Human-Centered Design and User Experience, Vol. 114, 2023, 495–504
https://doi.org/10.54941/ahfe1004268

Exploring Human-Like Behavior Explanation on AI Speaker Recognition as Communicable Partners Across Age Groups

Yukiko Nishizaki and Takumi Uchitani

Kyoto Institute of Technology, Japan

ABSTRACT

This paper investigated whether explaining Human-Like AI speakers behave like humans leads to their perception as communicable partners, irrespective of users' characteristics. Two age groups, the elderly and younger participants, were randomly assigned to two conditions: human-like and machine-like. The experimental results revealed a significant Simon effect, indicating that participants in the human-like condition perceived the AI speaker as a more social communication partner than in the machine-like condition. This effect was consistent across both age groups, suggesting that individual characteristics did not influence the recognition of AI speakers as communicable partners. The findings from this study suggests that AI speakers could be perceived as communicable partners when users are informed of their ability to "think and judge by themselves like a human."

Keywords: AI speaker, Human-agent interaction, Elderly participants

INTRODUCTION

AI speaker devices been used on daily bases in various countries. Most AI speakers have a simple shape similar to conventional speakers because their primary function is to output audio. As a result, they differ significantly in appearance from communication-oriented robots, such as Pepper® (SoftBank Corp.) and AIBO® (SONY Corp.). The primary role of AI speakers is to play music, provide weather updates, and perform other relatively simple tasks.

As the use of AI speakers becomes more widespread, researchers are considering the possibilities for their use beyond simple voice presentation functions. This includes exploring their potential as communicable partners for various demographics, such as the elderly (Kowalski et al., 2019) and young children (Lovato et al., 2019). These studies have viewed AI speakers not only as tools for presenting information, but also as communication partners capable of sharing thoughts, akin to interactions with friends or family members.

However, there has been limited research on users' perceptions of AI speakers. Pradhan et al., (2019) which found that individual differences exist in how users perceive AI speakers, with some groups of users viewing them as machines, similar to radios, while others perceive them as social agents, resembling humans. These varying perceptions are influenced by users' past experiences. Consequently, these differences in how each user perceives AI speakers may impact their ability to engage with AI speakers as social entities. Further investigation is needed to understand and explore these user perceptions, which can provide valuable insights for optimizing the interaction and relationship between users and AI speakers.

Pitardi et al., (2021) highlighted that the continuous listening feature of AI speakers, who are always attentive to surrounding voices poses a challenge in establishing trust with users (Mclean et al., 2019). However, this issue primarily relates to users' trust in the manufacturers or brands behind the AI speakers rather than directly impacting trust in the AI speaker itself. As a result, it becomes crucial to investigate strategies for fostering a more profound and trusting relationship between AI speakers and users.

Verbal Instruction and Recognition to Agents

To effectively communicate as communicable partners, it is essential to recognize others as distinct entities. For instance, Stenzel et al., (2012) conducted an experiment using a cooperative task involving both humans and humanoid robots; hereinafter referred to as "humanoid robots". Participants were divided into two groups: the human-like group, where they were informed that the robot "thinks and judges by itself like a human," and the machine-like group, where they were informed that it "is programmed to operate like a machine." Additionally, the researchers utilized the Joint Simon task, first introduced by Sebanz et al., (2003), to assess how participants perceived the humanoid robot as an intentional agent, which was referred to as a "communicable partner". The Joint Simon task demonstrated a stronger Joint Simon effect when participants had a partner to perform the task with, highlighting the ability to recognize the task partner as another entity. The results of the experiment indicated that the Joint Simon effect was significantly more prominent in the human-like group than in the machine-like group. The researchers suggested that differences in recognizing autonomy based on verbal instructions influenced how participants perceived humanoid robots as communicable partners.

While Sebanz et al., findings were pivotal for the potential use of humanoid robots as agents which were similar to humans, their experiments were limited to robots with human-like appearances. Consequently, they did not explore robots with appearances unlike humans or animals, such as AI speakers. Furthermore, Sayago et al., (2019) emphasized the importance of investigating the relationship between AI speakers and the elderly, given the aging population. This study aims to focus on examining this relationship in-depth.

Recognition of Agent and Trust

Glikson et al., (2020) highlighted the potential to improve low trust levels in robotics AI through interaction while expressing their concern about the possibility of reducing high trust levels in other AI technologies, such as embedded AI and virtual AI on the screen, due to technology errors. The study also emphasized that low trust levels in AI agents not only lead to non-use but can also result in misuse and abuse. Additionally, a scenario-based study by Tussyadiah et al., (2020) reported a strong negative correlation between trust in robotics AI and Negative Attitudes toward the Robots Scale (NARS, Nomura et al., 2010).

These findings collectively suggested that prior perceptions and recognition of an AI agent significantly influence the level of trust in it, which needs to be carefully controlled. However, it is noteworthy that Glikson's study did not address agents like AI speakers, whose classification as either robotics AI or embedded AI remains ambiguous. Further research may be required to clarify the categorization and trust implications of AI speakers and similar agents in human-robot interaction.

Purpose of This Study

The recognition of agent autonomy by verbal instruction affected whether or not the agent was perceived as a communicable partner, and the classification of the agent users recognized affected the transition in its trust level. Nevertheless, it should have been mentioned whether the AI speaker, whose appearance did not resemble animals or humans, followed the same pattern. Additionally, it was not apparent that verbal instruction affected the recognition of agents regardless of the user's characteristics (such as age and experience).

Therefore, the purpose of this study was to clarify the impact of differences in recognition of AI speakers on communication and to obtain knowledge that could be used to consider the practical situation and design policy for using them as trusted partners. As a preliminary step, we focused on the first situation. We investigated verbal explanations about autonomy that made AI speakers recognized as communicable partners, similar to agents that appear like animals or humans, irrespective of the user's characteristics, such as age and experiences in using robots.

The authors hypothesize that: (1) participants who were given the explanation that the AI speaker "thinks and judges by itself like a human" would recognize it as a communicable partner more strongly than participants who were informed that it "is programmed to operate like a machine"; Moreover, (2) this difference would be independent of the user's characteristics, such as age and experiences in using robots.

METHODS

Participants

In this study, 80 participants included in two age groups: Elderly group (n = 40, 22 females, mean age = 77.4 ± 3.79 years); and Younger group (n = 40, 19 females, mean age = 21.3 ± 1.37). To exclude the possibility of including those with special knowledge of AI, students majoring in information engineering were notincluded as participants. Participants who use

AI speakers at home on a daily basis were also not included. The study was approved by ethics committee of the Kyoto Institute of Technology, and all participants signed an informed consent form prior to participation.

Experimental Design

The following variables were manipulated in 2 × 2 factorial design: age (elderly, younger) and condition (human-like, machine-like). In the human-like condition, participants were given the explanation that an AI speaker used in the experiment would think and make decisions on its own, just like a human. In the machine-like condition, participants were explained that an AI speaker was programmed to operate like a machine. Each age group participants were randomly divided into two conditions. Elderly participants were assigned to 20 for the human-like condition (11 females, mean age = 74.3 ± 4.1 years) and 20 for the machine-like condition (11 females, mean age = 74.6 ± 3.6 years). Younger participants were assigned to 20 for the human-like condition (8 females, mean age = 19.9 ± 4.2 years) and 20 for the machine-like condition (11 females, mean age = 21.4 ±1.6 years).

Experimental Task

This study employs the "Joint Simon task" to determine whether participants would perceive the AI speaker as a reliable communication partner. The Joint Simon task (also called the social Simon task) is a variant of the Simon task (Simon, 1969), a cooperative assignment performed by two participants. Dolk et al., (2014) explained it that has been developed to investigate how and to what people mentally represent their own and other persons' action/-task and how these cognitive representations influence an individual's own behavior when interacting with another person. The joint Simon task is not only a task that represents joint action between people; instead, it is also used in studies that show the state of collaboration between people and robots or Agents.

In basic Simon Task, two words were displayed on the right or left side of a centralized cross. Then, participants were required to press "j" key (left side of the keyboard) if they saw the word (even if it appeared to the right of the cross) and to press "k" key (right side of the keyboard) if they saw the other word (even if it appeared to the left of the cross). Since the reaction to the button location is promoted by the stimulus unrevealed to the task (the side that words were displayed), the reaction time is shortened if the side of the button to be pressed (left or right) and the side of the stimulus (left or right) match. and lengthened if they do not match. This difference was called "Simon effect".

As the task partner was an AI speaker, auditory stimuli were used instead of visual stimuli, referencing Vu (2003). The auditory stimuli were played from speakers located on either side of the participants. Both the participants and the AI speaker, serving as the agent, were required to perform the task collaboratively.

The auditory stimulus consisted of the words "A" or "I". Two types of keys were provided for responding to "A" or "I". A key with a red sticker for "A" was positioned on the left side of the key, as seen by the participant, and a key with a blue sticker for the "I" response was positioned on the right side. The participant's role was to respond to "A". Whenever the participant heard "A" from either loudspeaker, they were instructed to press the key with the red sticker.

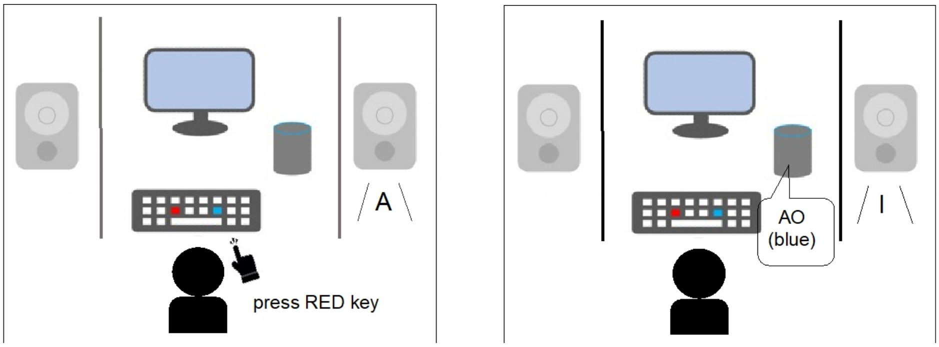

Figure 1: Schematic of the joint Simon task (Left picture: role of the participant, right picture: role of the AI speaker).

The AI speaker, acting as the agent and manipulated by the experimenter as a "Wizard-of-Oz," was responsible for responding to "I". (see Figure 1). The agent responded with "ao" (meaning blue in Japanese and having a blue sticker on the key) regardless of which loudspeaker emitted the "I" stimulus.

The word "A" or "I" was presented 1100 ms after the participant pressed the space key or 3000 ms after the previous trial. "The congruency condition" was defined as the left and right speakers emitting the auditory stimulus corresponding to the left and right sides of the key. It referred to a condition in which the left speaker emitted the letter "A," and the participant was required to respond to the key with the red sticker on its left side, as perceived from the participant's perspective. On the other hand, "the incongruent condition" was defined when the left and right speakers emitted the auditory stimulus, but the left and right keys did not match. It occurred when the right speaker emitted the letter "A," and the participant was required to respond to the key with the red sticker on the left side of the speaker, as seen by the participant. Figure 1 is a conceptual view of the Joint Simon Task.

Procedure

Alexa (voice assistant of Amazon services) was utilized as the agent. To prevent the potential for some participants to be aware of Alexa being an AI speaker and its functionalities, we applied a lab sticker over the logo to conceal its identity as Alexa (See Figure 2). The AI speaker function was not employed in this experiment; instead, it solely played the PC's voice as the primary speaker. For the agent's voice, voiceroid (developed by AHS) was employed.

Before the experiment, all participants completed the NARS (Negative Attitudes toward Robots Scale) and the Robot Anxiety Scale (RAS, Nomura et al., 2010). The NARS measured negative attitudes toward robots, while the RAS assessed anxiety toward robots.

The experimental setup was as follows: The agent, keyboard, and display were positioned on a desk. Speakers emitting the task's audio were placed on either side of the participants, separated by partitions that were not visible to the participants. All participants underwent four trials of the Joint Simon task, with each trial consisting of 50 individual trials. Before each trial, the participants were informed that the AI speaker would behave either like a human or like a machine, depending on the condition. They were instructed to perform the task as quickly and accurately as possible, and a two-minute break was provided between trials.

Figure 2: AI speaker as the agent used in experiment.

RESULTS

Joint Simon Effect

In the Joint Simon task, reaction time (RT) in milliseconds was measured from when the speaker emitted the sound until the participant pressed the key. These RT values were averaged across all other blocks of trials, excluding data from the first trial of each block and trials in which participants failed to respond within 3000 ms.

The keys with red stickers were positioned on the left side, while the keys with blue stickers were placed on the right. The congruent condition was defined when the keys' left and right sides matched the speaker's left and right sides from which the audio was presented. The incongruent condition was defined when the critical sides did not match the speaker sides. The Simon effect was calculated by subtracting the congruent condition's reaction time from the incongruent condition's reaction time.

In the human-like condition, the mean of the elder group's RT was 39.80 ($SD = 9.34$) and the mean of the younger group's RT was 26.18 ($SD = 6.46$). In the machine-like condition, the mean of the elder group's RT was 29.35 ($SD = 6.78$) and the mean of the younger group's RT was 10.27 ($SD = 5.42$). The RTs were found to follow a normal distribution, so the data were subjected to a two-way analysis of variance (ANOVA) with the factors being the condition (human-like, machine-like) and age (elder/younger). It is noteworthy that the Simon effect is known to be amplified in the presence of a collaborative partner, referred to as the joint Simon effect. The results showed a significant difference in the condition ($F(1, 76) = 4.17$, $p = .05$) and the age ($F(1, 76) = 4.69$, $p = .05$). However, there was no significant interaction between the condition and age ($F(1, 76) = 0.27$, $p = .60$). The results are shown in Figure 3.

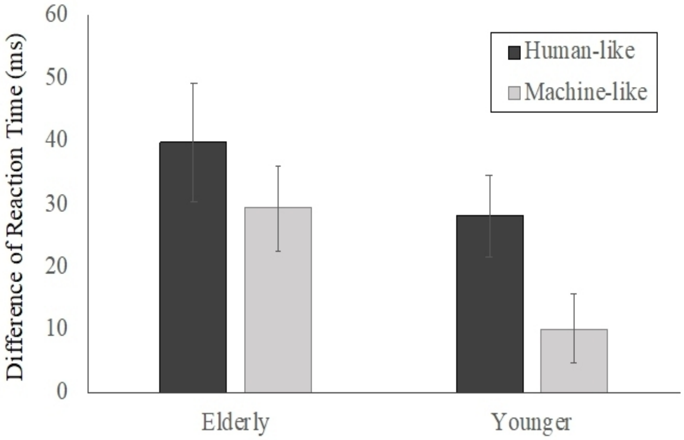

Figure 3: The result of the joint Simon task.

Influence of Attitudes on Robots'

We analysed the relationship between individual differences in how participants perceived the robot and the Simon effect. First, the means and standard deviations (SDs) of the NARS and the RAS scores for the elderly and younger groups are presented in Table 1. There were no significant differences between the elderly and younger groups on either of the scales.

Table 1. The NARS and the RAS scores in two groups.

	NARS		RAS	
	Mean	**SD**	**Mean**	**SD**
Elderly	2.87	0.49	3.70	0.73
Younger	2.76	0.59	3.50	0.97

Next, a correlation analysis was conducted between the NARS and RAS scores, respectively, and the Simon effect (reaction time). The results showed no significant correlation for each scale (NARS: $r(78) = 0.34$, $p = .73$; RAS: $r(78) = 0.45$, $p = .66$).

DISCUSSION

This study aimed to investigate whether explaining human-like AI speakers which behave like humans would lead to their perception as communicable partners, irrespective of users' characteristics, such as age and experience with using robots. We formulated the following hypotheses: (1) Participants informed that the AI speaker “thinks and judges by itself like a human” would perceive it as a more vital communicable partner compared to those informed that it “was programmed to operate like a machine.” (2) These differences in perception would remain consistent regardless of the user’s characteristics, including age and experience with using robots.

The experimental results demonstrated that the difference in reaction time between the congruent and incongruent conditions was more pronounced in the human-like condition than in the machine-like condition, with a significant Simon effect observed, thereby supporting our hypotheses. Additionally, this trend was consistent across the elderly and younger groups, suggesting that individuals in the human-like condition perceived the AI speaker as a more social communication partner. Significantly, individual differences in the perception of the robot did not influence this effect. Consequently, in initial encounters, AI speakers could be perceived as communicable partners akin to social robots whose appearances resemble animals or humans when users are informed that the AI speaker can “think and judge by itself like a human.”

On the other hand, there was a significant difference in the Simon Effect between the age groups. However, the interaction effect was insignificant. Figure 3 showed a similar tendency between the elderly and the younger group, suggesting that the difference in the Simon Effect is more strongly influenced by age-related differences in attentional function than by differences in how the AI speaker was recognized. The effect of age-related differences on performance in the Simon Task was previously described by Yano et al., (2010).

In conclusion, providing explicit verbal explanation for AI speakers could autonomously possibly provide enhanced users’ recognition of them as communicable partners in the early-stage communication, regardless of users’ characteristics, such as age, experience with using robots, and attitude towards robots. Additionally, it is worth noting that in this study, the actual agent used in the experiment was a regular speaker presented as an AI speaker with the explanation that it behaves like a human, and participants recognized it as an AI speaker without any doubt. Therefore, this finding may have implications for designing interactions with other devices whose appearances do not resemble animals or humans, such as autonomous cars, in order to effectively engage with people.

Limitations of the Current Study

In this study, we explored the recognition of AI speakers in a first-meeting situation. However, our ultimate goal is to utilize AI speakers as trustworthy partners, akin to friends or family members, in the future. The present study represents only a preliminary step towards building such relationship. Therefore, to investigate the development of intimacy between AI speakers and users, it is essential to examine how to instil trust in users toward AI speakers from the first interaction through continuous engagement (Glikson et al., 2020).

Moreover, we acknowledge a limitation of the Wizard-of-Oz method used in the laboratory. It prevented researchers from investigating the long-term effects of recognition that may arise when users interact with AI speakers in their daily lives over an extended period. To address this limitation, it is necessary to conduct field-based experiments where users communicate with actual AI speakers in real-life scenarios and live with them for an extended duration, such as over a month while being informed that the AI speaker can behave like a human or is operated like a machine.

REFERENCES

Dolk, T., Hommel, B., Colzato, L, S., Schütz-Bosbach, Si., Prinz, W., and Liepelt, R. (2014) The joint Simon effect: A review and theoretical integration, Frontiers in Psychology, Vol. 5, doi: 10.3389/fpsyg.2014.00974.

Glikson, E., and Woolley, A. W. (2020) Human trust in artificial intelligence: Review of empirical research, Academy of Management Annals, Vol. 14, doi.org/10.5465/annals.2018.0057.

Kowalski, J., Jaskulska, A., Skorupska, K., Abramczuk, K., Biele, C., Kopeć, W., and Marasek, K. (2019) Older Adults and Voice Interaction: A Pilot Study with Google Home, CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, May 2019 Paper No.: LBW0187, pp. 1–6.

Lovato, S. B., Piper, A. M., and Wartella, E. A. (2019) Hey google, Do unicorns exist?: Conversational agents as a path to answers to children's questions, Proceedings of the 18th ACM International Conference on Interaction Design and Children, pp. 301–313.

McLean, G., and Osei-Frimpong, K. (2019) Hey Alexa examine the variables influencing the use of artificial intelligent in-home voice assistants, Computers in Human Behavior, Vol. 99, pp. 28–37.

Nomura, T., Kanda, T., Suzuki, T., Yamada, S., and Kato, K. (2010) Human Attitudes, Anxiety, and Behaviors in Human-Robot Interaction (HRI), Proceedings of 26th Fuzzy System Symposium, pp. 554–559 (in Japanese).

Pitardi, V., and Marriott, H. R. (2021) Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice-based artificial intelligence, Psychology & Marketing, 38, pp. 626–642.

Pradhan, A., Findlater, L., and Lazar, A. (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, Vol. 3, issue CSCW, Article 214, pp. 1–21.

Sayago, S., Neves, B. B., and Cowan, B. R. (2019) Voice assistants and older people: some open issues. Proceedings of the 1st International Conference on Conversational User Interfaces, pp. 1–3.

Sebanz N., Knoblich G., Prinz W. (2003). Representing others'actions: just like one's own? Cognition 88, B11–B21.

Simon, J. R. (1969) Reactions toward the source of stimulation, Journal of Experimental Psychology, 81, pp. 174–176.

Stenzel A., Chinellato E., Tirado Bou M. A., del Pobil Á. P., Lappe M., Liepelt R. (2012) When humanoid robots become human-like interaction partners: co-representation of robotic actions, Journal of Experimental Psychology: Human Perception and Performance, vol. 38, pp. 1073–1077.

Tussyadiah, I. P., Zach, F. J., and Wang, J. (2020) Do travelers trust intelligent service robots?, Annals of Tourism Research, Vol. 81, March, 102886, doi.org/10.1016/j.annals.2020.102886.

Vu, K.-P. L., Proctor, R. W., and Urcuioli, P. (2003) Transfer effects of incompatible location-relevant mappings on a subsequent visual or auditory Simon task, Memory & Cognition, Vol. 31, pp. 1146–1152.

Yano. M. (2010) Congruency sequence effects in cognitive interference paradigms and negative priming, Proceedings of the Annual convention of the Japanese Psychological Association (in Japanese).