–658(2021).
13. Li, X., & Sung, Y. Anthropomorphism brings us closer: The mediating role of psychological distance in user–AI assistant interactions. *Computers in Human Behavior*, 118-119(2021).
14. Klein, K., & Martinez, L. F.. The impact of anthropomorphism on customer satisfaction in chatbot commerce: An experimental study in the food sector. *Electronic Commerce Research*, 2789–2825(2022).
15. Pfeuffer, N., Benlian, A., Gimpel, H., & Hinz, O. Anthropomorphic information systems. *Business & Information Systems Engineering*, 523–533(2019).
16. Roy, R., & Naidoo, V. Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research*, 23–34(2021).
17. Araujo, T. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 183–189(2018).
18. Araujo, T. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183–189(2018).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

