What If Conversational Agents Became Invisible? Comparing Users' Mental Models According to Physical Entity of AI Speaker

SUNOK LEE, Department of Industrial Design, KAIST, Republic of Korea
MINJI CHO, Department of Industrial Design, KAIST, Republic of Korea
SANGSU LEE, Department of Industrial Design, KAIST, Republic of Korea

The popularity of conversational agents (CAs) in the form of AI speakers that support ubiquitous smart homes has increased because of their seamless interaction. However, recent studies have revealed that the use of AI speakers decreases over time, which shows that current agents do not fully support smart homes. Because of this problem, the possibility of unobtrusive, invisible intelligence without a physical device has been suggested. To explore CA design direction that enhances the user experience in smart homes, we aimed to understand each feature by comparing an invisible agent with visible ones embedded in stand-alone AI speakers. We conducted a drawing study to examine users' mental models formed through communicating with two different physical entities (i.e., visible and invisible CAs). From the drawings, interviews, and surveys, we identified how users' mental models and interactions differed depending on the presence of a physical entity. We found that a physical entity affected users' perceptions, expectations, and interactions toward the agent.

CCS Concepts: • **Human-centered computing** → **User studies**.

Additional Key Words and Phrases: Ubiquitous smart home, conversational agent, voice user interface, invisible intelligence, drawing study

ACM Reference Format:
Sunok Lee, Minji Cho, and Sangsu Lee. 2020. What If Conversational Agents Became Invisible? Comparing Users' Mental Models According to Physical Entity of AI Speaker. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 88 (September 2020), 24 pages. https://doi.org/10.1145/3411840

1 INTRODUCTION

Due to its seamless and unobtrusive input method, the use of voice user interface (VUI) has increased in the form of AI speakers as hubs to control smart homes [11, 39, 45]. Because commercialized AI speakers consist of a physical device with a speaker and microphone, users need to face the device for better communication, as people do when they communicate face-to-face. Users interact with CAs in stand-alone devices, such as Alexa in Amazon Echo, which makes users look at where the device is installed and listen in the direction of the sound coming from the device. This means that the conversational agent (CA) in the physical device leads to visually and auditorily directed interaction. However, because the directed communication is based on where the device is installed, there are limitations in the agent's immediate interaction with the user, and it is difficult to notice the changing context of the user in real time at home [34].

Authors' addresses: Sunok Lee, sunoklee@kaist.ac.kr, Department of Industrial Design, KAIST, Republic of Korea, Daejeon, Republic of Korea, ; Minji Cho, mjcho@kaist.ac.kr, Department of Industrial Design, KAIST, Republic of Korea, Daejeon, Republic of Korea, ; Sangsu Lee, sangsu.lee@kaist.ac.kr, Department of Industrial Design, KAIST, Republic of Korea, Daejeon, Republic of Korea,

2474-9567/2020/9-ART88 $15.00
https://doi.org/10.1145/3411840

Thus, researchers have proposed that technology for smart homes should blend into the background without visual interference (e.g., invisible intelligence) for better user support [14, 53]. In addition, the development of ambient intelligence technology can be one direction for better interaction in ubiquitous smart homes [1, 13, 41]. The concept of invisible intelligence refers to the physical disappearance of computers from users' view [53]. This includes the miniaturization of devices and their integration into other everyday artifacts [48] or the home environment to act as intelligent agents that use prediction algorithms to predict behaviors [14]. Such devices can allow people to receive information without interacting with a physical or possibly obtrusive entity. Moreover, along with the development of invisible and ambient intelligence, recent studies have emphasized the possibility of ambient interaction, which would aim at interpreting what is going on in an environment, assisting when asked, or taking initiative when it could help users with nonintrusive technology [43]. This ambient interaction would differ from traditional human–computer interaction, which involves explicitly addressing a computer by interacting with a physical stand-alone device. Through the possibility of invisible intelligence, we assumed that an invisible agent could be one CA design direction for ambient interaction. However, an invisible CA has not been commercialized yet, and thus it is difficult to answer the following new questions with existing knowledge: How can people perceive and interact with invisible devices? What other advantages do invisible CAs have compared to CAs with physical entities? How can sensor-based, invisible interfaces be designed for implicit interaction with humans?

To find answers of these questions, it is critical to identify users' mental models and determine the design direction of the invisible CA. Similar to previous studies on CAs embedded in AI speakers, we emphasized the importance of understanding users' mental models for designing CAs [7, 11, 16, 28, 30, 39]. These studies attempted to understand users' mental models mainly through verbal approaches [11, 30]. However, because it is unclear how each user will accept the new concept of invisible agents, when forming a mental model toward the unsubstantial system by comparing it with current CAs in physical devices, limitations can hinder discovering users' expectations for a new concept with only a verbal approach. For this reason, it is necessary to explore users' mental models of invisible agents and to find out the difference between existing CAs in physical entities—such as AI speakers—for future design possibilities and development directions combining visual and verbal approaches.

Thus, we conducted a drawing study to understand the differences between the two CAs with different physical entities, as well as how their physical visibility affects users' mental models and interactions. To achieve that goal, 30 participants interacted with CAs with a physical presence and an invisible CA, and we then identified the users' mental models through their drawings, in addition to interviews and a survey. We explored how users' mental models and interactions are different in directed interaction with a visible agent than they are in ambient interaction with an invisible agent. The findings describe users' perceptions, expectations of the role of the agent, and interaction in terms of physical entities to discover the potential of the invisible agent. Based on this finding, physical presence is an important element in CA design, and we found features of invisible agents compared with CAs in AI speakers that were not revealed in previous studies. Finally, this study suggests that invisible agents can be one design direction in VUI to enhance user experience in the ubiquitous environment.

2 RELATED WORKS

As one of the primary benefits of a VUI is the ability to provide seamless interaction [12, 35, 36], CAs that use a VUI—such as AI speakers—have become popular in homes context to support various activities of users [11, 39, 45], among them finding recipes during cooking and controlling lights during reading [23]. A recent study on the context of AI speaker use also found that speakers were used most prevalently in homes. For the continued development of CAs to support better smart homes, many existing studies have attempted to explore not just users' experiences interacting with AI speakers in a ubiquitous home computing environment but also how to build, deploy, and interact with a ubiquitous smart home [2, 19, 22, 45]. Porcheron et al. [40] collected voice-recording data of actual Amazon Echo use in five participants' homes to understand how users incorporate the device into their everyday lives at home. Their study focused on the social relationship of human–Alexa interaction and identified collaborative activities of the user and the device in terms of taking turns talking and reactions of users to the device's responses. Park et al. [39] conducted a one-week participatory user study to discover multiple users' mental models toward the AI speaker as a key in the development of its family-oriented roles. The authors revealed seven domains of user expectations of CAs' roles that can be represented as family cohesion at home. Sciuto et al. [45] gathered the logs of 278,000 commands from 75 Alexa users to investigate their behaviors around Alexa, including the number and physical placement of devices and daily use patterns. Surprisingly, the use patterns showed that commands per day dropped after the first week, and eventually users' commands mainly involved playing musical requests. Furthermore, users experienced frustration due to limitations on fully hands-free and eyes-free interaction, even though VUI's primary advantage is its natural and unobtrusive interface [17]. Based on several studies that showed AI speakers were used for limited functions such as playing music [11, 45], we found that these devices cannot fully support a smart home with the strength of VUI. Among the limitations revealed from previous studies, we focused on two main issues: (a) limitations on the unobtrusive advantage of VUI, such as users' frustration at restrictions on fully hands-free and eyes-free interaction in a voice-controlled smart home and (b) the gap between users' mental models and actual use of CAs.

For unobtrusive interaction in a smart home, several studies have suggested that ambient intelligence with eye-free interaction is one possibility for resolving users' problems with limitations on seamless interactions. These studies proposed that ambient and invisible intelligence could respond to a user's seamless interaction and understand the user in any context [1, 5, 13, 14, 41, 48, 53]. Augusto and McCullagh [5] described the characteristics of systems with ambient intelligence through identifying scenarios of interaction with invisible intelligence for a better smart home and system flow. They suggested models supporting a network that deduces "contexts of interest" from the sensed environment and users. Their study highlighted an essential component of the invisible intelligence area as the distribution of technology that is intelligently orchestrated to allow an environment to benefit its users. Norbert and Nixon [48] investigated the possibility of disappearing computers through a study on invisible computing, with the hypothesis that the purpose of human–computer interaction research is the interaction not with the computer but with the information itself. They defined the term *invisible intelligence* as the physical disappearance of computers from users' view as well as the miniaturization of devices and their integration into other everyday artifacts. In addition, they suggested research agenda for interaction design issues when computers are disappeared from the scene (or in other words become invisible): Shouldn't the computer move into the background or out of view and understanding how people can interact with invisible devices. In addition, Weiser and Brown proposed [53] calm technology as a new approach to fitting technology to users' lives. They mentioned that embedding calm technology into the periphery could easily center user attention. This is because by deploying the technology within the users' periphery, they are provided with information without excessive burden and can control more things than the user can. Therefore, the authors suggested that calm technology is to put us at home, in a familiar place to solve the problem of information overload. In addition, they proposed properly considering the balance of the center and the periphery where calm technology is applied when designing a better ubiquitous computing environment. Cook et al. [14] designed the house environment itself to act as an intelligent agent without physical entity through presenting the MavHome smart home architecture, which allows a home to act as an intelligent agent that uses prediction algorithms to predict inhabitants' behaviors for the automated environment. They focused on the creation of an environment that acts as an intelligent agent, perceiving the state of the home through sensors and acting upon the environment through device controllers. The role of the MavHome agent involved lower-level agents responsible for subtasks within the home. Results indicate that the predictive accuracy was high even in the presence of many possible activities from synthetic and collected data at a smart home. In addition, they showed that prediction algorithms play critical roles in an adaptive and automated environment such as MavHome. These studies provided insights into the new possibility of better interaction and usability of CAs at home. In user evaluation studies on invisible systems, Wang et al. [52] focused on exploring how each physical entity affects users' perceptions and interactions. This study designed four agent appearances—voice-only, nonhuman (device shape), and two types of embodied agents—and then evaluated the overall user experience through interacting with the four agents by a given task. They found that voice-only agents helped users focus on the task without visual distractions. To investigate how visual entities and social behaviors influence the perception of virtual agents, Kim et al. [25] conducted a lab-based study using an interaction scenario with three types of agents (from one with no visual appearance to an animated 3D human). They found positive effects of visual embodiment on the users' sense of engagement, social richness, and social presence with the agent. However, these studies were conducted in a virtual environment through augmented and virtual reality systems and did not deeply explore the potential of invisible agents. In addition, Luria et al. [31] compared an embodied social robot with three different forms of interfaces (i.e., a voice-controlled speaker device, a wall-mounted touch screen, and a mobile application) for designing a smart home control interface. Participants were asked to simulate situations in which they controlled multiple devices in a smart home environment during cognitive tasks to simulate distractions in the home. Based on analyzing qualitative and quantitative results, they suggested that embodied social robots could provide an engaging interface with high situational awareness, and the voice interface had the advantage of hands-free and ubiquitous control. Moreover, they raised questions about non-visibility of VUI on low situation awareness, users' discomfort, and users being less familiar with voice interface compared with social robots. However, Luria et al. compared the user experience with other interfaces centered on social robots, but did not investigate invisible intelligence without physical entities. Based on the insights from previous studies, our study involved user experience of an invisible agent in a real environment and to reveal the agent's features in comparison with visible agents.

In terms of the gap between users' mental models and actual use of CAs, several studies have revealed that AI speakers are reduced to music players because users' mental models, including their expectations, are not aligned with actual user experiences; therefore, users feel the AI speakers have limitations and become disappointed [6, 7, 16, 30]. By exploring associated obstacles or difficulties, Cho et al. [11] analyzed how people use Amazon Echo in their daily lives and conducted a long-term user study in which they followed eight households. A series of diaries, surveys, and interviews with first-time users identified obstacles that caused users to forget the presence of the CA at home. The findings revealed that users expected Alexa to be a partner with whom to have human-to-human conversations and build a long-term relationship, but the actual capability of Alexa did not match users' mental models, leading them to disappointment. Beneteau et al. [6] conducted an in-the-wild study in which 10 families adopted an Amazon Echo Dot. Through audio recordings of 10 families' interactions with the devices and interview data, they found that device capabilities were not always in-line with family expectations of learning and use. Luger and Sellen [30] conducted a comprehensive study to understand what factors affect everyday use of dialogue-only CA systems. Their study showed that users had mismatched mental models of how their CAs worked, which were reinforced through a lack of meaningful feedback regarding system capability and intelligence; that is, a deep gulf between users' expectations and actual CAs was the main problem with CA use. The authors also identified that anthropomorphism contributed to the users' high expectations. As a result, they suggested ways to reveal system intelligence, including the interactional promise of humorous engagement, indication of capability through interaction, and redesigned system feedback. They explored user mental models for an agent embedded in a smartphone or AI speaker, yet no research has revealed a mental model for an invisible agent. Because of the low discoverability of the agent interacting through the VUI [15] can cause the user to have higher expectations. If the hardware is not visible, the visual cue disappears. There is a concern that the gap between mental models of invisible agents and actual use will be wider than the gap with existing agents. Thus, we intended to reveal the direction of CAs, which can provide better usability for users in smart homes through identifying mental models for visible and invisible agents that previous studies have not revealed.

Since mental models can include abstract images that are difficult to express in words, they are difficult to understand fully by a verbal approach such as an interview. Mental models for ambiguous and inexplicit systems such as invisible agents have an even greater likelihood of being hard to articulate. Therefore, we referred to studies that explored the mental models of users with an integrated visual and verbal approach. Lee et al. [28] asked participants with previous experience interacting with Amazon Alexa or Google Home devices to draw what they thought a CA looked like. The participants' drawings were categorized into four persona groups: human, speaker, system, and space object. These findings confirmed in a pictorial way the users' unique and diverse mental models and expectations [20, 21, 28]. Through drawing tasks and interviews, Xu and Warschauer studied the perceptions of three- to six-year-olds toward a speech-only "Google Home Mini" CA, examining domain membership, property attributions, and children's justifications of property attributions. Their findings on the children's mental models showed they categorized drawings of CAs as artifacts or humans, which was consistent with the traditional A-I distinction (distinction between living and nonliving things) proposed in developmental psychology research. Based on the children's interviews and drawings, the authors provided design implications to guide the development of CAs to support young children's cognitive and social development by considering the communication techniques that an agent should have as well as the appropriate contents of conversation between an agent and children [55]. However, these studies only focused on visible CAs, which were stand-alone AI speakers; they did not grasp users' mental models of invisible intelligence. Silver et al.[46] conducted a qualitative case study on children's perceptions of urban landscapes to create digital maps in the context of locative systems and wayfinding for children. Seventy children drew cognitive maps of their journeys from home to school. drawings, they provided the children's own accounts of their maps and generated a set of ten themes related to landmarks and design ideas for the creation of digital maps for children. Sciuto et al. [45] did not use drawings as the main basis for their study; they used a sketching application for a generative study inspired by a participatory design. The purpose of this activity was to evoke users' experiences before interviewing them. Participants, after sketching a concept of the Alexa ecosystem, created an image of Alexa using a police sketch application. Although the images themselves are not shown in the paper, users were able to express their perceptions of Alexa as explicit images. Kuzminykh et al. [26] conducted a qualitative multi-phase study seeking to identify patterns in users' anthropomorphized perceptions of three types of CAs (Alexa, Google Assistant, and Siri). The CA users created an approximate visual representation of Alexa, Google Assistant, and Siri using an open-source web avatar generator. Through analysis of visualization data and interviews to identify patterns in the purposefully anthropomorphized perceptions of CA technology, the consistent differences revealed through a comparative analysis of the three agents suggest the importance of this research direction for the design of CAs. However, these studies had limitations that made it impossible for users to visualize their mental model freely because they relied on applications that can only visualize humans. With reference to the studies, we found it necessary to understand users' mental models of an invisible agent to ascertain in which direction an invisible agent should be designed. Furthermore, we assumed that a drawing method would be suitable for exploring users' mental models for systems that have not yet appeared, such as an invisible agent. Therefore, we intended to discover the mental models users have of two different visible and invisible agents through a drawing approach and to explore how and why these models differ.

3 METHOD

3.1 Participants

To grasp users' first perceptions of visible and invisible CAs, we recruited 30 novice users (avg. age = 22.5 years old, SD = 3.32; 13 females and 17 males) who had never used an AI speaker. Half were to talk to a visible CA and the other half were to talk to an invisible CA. We posted in online communities, such as a university one, to recruit people who wanted to use an AI speaker for their homes. We tried to recruit relatively young participants, even if awareness of users of different ages is important, because 34% of U.S. 18–29-year-olds owned a smart speaker at the beginning of 2019 [51]. This is significantly higher compared to other age groups. Therefore, we thought that this age group would use this type of device in the future. We also focused on first-time users because the initial stage of CA use would affect users' mental models and later stages of use [12, 28, 36], and users in the initial stages of voice interaction need the most assistance to be led to proper use [11]. To compare clearly the new concept of invisible CA with that of the existing visible CA devices, users experienced with neither were more appropriate. Moreover, according to previous research, the first image of an agent rarely changes, being imprinted in users' minds as soon as they hear the agent's voice [36]. Therefore, we would be able to grasp the unbiased perceptions that users formed by the experimental setup we provided. All of our participants were Korean, and we conducted the study in Korean with native Korean speakers. Each participant was compensated with approximately U.S. $20.

3.2 Apparatus

Because the goal of this study was to compare natural user experiences with visible and invisible CAs, we used three commercialized AI speakers and invisible agents in the study environment. Figure 1 shows the apparatus that our study used. A CA without a physical entity—that is, an invisible agent—has not yet been commercialized, so we hid a stand-alone AI speaker in the study environment to make it invisible. Because the goal of the invisible agent is to assist users without visually distracting them from a task, we adjusted the volume and position during the setup process so that the participants did not know where the sound was coming from, causing them to feel as though the invisible agent was ambient. The participants accordingly perceived the sound as coming from the ceiling, rather than from where the agent was actually installed. Three AI speakers were installed in the lab environment in the case of visible devices, and three devices were hidden in the case of invisible agents.

In terms of AI speaker types, we used AI speakers from Naver Clova [38] which are some of the most advanced

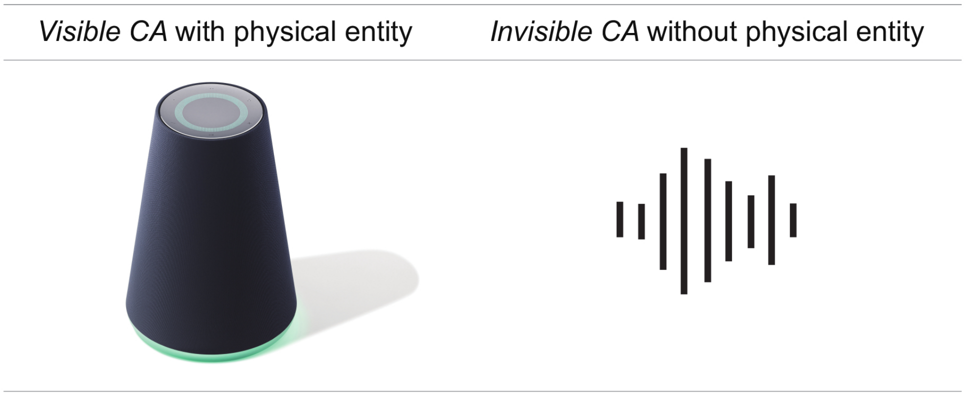

Visible CA with physical entity
Invisible CA without physical entity

Fig. 1. Apparatus for study: visible CA (left) invisible CA (right)

agents in South Korea, provide many functions, and connect to various IoT devices and apps to control a house. The Clova as a stand-alone AI speaker is typically shaped like a cylinder and includes a visual cue (Figure 1). Even if the Naver Clova had released an embodied device in which an agent is represented, such as JIBO [44], would have excluded the embodied AI speaker from the experiment to focus on the physical entity itself and ensure that participants would not be affected by appearance and embodiment level. The fact that Clova does not offer much functionality compared to global mainstream devices could have a slight effect on overall user interaction. The Clova agent launched by Naver, the main search engine in Korea, provides information that fits the Korean context and was developed based on the Korean language, so it was more appropriate for Korean-speaking participants compared to other devices optimized for English.

3.3 Study Environment

Figure 2 shows the study environment, which includes the location of each device and interaction point. We did not specify the interaction point clearly for the participants but had them interact at whichever point was comfortable in each space. The environment reflected a smart home in which people have multiple speakers dispersed to increase the coverage of a voice agent. Referring to prior research on common locations for Alexa devices [45], we configured the lab environment into three spaces: a kitchen, a living room, and a home office. To provide the sense of a comfortable home and typical environment for AI speaker usage, we placed a sofa and table in the private room, a table and TV in the living room, and a desk and computer in the home office. The participants were asked to interact as naturally as possible as though they were at home, although the experiment was conducted in a lab environment that assumed a home environment. In the case of CAs with physical entities, three AI speakers were installed in each space (Figure 2 left). Participants in the visible agent group sat down and interacted with an agent by facing it. In the case of invisible agents, three devices were hidden in the same places (Figure 2 right) so that the agents could answer and participants could make requests from anywhere. Participants in the invisible agent group interacted with the agents in three types of spaces ubiquitously.

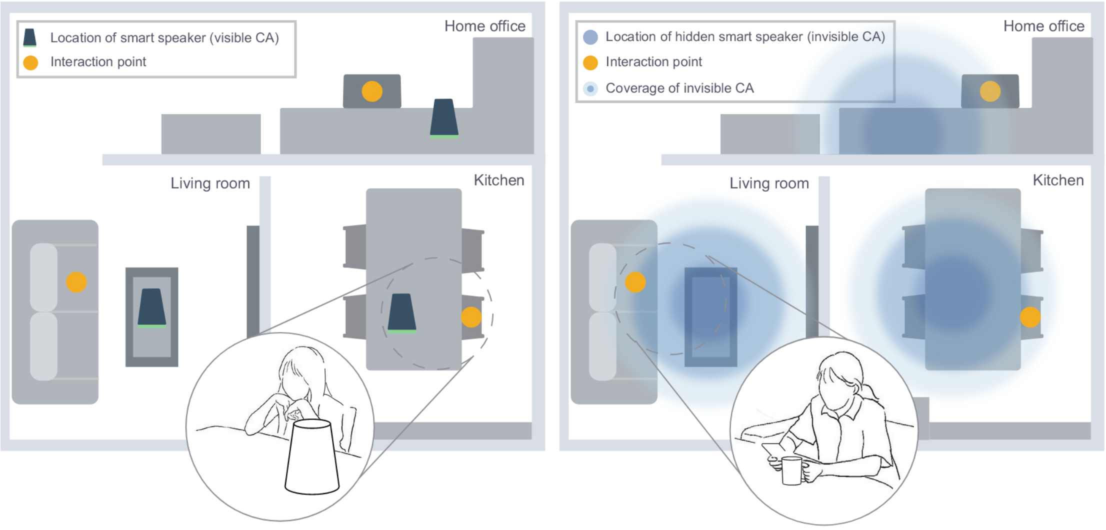


Fig. 2. Study environment of the visible CA (left). Study environment of the invisible CA (right).

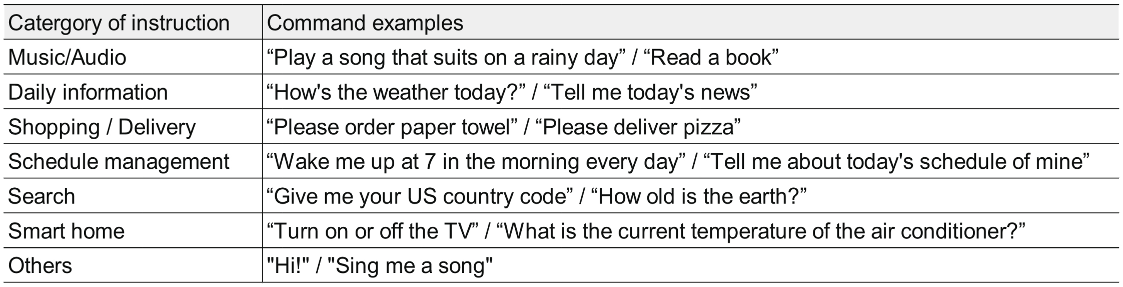
3.4 Procedure

3.4.1 Interaction Task. The study consisted of four sessions: a 5-minute tutorial, a 30-min interaction, a 40-minute drawing, and a debriefing interview. In the tutorial session, we encouraged participants to talk freely with the CA. The tutorial was intended to help first-time users become familiar with a VUI. After the tutorial, we asked participants to interact with CAs in the three spaces in whatever order and to whatever extent they wanted. For a better understanding of interaction, we gave participants printed instructions in seven categories with 65 command examples (Table 1). In addition, we asked them to interact with the CAs as if they were at home.

3.4.2 Drawing Task. After interacting in the three spaces until participants formed an image of the agents, the participants came back to the kitchen (Figure 2) and were given a drawing task. We asked them to draw their perception of the agent, instructing them to "draw a CA or CAs in this space." The purpose was to see whether the participants considered the agent singular or plural. We additionally said, "You can also draw a relationship between the CA and you," because such a picture could allow us to determine the relationship between agents and users. For drawing material, participants used unlined A4 size paper, which is common in many types of drawing studies. Because a comfortable environment and familiar drawing materials are important factors for relieving pressure when the users draw [9, 28, 37], we let them draw with various items they felt more comfortable with (e.g., pens or colored pencils). Moreover, participants were asked to focus on visually expressing their perception freely rather than worrying about drawing skills.

3.4.3 Debriefing Interview and Evaluation. After the drawing task, we conducted debriefing interviews to understand participants' mental models and interactions through observations based on the drawings. Regarding the perception of the CA, we asked users to describe their drawings with such prompts as "Explain your drawing" and "Why did you draw these elements?" For mental model and interaction, we asked, "What did you expect about the CA that you depicted?" "Why did you act in the way you did when you talked to the agent in the living room?" and "Are there any good or bad things while talking to an agent?" After the interview, a survey was conducted on the overall user experience of the CA. Participants then rated their user experience (presence, trust, preference, continued use, satisfaction, and intimacy) of each CA in the expanded space on a 7-point Likert scale. Data generated from the evaluation were intended to supplement findings from the drawings and interview data. We used the scale to examine the differences between each physical entity with regard to the overall user experience. The presence scale is used to obtain a sense of the relationship between the agent and the user [52]. Trust and preference motivate maintenance of a relationship [8, 24, 52], and continued use and satisfaction indicate general user satisfaction [8, 52]. Intimacy refers to the emotions that explain the user's retention [11, 29].

Table 1. Printed instruction in seven categories



Catergory of instruction	Command examples
Music/Audio	"Play a song that suits on a rainy day" / "Read a book"
Daily information	"How's the weather today?" / "Tell me today's news"
Shopping / Delivery	"Please order paper towel" / "Please deliver pizza"
Schedule management	"Wake me up at 7 in the morning every day" / "Tell me about today's schedule of mine"
Search	"Give me your US country code" / "How old is the earth?"
Smart home	"Turn on or off the TV" / "What is the current temperature of the air conditioner?"
Others	"Hi!" / "Sing me a song"

3.5 Data Analysis

Based on the presence or absence of a physical entity, we transcribed a total of 22.7 hours of recorded interviews and analyzed 30 drawings along with observation records to organize the data into emerging themes. For the drawings, data were coded in relation to how the physical entity was exhibited in the drawings. The main priority of our analysis was to examine each detail or element of the drawings to understand the participants' perceptions. For the interviews, we conducted open coding, where we identified and coded concepts that were significant in the data from the mental model and interaction in terms of the physical entity through the open coding process. We created 467 codes that we grouped into categories using axial coding. We then discussed and refined the codes to reach agreement ($K > .84$), and we excluded data instances without agreement. The interview data from the findings supplemented the elements of the drawings and allowed us to determine each feature of visible and invisible CAs. In addition to quantitative analysis, we ran an independent t-test to examine the effect of the physical entity on the overall user experience. There is no significant difference was found between visible and invisible agents regarding overall user experience. Thus, quantitative data were only used to supplement the findings regarding the participants' interactions.

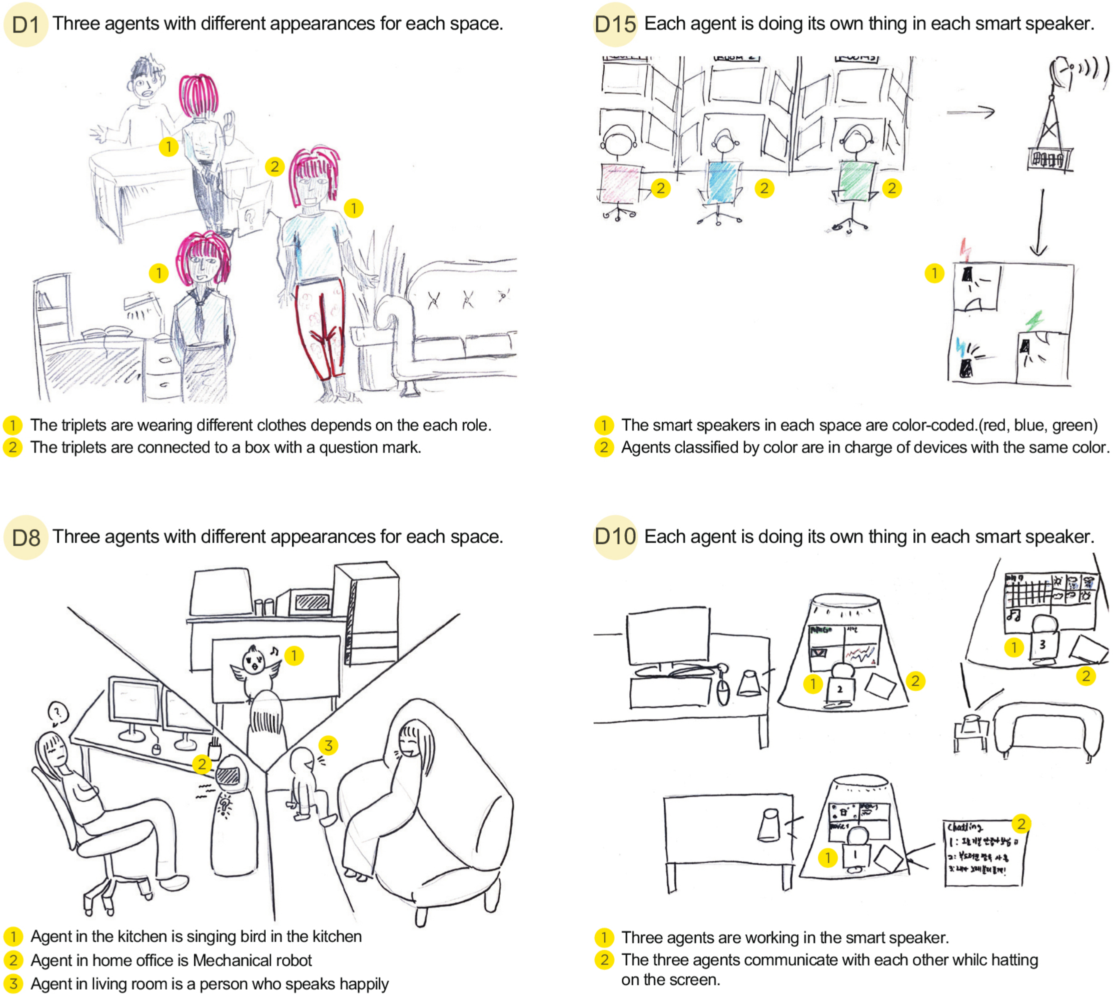
4 FINDINGS

From the qualitative analysis, which mainly focused on comparing each agent, we found that participants perceived CAs with physical entities and invisible CAs differently. In particular, participants' drawings showed how the different physical entities of the CAs affected their perceptions (Figure 3 and 5). The data from observations and interviews allowed us to grasp the mental models and interactions linked to the users' perceptions. In addition, for the survey data, we analyzed the average of the overall CA experience trends for the two physical entities (Table 2). Through findings related to perception, the mental model, and interaction, we can discover the unexplored features of invisible CAs compared with CAs in the stand-alone device. In next section, we refer to each drawing as D and the participant who drew it as P. Both have the same number.

4.1 Perception

4.1.1 Perception toward Visible CA. After three interactions with visible CAs, the majority of the participants depicted three agents in the drawings. In particular, the drawings of the 11 participants who interacted with an visible CA showed the participants expressed the three identical AI speakers as three different agents with different appearances and roles that contributed to the overall persona of each CA (Figures 3 and 4). In other words, the participants perceived that interacting with the three AI speakers meant talking to three individual agents, because even though the devices had the same form, wake-up words, and voice in the experimental setting, the users perceived each agent differently. In D19 and D30, in which three AI speakers were represented as three different agents, three agents were drawn with different colors or appearances for each space. In addition, these drawing illustrated that each agent played a different role depending on the space in which the interaction happened. We thought that each device would make the participants feel that the agent was an independent physical conversation counterpart.

For example, in D8 (Figure 3), the agents were expressed as a pet, a robot, and a person in each space in which the AI speaker was installed. In addition, this drawing revealed that the agents were given different roles depending on the purpose of the space. In the drawing, a cheerful-looking bird is in the living room, and an anthropomorphic agent and the user talk pleasantly in a room with a sofa. In contrast, in an office space with a desk, a mechanical robot is communicating with the user. In another case, D1 (Figure 4), the agent had the same face and hair color in all of the spaces but was expressed as wearing different outfits. Furthermore, this participant (P1) explained that these three "people" were triplets. We found that the identical exterior design


D1 Three agents with different appearances for each space.

1 The triplets are wearing different clothes depends on the each role.
2 The triplets are connected to a box with a question mark.

D15 Each agent is doing its own thing in each smart speaker.

1 The smart speakers in each space are color-coded.(red, blue, green)
2 Agents classified by color are in charge of devices with the same color.

D8 Three agents with different appearances for each space.

1 Agent in the kitchen is singing bird in the kitchen
2 Agent in home office is Mechanical robot
3 Agent in living room is a person who speaks happily

D10 Each agent is doing its own thing in each smart speaker.

1 Three agents are working in the smart speaker.
2 The three agents communicate with each other whilc hatting on the screen.

Fig. 3. Drawings representing three agents, each playing a different role

of the AI speakers was reflected as three people with the same appearance in the picture; these people wore different clothes and performed different roles. In addition, in D15 (Figure 4), the agents of the same appearance were represented by various colors, such as red, green, and blue, respectively. These participants explained that they drew three different agents because three physical devices were visible, so he matched the number of agents with the number of devices.

> *Although there are three devices of the same shape, each one seems to be different depending on where they are placed individually.* (P8)

> *The devices were separate, so I perceived them as different beings.* (P10)

4.1.2 Perception toward Invisible CA. Contrary to the visible CAs, the participants expressed the agents as a single agent after three times of interactions with the invisible CA. Twelve of the 15 drawings representing the invisible CA were drawn as one agent controlling the entire space. Because the voice of the agent was the same everywhere and the device was not visible in any of the spaces, the participants focused more on the consistency of the voice regardless of the number and location of the interactions. In the drawings, each participant perceived the agent's persona differently, such as a person, a computer server at a headquarters, and a brain, but all participants perceived the agent as one consistent agent throughout the space. All 15 drawings represented the CA as looking down on the entire space from above.

For example, P16 visualized the agent as a server at the headquarters of the company "Naver" that produced the AI speaker and displayed a structure in which information was transmitted to the user through the speaker from where the server was located (D16 in Figure 5). Another participant who said that the space itself was an agent drew the CA as a brain-shaped central processing unit surrounding the whole space and expressed the control range of the agent for each space with a speech bubble (D23 in Figure 5). This participant thought that the available range of the interaction with the agent was like a big transparent balloon. We thought that because the device was invisible, the participant perceived the space itself as an available range of communication, unlike a perception of the visible device. In addition, P21, who portrayed the agent as a woman with large ears, thought that the agent beyond the ceiling delivered an answer through the speaker to each space where the participant's voice input was. In addition, the ears were large because the agent listened carefully to the user's voice anywhere. Accordingly, participants who visualized one agent controlling the entire space perceived that they could interact with the invisible agent anywhere without considering the location of the device.

> *I feel like there is a single agent connected to the server in the sky. Actually, rather than having three agents, I feel like one agent can go anywhere.* (P17)
>
> *I think this is one person. Because it doesn't change the tone, it doesn't change the voice, and the personality is not different (....) The reason I drew the big ears is because I feel like she is listening to me very carefully with huge ears.* (P21)

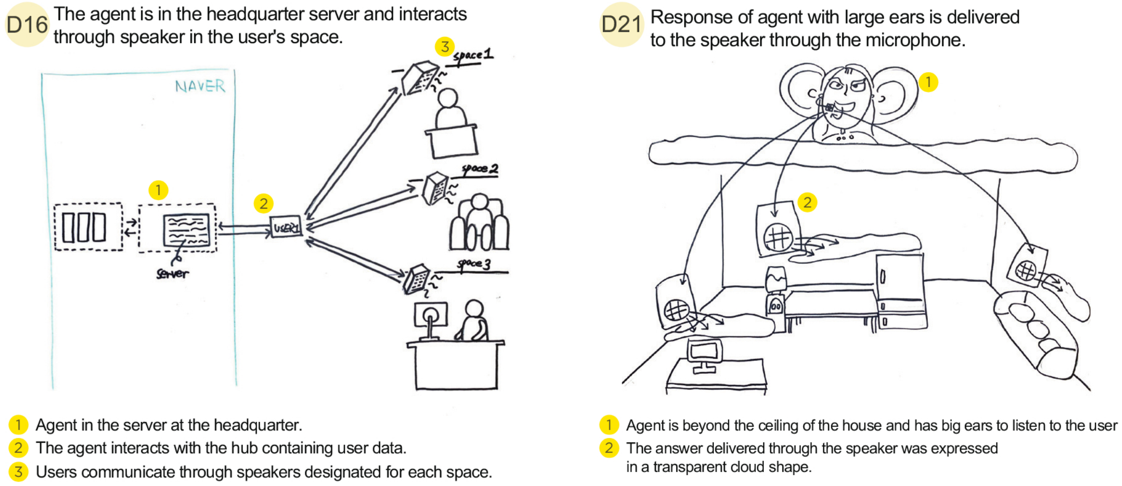
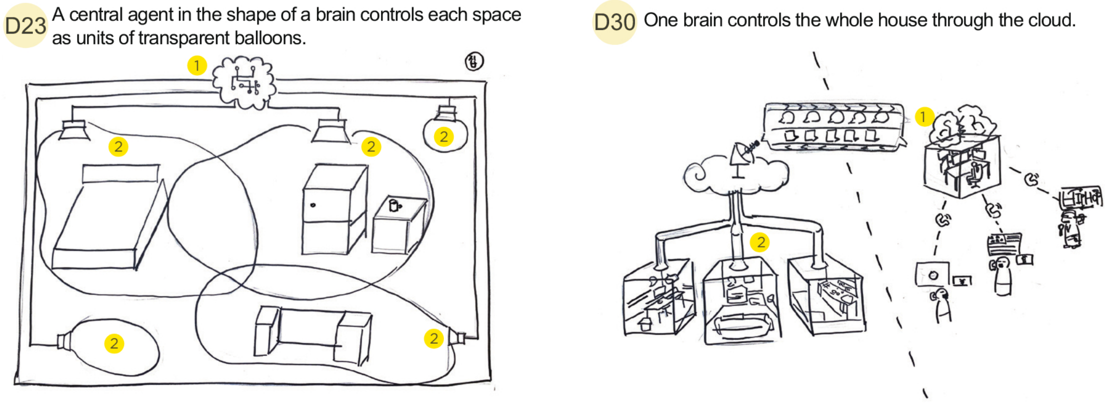

D16
The agent is in the headquarter server and interacts through speaker in the user's space.
NAVER
1
2
3
server
user
space1
space2
space3
1 Agent in the server at the headquarter.
2 The agent interacts with the hub containing user data.
3 Users communicate through speakers designated for each space.
D21
Response of agent with large ears is delivered to the speaker through the microphone.
1
2
1 Agent is beyond the ceiling of the house and has big ears to listen to the user
2 The answer delivered through the speaker was expressed in a transparent cloud shape.

Fig. 4. Drawings of the participants' perception of an invisible agent



D23 A central agent in the shape of a brain controls each space as units of transparent balloons.

1 One agent was expressed as a house itself.
2 The control range for each space is divided into transparent balloon sections.

D30 One brain controls the whole house through the cloud.

1 The agent is in a brain-shaped box beyond the cloud.
2 Information coming from the cloud moves through a pipe connected to each place.

Fig. 5. Drawings of an invisible agents where one agent controls the entire space

4.2 Mental Model

As described in the perception of the CA for each visible entity above, different perceptions naturally led to different mental models of the CAs. Through the participants' interviews and explanations of their drawings, we found a relationship between the perception of each agent and the mental models of the different expectations regarding the visible and invisible agents' roles.

4.2.1 Mental Model on Visible CA. Participants who communicated with the visible agents perceived three agents that each play a role with own expertise depending on the space. The perceptions of the three visible agents naturally influenced this expectation. Even if all AI speakers had the same functions, wake-up words (which were the names of the agents), and voices in the study setup, the participants actually expected the agents to have their own expertise in each role. For example, D8 illustrated the agents as strict but smart agents in the home office and as more active and bright agents in the living room. Then, some of the participants wanted three agents with individual personas for different roles. This expectation demonstrated that the users' perceptions of individual agents for each device were naturally connected to different personas, roles, or functionalities. Some participants even said there was no reason to install all three identical devices at home and communicate with multiple agents if they were all the same.

> *If there are several [agent], it would be better to organize agents' roles using each specialty rather than one to cover everything. (P6)*
>
> *If each has a different role, there may be a reason to use multiple AI speakers.* (P11)

On the other hand, some participants had negative impressions of interacting with multiple agents. These participants said that if the experimental environment were a real home, a mental load would be required to talk to several agents in their private spaces. For these participants, the interaction was similar to communicating with people, where more effort is required to get to know several people than it is to know one person. In addition, the fact that multiple agents archived users' conversation histories was similar to many people knowing one person's secrets. Furthermore, some participants were concerned about interactions between agents in relation to their privacy, which caused negative perceptions that affected their trust in the agents. P8 thought that if the agent shared a secret and communicated it to the other agents and not him, then he would distrust the agent and feel alienated from it.

> *I feel like I have to try to get to know this [agent] through conversation, but if there are several of them, it is a huge burden that I have to try to get to know three or four people.* (P8)
>
> *Even if different agents communicate with each other through something like a hub, it's likely to be creepy or less reliable because I'm afraid that these agents will communicate my private information that I don't want to share with other people.* (P12)

4.2.2 Mental Model on Invisible CA. Participants who interacted with the invisible agent expected that one agent controlling the entire space would interact with the same persona in a consistent tone and voice no matter where the participants were communicating. The perceptions of a single invisible agent naturally influenced this expectation. In addition, because no physical device spatially limited the agent, the participants thought the agent could detect a user's behavior and status without space limitations everywhere in the house. Therefore, they wanted the agent to provide the most appropriate information based on its awareness of the user's context. Most of the participants expressed high satisfaction with the invisible agent after remembering the previous conversations and responding to their requests based on the content. They even took personalized responses from agents for granted. This was contrary to the mental model for the visible agent, where participants were concerned that their conversation records were shared with agents in other spaces.

> *On the sofa, when I said to the agent, "I was having a hard time today," and then when I moved and sat at the desk and asked it to play music, the agent said, "I'll play music that feels good." That is so touching even though it was a brief moment, I felt like I was emotionally interacting with this agent.* (P28)

However, some participants (P17, P19, and P20) shared their worries about one invisible agent that knew all of their information and stated they expected high privacy protection. Their reason was that if the agent, who was always watching the user, shared private information with another system next door or a corporate server, then it would be difficult for them to trust the agent due to their anxiety about the exposure of their personal information. Therefore, in terms of information sharing, their concerns were similar to those associated with visible agents, but invisible agents could spark even greater user concerns compared with visible CAs. This is because the participants imagined that invisible agents might know about and monitor all of the users' every single movements. In addition, because the device was invisible, the participants did not know whether the agent was looking at them. Therefore, the participants expected the agent to keep all of their information secure. They thought that the home interacting with the agent was a private space, so the agent should be personalized for the participant. Through this, we found that invisible agents could be more appropriate to use as personalized agents that respond based on a user's information in the home context rather than in a public space. In addition, invisible agents may have more opportunities to use personal information than visible agents do from the expectations of the participants. However, in terms of privacy protection, an invisible agent must build trust with a user to align with the user's mental model.

> *This [agent] knows all of the things I do, but I don't want my information to be shared with the next-door neighbor or headquarters like Google. Then, I will never believe this agent and will never use it.* (P20)

4.3 Interaction

To investigate if there were any differences in the interaction according to the CAs' physical presence, we analyzed the observation data of participants' interaction. We found that the participants interacted differently in terms of the agents' visibility. Specifically, participants who interacted with the visible agent engaged in directed interaction based on where the device was installed, and participants who interacted with the invisible agent engaged in ambient interaction without space constraints compared with the visible agent. Table 2 lists the six evaluation criteria for the user experience for each agent. The independent t-tests revealed no significant effect of the agent on different physical entities. Thus, because the physical entity of an agent did not have much of an effect on the overall CA experiences, the survey results were used to supplement the observations and interview data.

4.3.1 Directed Interaction with Visible Agent. In the observations and interviews, we found that if a physical device was in the space, the participants tended to stare at the device or to interact with the agent with their bodies directed toward the device. In particular, when participants sat down at the interaction point, they naturally sat facing the AI speaker, and when starting conversations, they looked at the visual cue of the AI speaker and grasped the direction of the conversation. Even if the participants did not anthropomorphize the agent in the drawings, they treated the physical device as a conversation partner that could see and talk. Therefore, we found that the physical entity caused directed interaction and made people feel comfortable with talking. We also assumed that the reason people were constantly looking for devices to keep an eye on is because they were used to seeing people during conversations.

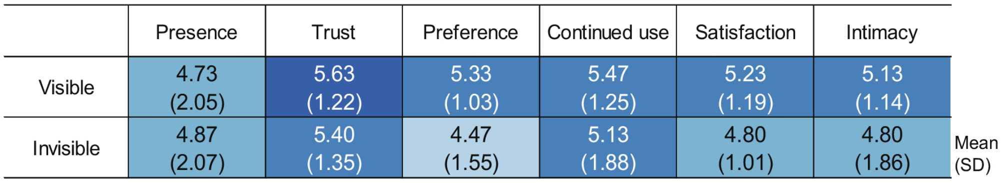
> *I felt comfortable because I had someone to talk to.* (P9)
>
> *This form makes me feel like I'm having a conversation with someone. If this disappears, it would be awkward because I would be talking to nothing, which would definitely not be a natural conversation.* (P10)

Furthermore, the visual cue that represented the agent's wake-up words and caused it to listen to the user—one of the factors that caused the participants to look at the device—increased the learnability of turn-taking with agents and the discoverability of the VUI, especially for the first-time users. Because the VUI is mainly auditory, less visual affordance makes it especially difficult to understand how early-stage users interact with agents. In addition, visual cues can give users the confidence needed to communicate with agents about uncertain input methods, and the physical entity can make the user feel that the conversation is real because the device responds to his or her voice. As we can see from the survey results, participants rated the overall user experience slightly higher. Through the participants' survey and interview data, we can see that the visible agent with visually directed interaction is familiar to users. As a result, directed interaction with the invisible agent not only makes users feel that the agent is a conversation partner but also offers the advantage of providing initial usability.

> *When the light comes on, I can see that the agent is listening to me, and if there is no visual feedback, I wouldn't be really sure.* (P8)

4.3.2 Ambient Interaction with Invisible Agents. In the observations of interactions with the invisible agent, we found that some participants looked at their mobile phones for personal texting and walked around with them, talking in a place not specified by the experimenter as they talked to the invisible agent. In other words, when interacting with an invisible agent, participants were not conscious of agents without physical entities. They felt that they could talk to the agent and listen to the answer anywhere, so that nondirected interactions are

Table 2. Mean comparison of 6 user experience criteria depends on agent's visibility



	Presence	Trust	Preference	Continued use	Satisfaction	Intimacy
Visible	4.73 (2.05)	5.63 (1.22)	5.33 (1.03)	5.47 (1.25)	5.23 (1.19)	5.13 (1.14)
Invisible	4.87 (2.07)	5.40 (1.35)	4.47 (1.55)	5.13 (1.88)	4.80 (1.01)	4.80 (1.86)

Mean
(SD)

possible while looking at mobile phones or other places. Thus, those who interacted with the invisible agent did so in a way that is opposite of those who interacted with the visible agent, which made them perceive agents as the conversation partners. This interaction was naturally connected to the expectation and mental model of the invisible agent in previous findings. These participants explained the benefits of a conversation with an invisible agent compared with Siri embedded in a mobile phone. They said that unlike with Siri, interacting with the invisible agent was more comfortable due to the lack of visual interference. For example, when they asked Siri something on their phones while they were doing other things, they needed to constantly look at the screen and check the status of the device while doing their main tasks. Through this, we assumed that the invisible agent has the advantage of enabling parallel task execution, because in the case of an agent without a device, it is not necessary to look at it, so subtasks can be carried out while the user continues to do his or her previous or main work.

> *When I can see it, I have to consciously stare and talk, and keep checking to see if the agent is listening to me or not, and I guess this will interfere with usability. But, if I don't need to see the agent, then I can do two things at the same time.* (P17)

On the other hand, conversations with invisible agents were unfamiliar interactions and caused cognitive discrepancy. This is because although the participants interacted with the agent using their voices, they could not see its physical entity, and the source of the sound was unknown. This makes it difficult for people to feel that the agent was the subject of the conversation, which may be because the invisible agent has not been commercialized yet, so participants may be unfamiliar with it or the cognitive discrepancies that come from talking to an invisible entity. Furthermore, contrary to the benefits of a visible agent, an invisible agent has no visual cue, so participants struggled to learn the turn-taking timing at first. Compared with the case of the visible CA, when the participants called Siri on their smartphones, or when Siri listened to them, the status of Siri was displayed on the screen, so they knew the timing for talking. However, in the case of an invisible agent, because no device shows the agent's status, it is difficult to understand the timing of a conversation. Therefore, we found that low learnability due to the lack of a visual cue would be a barrier to providing an initial positive experience for first-time users of an invisible agent.

> *I felt uncomfortable and a sense of cognitive discrepancy with talking to an invisible agent because there's a sound, and something is coming from somewhere, but there is nothing—just empty space.* (P21)

> *This is uncomfortable because Siri lights up but this one does not. I have no idea how loud I should speak and where to speak. If this is visible, I can know how I should talk to the agent, but I don't know where it is... (Is it under there or up there?) I don't know how loud I should speak.* (P16)

However, as we can see from the survey results, the overall experience with the invisible agent was slightly lower on average but did not seem to differ significantly from that of the visible agent. In addition, the participants (P19 and P21) who felt cognitive discrepancy toward the invisible agent explained that through repetitive interactions, such cognitive discrepancy gradually disappeared. Then, we assumed that this cognitive discrepancy and low learnability could be resolved through the interaction design approach. Therefore, the cognitive discrepancy is an important issue that needs to be considered to provide positive experiences with invisible agents in the future.

5 DISCUSSION

In light of our findings, the participants' mental models toward the agents and interactions differed depending on the presence or absence of the physical entity of the CA. This is because visible agents pull the participants' visual and auditory senses toward the physical device, which leads to directed interaction. On the other hand, an invisible agent without visibility leads to ambient interaction, which makes people talk to the CA and listen to its response from anywhere. Through this study, the physical entity's visibility influenced the mental model of the user, and the physical entity played an important role in the user experience. By revealing the relationships among the perception, mental model, and interaction depending on the presence or absence of visible physical devices, we suggest future design directions that support better voice-controlled smart homes.

5.1 Directed Interaction with a Visible Agent vs. Ambient Interaction with an Invisible Agent

5.1.1 Directed Interaction with a Visible Agent. In directed interaction with the visible agent, which had a physical presence, participants perceived the number of agents as being the same as the number of devices and expected each agent to have its unique expertise, rather than equal ability. The users' perception of equal numbers of devices and agents aligns with the concept of "one-for-one," a human-inspired model in which each social presence has a single body, as identified in previous studies [32]. According to our research, when participants interact with the same agents (Clova) embedded in multiple AI speakers with the same shape, they perceived each AI speaker as an independent entity because the physical devices were separated: "Although there are three devices of the same shape, each one seems to be different depending on where they are placed individually" (P8). However, previous studies have shown that users can be confused when interacting with multiple agents if the definition of the relationship is unclear. For example, different roles and expertise can be assigned to agents for each device to reduce user confusion [10]. Therefore, when designing a ubiquitous environment built with multiple devices like IoT, designers should define a clear relationship between the user and the agent when trying to embed different roles of multiple agents. For example, there should be no confusion when users request a specific function among multiple agents through users' awareness of each role of the multi-agent at home or the agent's clear communication of its role.

In addition, our results showed that several participants had negative perceptions of the existence of multiple agents. The participants worried that multiple agents could share the users' private information among each other without the users' knowledge or permission: "Even if different agents communicate with each other through something like a hub, it's likely to be creepy or less reliable because I'm afraid that these agents will communicate my private information that I don't want to share with other people" (P12). This is similar to the findings revealed by Luria et al. [32]: agents interacting with each other without user input evoked negative feelings in users, such as exclusion and social isolation. Therefore, designers need to be cautious when trying to embed different roles of multiple agents into a smart home environment to prevent this negative user experience.

In terms of having a physical counterpart to see and communicate with, participants interacted easily with the visible agent by considering that interaction to have the same characteristics as human–human communication: "I felt comfortable because I had someone to talk to" (P9). This is because people are naturally used to looking at the person they are talking to when communicating. In addition, quantitative data showed that users evaluated the user experience more positively for familiar interactions with visible agents than they did for invisible agents, which means directed interaction with an agent inside a physical entity is similar to communicating with people face-to-face. Through these findings, the presence of a physical entity will help users who are unfamiliar with or find it difficult to adapt to new technologies or who prefer more human-like communication. For example, in the science fiction drama *Years and Years* [50], which speculates about the future based on possible technology, when an AI speaker is converted into an invisible agent in a house due to technological development, the grandmother—one of the characters—says she wants the physical AI speaker back to stare at and talk to it. Therefore, as previous studies have shown [31, 52], a physical entity such as embodied robot can provide familiar interactions to users. Then, it will be easy to have social interaction resembling human communication, and the learnability of VUI can be increased. Thus, physical entities with diverse forms can be used appropriately depending on the users and context.

Additionally, our study showed that visual cues provoked visual attention, which allowed users to know how to communicate with the agents. This advantage of learnability was particularly useful for first-time users. These properties of a visible agent can support learnability for an initial user who needs to understand the agent's turn-taking or for people who are not familiar with interactions with an agent, such as older people who are more familiar with physical counterparts seen in human communication. Therefore, designers can use the physical entity of visible agents appropriately as a design element according to the users' characteristics.

5.1.2 Ambient Interaction with Invisible Agent. In ambient interaction with invisible agents without physical presence, the participants freely interacted with the agent without considering the device's location, such as by starting a conversation with the agent while moving to another space. This nondirectional interaction unconstrained the users' interaction from the device's location. Therefore, ambient interaction naturally connected to a user's perception and was caused by a user's perception that one agent controlled the entire space. In addition, the participants perceived that the agent was looking at them anywhere they were in the setting, so they expected the agent to know all of their information and be optimized for them. Their expectation was that they could interact with personalized agents everywhere, which is similar to the goal of ambient intelligence [14, 48, 53] and J.A.R.V.I.S from the movie Iron Man [33]. In addition, it is aligned with the "one-for-all" concept of a previous study in which a singular agent inhabits multiple devices simultaneously, or "re-embodiment," which is when a singular agent can cross from one device to another across a task or a service [32]. Accordingly, the participants felt comfortable with the familiar agent's and agents' responses connected to users' daily behavior.

Many current systems aim to be single-agent systems with one persona controlling multiple devices in the house. For example, one embeds Google Assistant or Siri in various devices (a TV, a smartphone, and a smart speaker) [23, 47] to respond to user commands. Although current systems are designed to respond to the device closest to the user in a multi-device environment [3, 49], according to the drawings representing the visible agent in our findings, it can be difficult for users to perceive "one for all" when multiple devices are physically separated. Therefore, if designers intend to allow a single agent to control an entire house, then the existence of hubs representing a single agent in various devices needs to be clearly defined to make users recognize that there is one agent.

Furthermore, several participants perceived the single agent to be an omnipotent agent that knew a user's entire context and could respond anytime, anywhere. However, the current agent does not fulfill the users' expectations due to technological limitations. Although this gap is not easy to fill, this is not to say that one should just wait for technology to advance until J.A.R.V.I.S. becomes available [11]. Furthermore, it is impossible for agents to know and respond to all user contexts. Therefore, the user-centered design direction could address these challenges. For example, the agent could guess the user's status based on leveraging knowledge about the place, where the device is installed, or the user's location. Then, depending on the location, the agent can provide a response using the emotions that users have mainly expressed and the information that they had mainly wanted in a specific place. Although it is difficult for an agent to grasp subtle emotional changes from users' speech, it may be able to grasp users' needs in each space through accumulated data.

5.2 Design Direction of Invisible Agents for Ubiquitous Environments

In our previous discussions, we described features of visible and invisible CA and suggested that a physical entity should be considered in CA design in the future. In the case of a visible agent, the weakness of low learnability can be overcome through physical presence or visual cues. The visible agent has been used for various purposes in the home environment in the form of an AI speaker. Based on the trends, guidelines and considerations focused on the visible agent in terms of interaction design have been explored. However, although an invisible agent has the advantages of multitasking and ambient interaction, it has not yet been commercialized and explored in terms of interaction design. In our study, we not only ascertained users' mental models of invisible agents but we also explored users' interactions with an invisible agent. Based on our findings, we discuss the design direction of invisible agents that are not dealt with in terms of user-centered design in the previous studies as compared to AI speakers with physical entities. In addition, we discuss the tradeoffs between invisible and visible agents and explore future design directions for the ubiquitous home.

According to Edwards and Grinter, a smart home as a viable place to live should understand the users' context, respond to the users' natural and seamless interaction input, and support users without interfering with their previous behavior. In addition, ambient intelligence should serve people unobtrusively [1, 22, 48]. We found that ambient interaction with an invisible agent could support a better ubiquitous environment based on the relationship between a feature of ambient intelligence and users' mental model of invisible agents. This is because ambient interaction with an invisible agent leads to natural multitasking at home, where various activities occur. Additionally, users in the home may need more multitasking capability so they can focus on a main task, and agents can support users' subtasks more frequently in an unobtrusive way. For example, a user can naturally ask for recipes while cooking and not be interrupted by having to look at an AI speaker. Invisible agents in the ubiquitous environment can enable such features for natural multitasking.

This study showed that users expected an invisible agent to make full use all of the users' context and information as an omnipotent agent because the agent could detect the users' behavior and status without space limitations, everywhere and anywhere in the home. In addition, the participants said that if the environment in which the user communicates with the agent is a public space, then they would not expect the agent to make full use of their information because of privacy issues. Therefore, we assumed that one's home would be the appropriate environment for the agent to use personal information positively. This invisible agent can provide some services autonomously in response to perceived needs and accept user input through voice commands along with the development of natural language processing. However, privacy is a continuous issue in VUI that only interacts with voice [4, 27]. Therefore, even if the users were more tolerant of an invisible agent using their personal information, both invisible and visible agents must build a trust relationship with users regarding the privacy issue. Moreover, CAs need to provide users with transparency about the accumulated personal data from home, such as a data policy and means of protection. Moreover, companies need to integrate their conversation privacy dialog in an easily accessible way through natural interaction by clearly defining in their privacy policies or terms of use how users' data will be used [27].

To provide a positive user experience, the invisible agent needs to solve problems related to the users' cognitive barriers. Our study showed users faced awkwardness and ambiguity due to the absence of an entity from which the sound came: "I felt uncomfortable and a sense of cognitive discrepancy with talking to an invisible agent because there's a sound, and something is coming from somewhere, but there is nothing—just empty space" (P21). Our findings align with previous studies that have also revealed cognitive discrepancy must be reconsidered for better user experience [48]. To reconsider the issue, it is important to understand users' perceptions of invisible agents. In addition, a process that allows the user to adapt to the agent cognitively must be considered during the initial when a user forms his or her mental model of the invisible agent. For this purpose, the agent can refer to a self-disclosure strategy that clearly reveals the users' information used in human communication [18]. In addition, while visible agents can visually indicate inactivity or pull users' attention, invisible agents are difficult to convey this kind of information. Therefore, the traits of a visible agent (e.g., visual cues) can be applied to an invisible agent to enhance learnability and discoverability. If the interaction design issues regarding user–invisible agent relationships are not addressed as ambient intelligence technology advances, then it will be difficult for an invisible agent to lead to a positive user experience. This is because the gap between the users' mental model and an agent can be the main barrier in using a VUI [11]. Therefore, a voice-controlled ubiquitous environment could be reified by considering possible future problems in interaction design, such as cognitive discrepancy and lower discoverability.

In discovering these possibilities and design directions, we assumed that an invisible agent could be a new direction that supports smart homes along with an agent with a physical entity. This is because the users' mental model of invisible agents aligns with the completely autonomous environment targeted by academics and companies. Therefore, we suggest that an invisible agent can be the future direction of VUI for a successful ubiquitous environment because an invisible agent can make ambient interaction possible without space limitations. Then, if designers carefully consider the gap between users' mental models and the actual usage of an invisible agent in terms of interaction design, it would help users achieve more natural multitasking based on the grasp of all users' needs. A personalized agent such as the movie Iron Man would be possible through ambient interaction without space constraints. Moreover, when comparing these invisible features to those of visible agents in previous studies, invisible agents may have limitations on social interaction rather than visible agents such as social robot when interacting with people who are not familiar with using VUI (e.g., children or seniors) [20, 42, 54, 56]. Therefore, depending on the role of the agent, end-user, and situation, the physical form of an agent should be properly considered. Based on this study, we found a CA's physical entity influenced the user experience. The advantages and disadvantages of interaction and the difference between the mental models of each user depend on the presence or absence of a physical entity such as features of the visible agent's high learnability and the possibility of social interaction through directed interaction. In addition, the features of an invisible agent are capable of ambient interaction. Based on these considerations, we discovered that tradeoffs between visible and invisible agents depend on the situation of the smart home.

6 LIMITATIONS AND FUTURE WORK

In our study, we conducted a drawing study to compare users' mental models toward two agents: a visible agent and an invisible agent. In terms of the drawing method, because the perception toward CAs is derived from a small sample of users, it is difficult to say that the perceptions we have found represent whole characteristics. However, the result is meaningful, because we have seen users' diverse perceptions. In addition, although drawings offer various advantages, it is difficult for some adults to draw. However, our results were irrelevant to drawing skills, and we were able to overcome those difficulties with a comfortable environment and materials. In terms of the invisible agent, we adjusted the volume of AI speakers to implement an invisible agent that was not yet commercialized. This has limitations when compared with the features of invisible agents to be developed in the future, as highlighted in previous studies. To allow users to feel they are actually using an invisible agent, we plan to add an additional study. In this study, since we identified the mental models of young users, the primary users of AI speakers, we need further exploration of users with a wide range of ages. In addition, the current study is limited in that the interaction and evaluation were conducted in a lab environment. User interaction varies in a real smart home in which a user actually lives. Understanding the user in the wild would result in different user expectations. Because users cannot assume a lab environment is completely same as their homes, they should actually interact with the agent in environments where they live. We believe that our study provides insights into envisioning the features of an invisible agent compared with current AI speakers, and it suggests a design direction to enhance the user experience in a smart home. However, further research could be conducted to deal with elements that help with the development direction of an invisible agent based on empirical data.

7 CONCLUSION

CAs have been incorporated into our lives in a variety of forms, and in the future, they may be invisible without a physical device. In this study, we conducted a qualitative study of 30 users' drawings to explore their mental models of visible and invisible agents, as well as the design direction leading to an agent for a better voice-controlled home environment. Our study showed how CAs with different physical presences affect users' mental models and interactions. We described the characteristics of a current stand-alone device and a future invisible agent for use in future smart homes. Because the findings from users' drawings, interviews, observations, and a survey differentiated depending on the CA's physical presence, our study showed that the characteristics of each agent should be appropriately used depending on the situation in which the directed interaction and ambient interaction are needed to support the user. Furthermore, unconditionally focusing on technological advancement was not a solution for bridging the gap between the mental models and actual use with users. Therefore, technology development, such as invisible intelligence, must be considered in terms of the interaction design. In this study, through a user-centered approach, we proposed design directions for CAs that better support users and voice-controlled homes.

REFERENCES

[1] Emile Aarts and Reiner Wichert. 2009. Ambient intelligence. In *Technology guide*. Springer, 244–249.

[2] Gregory D Abowd and Elizabeth D Mynatt. 2000. Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 1 (2000), 29–58.

[3] Amazon Alexa. Accessed 2020. Echo Spatial Perception (ESP). https://developer.amazon.com/blogs/alexa/post/042be85c-5a62-4c55-a18d-d7a82cf394df/esp-moves-to-the-cloud-for-alexa-enabled-devices.

[4] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput.-Hum. Interact.* 26, 3, Article 17 (April 2019), 28 pages. https://doi.org/10.1145/3311956

[5] Juan Carlos Augusto and Paul McCullagh. 2007. Ambient intelligence: Concepts and applications. *Computer Science and Information Systems* 4, 1 (2007), 1–27.

[6] Erin Beneteau, Yini Guan, Olivia K. Richards, Mingrui Ray Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2020. Assumptions Checked: How Families Learn About and Use the Echo Dot. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 3 (March 2020), 23 pages. https://doi.org/10.1145/3380993

[7] Holly P Branigan, Martin J Pickering, Jamie Pearson, Janet F McLean, and Ash Brown. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121, 1 (2011), 41–57.

[8] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. 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, 1–11. https://doi.org/10.1145/3290605.3300270

[9] Robert C Burns and S Harvard Kaufman. 2013. *Action, Styles, And Symbols In Kinetic Family Drawings Kfd.* Routledge.

[10] Ana Paula Chaves and Marco Aurelio Gerosa. 2018. Single or Multiple Conversational Agents? An Interactional Coherence Comparison. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173765

[11] Minji Cho, Sang-su Lee, and Kun-Pyo Lee. 2019. Once a Kind Friend is Now a Thing: Understanding How Conversational Agents at Home Are Forgotten. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1557–1569. https://doi.org/10.1145/3322276.3322332

[12] Michael H Cohen, Michael Harris Cohen, James P Giangola, and Jennifer Balogh. 2004. *Voice user interface design.* Addison-Wesley Professional.

[13] Diane J Cook, Juan C Augusto, and Vikramaditya R Jakkula. 2009. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing* 5, 4 (2009), 277–298.

[14] D. J. Cook, M. Youngblood, E. O. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja. 2003. MavHome: an agent-based smart home. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003).* 521–524.

[15] Eric Corbett and Astrid Weber. 2016. What Can I Say? Addressing User Experience Challenges of a Mobile Voice User Interface for Accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 72–82. https://doi.org/10.1145/2935334.2935386

[16] Benjamin R Cowan, Holly P Branigan, Mateo Obregón, Enas Bugis, and Russell Beale. 2015. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human- computer dialogue. *International Journal of Human-Computer Studies* 83 (2015), 27–42.

[17] 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.

[18] Paul C Cozby. 1973. Self-disclosure: a literature review. *Psychological bulletin* 79, 2 (1973), 73.

[19] Andy Crabtree and Tom Rodden. 2004. Domestic routines and design for the home. *Computer Supported Cooperative Work* 13, 2 (2004), 191–220.

[20] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. " Hey Google is it OK if I eat you?" Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children*. 595–600.

[21] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google is It OK If I Eat You?": Initial Explorations in Child-Agent Interaction. Association for Computing Machinery, New York, NY, USA.

[22] W Keith Edwards and Rebecca E Grinter. 2001. At home with ubiquitous computing: Seven challenges. In *International conference on ubiquitous computing*. Springer, 256–272.

[23] Google. Accessed 2020. Google connected home. https://store.google.com/category/connected_home.

[24] Kerstin Heuwinkel. 2013. Framing the Invisible–The Social Background of Trust. In *Your Virtual Butler*. Springer, 16–26.

[25] K. Kim, L. Boelling, S. Haesler, J. Bailenson, G. Bruder, and G. F. Welch. 2018. Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 105–114.

[26] Anastasia Kuzminykh, Jenny Sun, Nivetha Govindaraju, Jeff Avery, and Edward Lank. 2020. Genie in the Bottle: Anthropomorphized Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376665

[27] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (Nov. 2018), 31 pages. https://doi.org/10.1145/3274371

[28] Sunok Lee, Sungbae Kim, and Sangsu Lee. 2019. "What Does Your Agent Look like?": A Drawing Study to Understand Users' Perceived Persona of Conversational Agent. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312796

[29] Yonnim Lee and Ohbyung Kwon. 2011. Intimacy, familiarity and continuance intention: An extended expectation–confirmation model in web-based services. *Electronic Commerce Research and Applications* 10, 3 (2011), 342 – 357. https://doi.org/10.1016/j.elerap.2010.11.005

[30] 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

[31] Michal Luria, Guy Hoffman, and Oren Zuckerman. 2017. Comparing social robot, screen and voice interfaces for smart-home control. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 580–628.

[32] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of Social Presence for Robots and Conversational Agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 633–644. https://doi.org/10.1145/3322276.3322340

[33] Iron Man. Accessed 2020. J.A.R.V.I.S. https://en.wikipedia.org/wiki/J.A.R.V.I.S.

[34] Sven Mayer, Gierad Laput, and Chris Harrison. 2020. Enhancing Mobile Voice Assistants with WorldGaze. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376479

[35] Chelsea M. Myers, Anushay Furqan, and Jichen Zhu. 2019. The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface. 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, 1–9. https://doi.org/10.1145/3290605.3300277

[36] Clifford Ivar Nass and Scott Brave. 2005. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press Cambridge, MA.

[37] Margaret Naumburg. 1966. *Dynamically oriented art therapy: Its principles and practices, illustrated with three case studies*. Grune Stratton.

[38] Naver. Accessed 2020. Clova. https://clova.ai/ko.

[39] Sunjeong Park and Youn-kyung Lim. 2020. Investigating User Expectations on the Roles of Family-Shared AI Speakers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376450

[40] 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174214

[41] Carlos Ramos, Juan Carlos Augusto, and Daniel Shapiro. 2008. Ambient intelligence—the next step for artificial intelligence. *IEEE Intelligent Systems* 23, 2 (2008), 15–18.

[42] L. Ring, B. Barry, K. Totzke, and T. Bickmore. 2013. Addressing Loneliness and Isolation in Older Adults: Proactive Affective Agents Provide Better Support. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 61–66.

[43] Giuseppe Riva, Francesco Vatalaro, and Fabrizio Davide. 2005. *Ambient intelligence: the evolution of technology, communication and cognition towards the future of human-computer interaction.* Vol. 6. IOS press.

[44] Jibo robot. Accessed 2020. https://www.jibo.com/.

[45] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. Association for Computing Machinery, New York, NY, USA.

[46] Cláudia Silva, Catia Prandi, Marta Ferreira, Valentina Nisi, and Nuno Jardim Nunes. 2019. See the World Through the Eyes of a Child: Learning from Children's Cognitive Maps for the Design of Child-Targeted Locative Systems. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 763–776. https://doi.org/10.1145/3322276.3323700

[47] Apple Siri. Accessed 2020. https://www.apple.com/siri/.

[48] N. Streitz and P. Nixon. 2005. The Disappearing Computer. *Communications - ACM* 48, 3 (2005), 32–35. https://strathprints.strath.ac.uk/2563/

[49] Support.Google. Accessed 2020. Google Home phone respond to Ok Google. https://support.google.com/googlenest/answer/7257763.

[50] British television drama series. Accessed 2020. Years and Years. https://en.wikipedia.org/wiki/Years_and_Years_(TV_series).

[51] Voicebot.ai. Accessed 2020. Voice Assistant Demographic Data – Young Consumers More Likely to Own Smart Speakers While Over 60 Bias Toward Alexa and Siri. https://voicebot.ai/2019/06/21/voice-assistant-demographic-data-young-consumers-more-likely-to-own-smart-speakers-while-over-60-bias-toward-alexa-and-siri/.

[52] Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring Virtual Agents for Augmented Reality. 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/10.1145/3290605.3300511

[53] Mark Weiser and John Seely Brown. 1997. *The Coming Age of Calm Technology.* Springer New York, New York, NY, 75–85. https://doi.org/10.1007/978-1-4612-0685-9_6

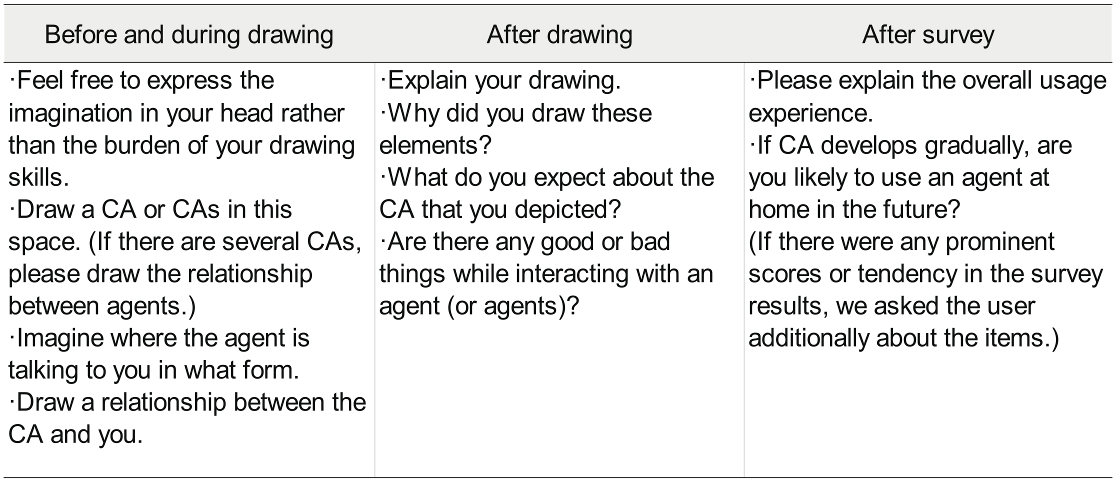
[54] Jacqueline M. Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal. 2018. Measuring Young Children's Long-Term Relationships with Social Robots. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (Trondheim, Norway) *(IDC '18)*. Association for Computing Machinery, New York, NY, USA, 207–218. https://doi.org/10.1145/3202185.3202732

[55] Ying Xu and Mark Warschauer. 2020. What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376416

[56] 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* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188575

A APPENDIX

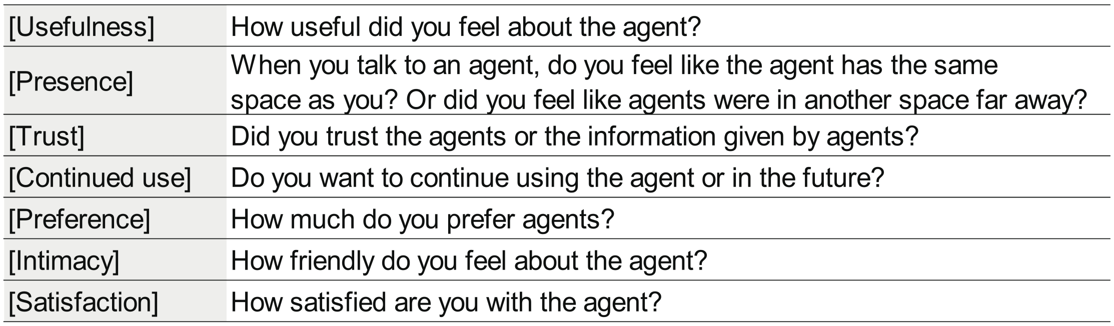
A.1 Session Script



Before and during drawing	After drawing	After survey
·Feel free to express the imagination in your head rather than the burden of your drawing skills. ·Draw a CA or CAs in this space. (If there are several CAs, please draw the relationship between agents.) ·Imagine where the agent is talking to you in what form. ·Draw a relationship between the CA and you.	·Explain your drawing. ·Why did you draw these elements? ·What do you expect about the CA that you depicted? ·Are there any good or bad things while interacting with an agent (or agents)?	·Please explain the overall usage experience. ·If CA develops gradually, are you likely to use an agent at home in the future? (If there were any prominent scores or tendency in the survey results, we asked the user additionally about the items.)

A.2 Survey Instruments

The survey was conducted at 7 Likert-scale using Google survey tools.



[Usefulness]	How useful did you feel about the agent?
[Presence]	When you talk to an agent, do you feel like the agent has the same space as you? Or did you feel like agents were in another space far away?
[Trust]	Did you trust the agents or the information given by agents?
[Continued use]	Do you want to continue using the agent or in the future?
[Preference]	How much do you prefer agents?
[Intimacy]	How friendly do you feel about the agent?
[Satisfaction]	How satisfied are you with the agent?

A.3 Additional Drawings

Visible CA
Invisible CA