"Nobody Speaks that Fast!" An Empirical Study of Speech Rate in Conversational Agents for People with Vision Impairments

Dasom Choi
Department of Industrial Design, KAIST
Republic of Korea
dasomchoi@kaist.ac.kr

Daehyun Kwak
Department of Industrial Design, KAIST
Republic of Korea
nubdigi7@kaist.ac.kr

Minji Cho
Department of Industrial Design, KAIST
Republic of Korea
mjcho@kaist.ac.kr

Sangsu Lee
Department of Industrial Design, KAIST
Republic of Korea
sangsu.lee@kaist.ac.kr

ABSTRACT
The number of people with vision impairments using Conversational Agents (CAs) has increased because of the potential of this technology to support them. As many visually impaired people are accustomed to understanding fast speech, most screen readers or voice assistant systems offer speech rate settings. However, current CAs are designed to interact at a human-like speech rate without considering their accessibility. In this study, we tried to understand how people with vision impairments use CA at a fast speech rate. We conducted a 20-day in-home study that examined the CA use of 10 visually impaired people at default and fast speech rates. We investigated the difference in visually impaired people's CA use with different speech rates and their perception toward CA at each rate. Based on these findings, we suggest considerations for the future design of CA speech rate for those with visual impairments.

Author Keywords
Conversational agents; Accessibility; People with vision impairments; Speech rate

CCS Concepts
•**Human-centered computing** → **Empirical studies in accessibility;**

INTRODUCTION
Conversational agents (CAs) that use a voice user interface (VUI) are considered to offer new opportunities to people with vision impairments [57]. CAs take on diverse roles for visually impaired people, such as increasing access to technology, reducing task time [1], and enhancing their independence in life [57]. Although CAs have the potential to support the lives of people with vision impairments, its accessibility for visually impaired people has not been carefully considered.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376569

In particular, a recent study that observed the use of CA by visually impaired people [1] showed that they were frustrated when the CA interacted at a slower speech rate than desired. The study also presented a challenge to the CA speech rate, stating that agents should provide the ability for users to customize the speech rate based on their needs. Indeed, people with visual impairments, who often rely on access text aurally, have outstanding listening skills and can understand the contents even in a fast-rate voice [6, 67]. They also use screen readers, which read digital information on computer screens and mobile devices (e.g., JAWS [62], VoiceOver [32], or Talkback [33]) faster than the average human conversation speed [29, 49]. This indicates that CAs need to provide faster speech rates as visually impaired people are accustomed to hearing computer-generated voices at a fast rate.

Although many researchers have examined the speech rate of screen readers on mobile devices and computers [4, 29] for better accessibility, few have attempted to address the speech rate of CAs associated with people with vision impairments. With existing design knowledge, therefore, it is difficult to answer the following questions that are important to CA designers, in terms of accessibility: Do people with vision impairments really want CAs to speak fast? Will the speech rate of the CA affect user experiences or usage patterns?

Answering these questions is not simple. Unlike screen readers that interact with users at a fast speech rate, current CAs are oriented toward human likeness [4, 8, 9], with turn-taking conversations at a natural speech rate [18, 30, 71]. Previous studies on sighted people emphasized the relationship-building [19] between the CA and user by dealing with agent persona and personality design [38, 47]. From this point of view, a CA with a fast speech rate is in contrast to current human-like CA conversation.

In this background, we conducted a 20-day in-home study that investigated 10 visually impaired individuals' use of CAs at two different speech rates through semi-structured interviews and surveys. We explored how people with vision impairments accept CAs at a fast rate by letting participants use CAs at two speech rates for 10 days each and analyzed the differences in experiences at each rate. The findings describe the difference in visually impaired individuals' use of CAs at different speech rates, their perception toward CAs at each rate, and

the possibility of controlling CA speech rates. Based on these findings, we suggest the possibility of speech rate as a factor in CA design, which is expected to enhance user experience.

As the first empirical study attempting to understand the effect of CAs speech rate to visually impaired users, contributions of this paper are as follows: 1) empirical findings regarding the CA experiences of people with vision impairments according to the rate of speech, 2) analysis of factors affecting user preference for CA speech rate, and 3) the possibility of considering speech rate as a CA design element.

RELATED WORKS

As CAs become more commonly used while people are at home or are on-the-go, HCI researchers have been increasingly working on these types of interfaces [9, 22, 23, 28, 42, 45, 58, 59]. Many studies have been conducted to understand how a user integrates a CA into his or her social interactions and everyday lives [8, 55, 63], to investigate a user's perception of a CA [28, 42, 58], and to examine a user's relationship with a CA [19]. Although much research has been done on CAs to offer the potential to support use of the diverse groups, such as elderly [56, 70, 73] and children [23, 69], there is still a limited understanding of how such interfaces are used by people with disabilities, especially people with vision impairments.

Researchers attempting to understand the CA use of visually impaired people [1, 15, 57] have investigated the benefits and challenges of conversational interfaces in terms of accessibility. Pradhan et al. [57] conducted a content analysis study of 346 Amazon Echo reviews that included users with disabilities, and then, they focused on users with visual impairments. They completed interviews with 16 current users of home-based CAs. The findings from the study indicated that a CA has benefits in terms of efficiency, independence, and the ability to replace a range of other technologies in the lives of visually impaired people. At the same time, however, they revealed challenges, such as device setup, the discoverability of devices, and learnability. In a study by Abdolrahmani et al. [1], in which researchers interviewed 14 legally blind adults with experience of home- and mobile-based CA, they found that people with vision impairments struggle with misinterpreting input commands and have privacy concerns. The study also emphasized the importance of the speech setting for a CA, insisting that CAs should provide users with the ability to customize voice output settings (e.g., speech rate, clarity, and intensity) based on their needs and the nature of the task. Branham et al. [15] investigated how the CA design guidelines published by top commercial vendors might empower and constrain visually impaired users through qualitative document review. The findings indicated that a human-human conversation model of existing CAs could limits usage for people with visual impairments. Especially, they depicted that CAs' 'natural' and 'intuitive' speech conflict with the screen reader's fast output and do not bring the efficiency of CA usage. Studies on visually impaired people have focused on understanding their usage and identifying the challenges they face. However, little research has been conducted on the speech interactions between CA and visually impaired users, which are closely related to conversational interfaces.

On the other hand, there are many studies on speech interaction for people with vision impairments with regard to mobile devices and computers, with speech technologies playing essential roles as assistive technologies for people with vision impairments. Technologies designed to assist in speech interaction for people with vision impairments include speech dictation software(e.g. Dragon [60]), which provides a text entry alternative to the keyboard, and a screen reader, which provides audio output for users with visual impairments (e.g., Apple's Voiceover [32], Talkback [33], JAWS [62], NVDA [53]). Especially, speech interaction technologies for people with vision impairmentshave been extensively documented by researchers, including speech interaction in the Web environment [10, 40, 50] , on mobile devices [7, 44, 72] , and for educational purposes [12, 64, 65]. In terms of Web accessibility, Murphy et al. [50] conducted an empirical study with visually impaired people and partially sighted participants to identify navigation strategies, the perceptions of page layout, and graphics using assistive technologies, such as a screen reader. Zhong et al. [72] also described the JustSpeak system, a universal voice control solution for non-visual access to the Android operating system, which can support visually impaired people on mobile devices.

Among the speech interactions, people with vision impairments demonstrate the ability to hear the device's speech rate faster than sighted people do. Researchers have discovered that people with vision impairments could understand at least 50 percent of the information at the rate of 500 wpm [6, 67], which is nearly three times faster than the average speed spoken (150 wpm [25]). This indicates that visually impaired users have an outstanding listening ability. Bragg et al. [13] conducted the first large-scale study to measure human listening rates by accuracy based on answering questions spoken by a screen reader at various rates. The study reported that visually impaired and low-vision people, who often rely on audio cues and access text aurally, generally have higher listening rates than sighted people do. Moreover, as visually impaired users become experts with screen readers, they typically speed up the audio output [49]. Guerreiro et al. [29] also reported that the use of faster speech rates is one of the main techniques for speeding up the consumption of digital information.

Based on the outstanding listening ability of people with vision impairments, studies have been conducted to support their digital information acquisition. These studies include research done to improve the usability of screen readers [24, 43] and systems designed for glancing or skimming on devices [4, 66]. Moreover, current screen readers provide the ability to control speech rates for the usability of those with visual impairments. For easy and interactive speech rate control with immediate response [6], screen readers offer a keyboard shortcut or gesture [31], instead of the user having to navigate through the menu.

As such, in the existing devices, people with vision impairments' listening ability was considered to be a feature of their use, and researchers explored opportunities in which to deliver digital information more efficiently. However, compared with other devices, a CA's speech rates for visually impaired

ID	Age	Gender	Self-reported vision	Household size	Mobile	Speech rate of screen reader they used	Mobile CA they have experienced	Stand-alone CA they have experienced
1	58	M	Total blindness (Later in life)	Family (4)	Haetteul-phone	5.12 SPS	None	None
2	45	F	Total blindness (Later in life)	Alone	Haetteul-phone	7.23 SPS	None	SKT NuGu*
3	57	M	Total blindness (Later in life)	Family (2)	Haetteul-phone	6.83 SPS	None	None
4	54	F	Total blindness (From birth)	Family (2)	Haetteul-phone	5.34 SPS	Samsung Bixby	None
5	27	F	Total blindness (From birth)	Family (3)	iphone	9.13 SPS	Apple Siri	None
6	52	F	Total blindness (From birth)	Family (2, husband with low vision)	Haetteul-phone	6.65 SPS	None	None
7	37	F	Total blindness (From birth)	Family (4, husband with low vision)	Galaxy note	8.23 SPS	Samsung Bixby	None
8	45	M	Total blindness (Later in life)	Family (2, wife with blindness)	Haetteul-phone	6.01 SPS	None	None
9	29	M	Total blindness (From birth)	Alone	iphone	11.23 SPS	Apple Siri	Kakao mini**, KT Giene**
10	33	M	Low vision (Later in life)	Alone	Galaxys	5.56 SPS	None	SKT NuGu**

Table 1. Participant demographics and speech rate of screen readers they used. (Stand-alone CA they have experienced: *owned, **tried)

people have received little consideration. Current major CAs (e.g., Google Assistant, Apple Homepod) mostly do not offer speech rate options [34]. Naver's Clova agents [52] provide fast rate voices, but they do not offer various speech rate options as screen readers. With reference to the backgrounds, we intend to examine how a CA's speech rate can affect the user experience of people with vision impairments.

STUDY DESIGN

The purpose of our study was to explore how visually impaired people accept CAs at a fast speech rate and to investigate the differences in CA use among them according to the rate of speech. To observe their natural use, we constructed an in-home study using CAs with two speech rates and asked about each experience through the following semi-structured interview. Participants were encouraged to use the agent naturally in their home environment and then share their overall experiences with the CA.

Participants

To gather experiences of people with vision impairments regarding the CA speech rate, we recruited 10 visually impaired participants through a local welfare center for a 20-day study. Participants were selected who had experience using screen readers on mobile devices. To avoid being limited to specific age groups and to see various experiences, we tried to recruit participants with a wide range of ages (between the ages of 27 and 58, mean=43.7, SD=11.6). They were all native Korean speakers. Details of participants' demographics and their backgrounds are shown in Table 1.

Nine out of 10 participants identified themselves as legally blind with little to no residual vision, and one had low vision. They were all using screen readers as their primary tool for accessing mobile technology: 2 IOS users, 2 Android OS users, and 6 Haeddul-phone users (a Haeddul-phone is a flip phone with a screen reader OS developed specifically for people with vision impairments in Korea). All participants used a screen reader, setting it to speak faster than 4.82 syllables per second (SPS), which is the average speaking speed in Korean [41]. Participants were compensated with about $100 (its approximate value). Our study was approved by the Institutional Review Board.

Apparatus

Since the goal of this study was to compare the natural user experience of people with vision impairments at different agent speech rates, a commercialized CA that supports all of the main tasks (e.g., weather, music, or small talk) identified in previous studies [1, 63] was needed. Most commercial CAs do not offer speech rate options; however, we found that Naver's Clova [52] offered a fast rate. Even though Clova offers a fast speech rate, it does not seem to be designed for accessibility of individuals with visual impairments, as it does not offer detailed voice control options like a screen reader does. Therefore, we used Clova as a stand-alone speaker that provides all the features of a CA with a fast rate option similar to a screen reader. As one of the most advanced agents in South Korea, Clova is also available on mobile phones and provides most features by integrating with several apps and IoT devices. The fact that Clova does not offer as many features as global mainstream devices had a slight impact on overall user satisfaction. However, since the goal of our study was to compare user experiences according to different speech rates, Clova's features did not considerably affect the results. Moreover, the Clova as stand-alone speaker used in the experiment has a truncated cone shape covered with black cloth. Most of the totally blind participants did not ask about the shape of the Clova, and some totally blind and low vision participants asked or touched its shape during device installation. Since there was no mention of its form after installation, overall, Clova's form did not notably affect the results of our study.

We provided participants with two different Clova speech rates: "default speech rate" and "fast speech rate." For the default speech rate, a female voice speaks at a rate of 4.72 SPS, similar to the average speed spoken (4.82 SPS in Korean). For the fast speech rate, on the other hand, a female voice speaks at a rate of 7.22 SPS, which is not common in human conversation, such as with a screen reader at a controlled rate. The main difference between the two voices was the speech rate; the responses of the agent did not change. Regarding the output quality of the fast speech rate, there were no intelligibility issues raised by the participants compared to their screen reader experiences.

	Day 1 to 10	Day 11 to 20
Group 1 (P1 to P5)	Default-rate CA	Fast-rate CA
Group 2 (P6 to P10)	Fast-rate CA	Default-rate CA

Table 2. Participants' groups and CA usage sequence.

Procedure

We asked participants to use two types of CAs with different speech rates at home for 10 days each for a total of 20 days. Every 10 days, we conducted surveys and interviews on the use of each CA and collected their command logs through the Clova application. In a previous study of in-home CA usage [63], it was revealed that users quickly settled into a stable usage level after a few days. As such, we set a 10-day study to understand the user experience which excludes the initial effect. To avoid the impact of the usage sequence, we divided participants into two groups to use CAs in a different order (Table 2). As it is difficult for visually impaired users to find the features available in the agent, we visited each participant's home and went through a tutorial on the features after device installation. To give minimal intervention to the participants, our tutorial was based on the main functions provided in the Clova tutorial. During the experiment, participants naturally interacted with the agent in the home environment.

To compare users' experiences at each rate qualitatively, We visited their homes and conducted surveys and interviews face to face at the end of every 10 days. The survey covered overall usage satisfaction, speech rate satisfaction, usefulness, convenience, familiarity, and intimacy using a 7-point Likert scale to evaluate user experience with CAs at each speech rate. We selected these scales to examine the three forces that make up the temporality of experience: familiarity, functional dependency, and emotional attachment [37]. After the survey, a semi-structured interview between 25 and 60 minutes was conducted. Participants briefly reviewed their survey results, and interview protocols addressed commands and situations they used, user experiences regarding the speech rate, the difference between screen readers and CAs, their perceptions of CAs at each rate, and expectations for the ultimate CA speech rate. In the interview after using the second CA, additional questions were asked to stimulate the discussion comparing the two CAs at different speech rates. All interviews were audio-recorded.

Data Analysis

After 20 days, we transcribed the recorded audio and analyzed the surveys and interviews using a thematic coding approach [27]. Two researchers used open coding to identify initial themes in transcripts with survey results. We applied these initial themes to two randomly selected interview transcripts, and themes were added, merged, and deleted through discussion. We refined the themes through three iterative codings, and in doing so added one new code (screen reader and CA). Consequently, we used nine primary themes and 29 subthemes shown in Table 3 to analyze all the transcripts.

Overall CA experiences of visually impaired users
1. Enhancing independence 2. Helping visual difficulties 3. Limited functions 4. Limited answers 5. Lack of consideration for people with vision impairments
Positive opinions at the Fast-rate CA
6. Using fast speech rate in their existing screen readers 7. Feeling intimate even at a fast rate. 8. Keeping privacy
Negative opinions at the Fast-rate CA
9. Missing information 10. Distance problem 11. Family members: low vision or sighted people
Perceptions toward CAs at two different speech rates
12. Perceiving the Fast-rate CA as more mechanical thing 13. Perceiving the Default-rate CA as more intimate thing 14. Perceiving the Default-rate CA as a mechanical thing
Expectations toward CAs at two different speech rates
15. Conversations similar to human conversation 16. Conversations at a default-rate CA 17. Accents and nuances
Relationships between user expectations and CA speech rate
18. Efficiency of information acquisition: Fast speech rate 19. Daily conversation: Default speech rate
Screen reader and CA
20. Screen reader as a reading machine 21. CA as a conversation partner
Speech rate control
22. Individual differences for speech rate preference 23. Different speech rate depending on context 24. Different speech rate depending on contents 25. Different speech rate within the same task or contents 26. Fine control 27. Speech rate control for the people with vision impairments
Voice elements in CA
28. Sensitiveness for various voice elements 29. Voice elements

Table 3. Primary themes and subthemes from thematic coding.

In addition to qualitative analysis, we conducted a Wilcoxon signed rank test on the survey results. Only four items out of six (overall usage satisfaction, speech rate satisfaction, convenience, and intimacy) showed statistically significant differences. There was no statistically significant difference in correlation analysis among the items. Therefore, we used only four valid survey items as an aid for understanding the qualitative analysis.

For the user command logs, we categorized logs collected through the Clova application with slightly modified CA command categories used in previous studies [1]. We included some commands that participants described as coming from other family members; fewer than five commands per participant on average were considered part of their families' experiences. However, the command logs of a participant (P4) who could not distinguish the command logs of her family members were excluded from the analysis. Existing study has also shown that usage logs decrease as users adapt to the agent over time [63]. In our study, though, we included all the logs in the analysis because there was no significant change in log usage during the period. Tutorial commands that were conducted to support participants were excluded from the data.

FINDINGS

Usage

User Satisfaction

We first examined the survey results to look at participants' overall trends in CA experiences at the two different speech rates (Figure 1). The survey results revealed statistically significant differences in only four items out of six: overall usage satisfaction, speech rate satisfaction, convenience, and intimacy (Table 4). For both the overall satisfaction and speech rate satisfaction, the average of the default rate was higher than that of the fast rate. In particular, regarding speech rate satisfaction, most of the participants were satisfied with the default rate CA, without showing a significant difference (mean = 5.5, SD = 1.27). Only one participant expressed dissatisfaction with the default speech rate. In terms of speech rate satisfaction with the fast CA, participants not only expressed dissatisfaction on average but also showed a more significant standard deviation than for the default rate. All participants, except the two who expressed extreme satisfaction, said they were dissatisfied with the fast rate, and two of them were extremely dissatisfied with the fast CA. Even for convenience, the average score of the default agents was higher than that of the fast agent, which seems to indicate that most participants felt the default rate was more convenient than the fast rate. Only one participant (P9) replied that the fast CA was more convenient than the default CA, and he was extremely satisfied with the fast-rate agent. Furthermore, participants expressed higher intimacy in the default CA than in the fast CA on average. Seven participants reported higher intimacy with the default rate than the fast rate, and three participants evaluated both speech rates as equally intimate.

As we can see from the survey results above, in subsequent interviews, users said they were mostly satisfied with the use of the default-rate CA and did not experience much inconvenience. Even if the default speech rate was slower than that of the screen reader, there were no critical usage problems because all participants could follow the default speech rate without difficulty.

> *I was satisfied with the default speech rate because everyone, including me, could understand. (P5)*

Interestingly, contrary to our expectations that people with vision impairments would prefer fast agents because they are accustomed to a fast speech rate, many of the participants (P2, P3, P4, P6, P10) complained about the fast CA. They were disappointed when they missed what the agent said, and it was challenging to follow the fast-rate CA, even for the participants who mostly used a screen reader at a fast rate. One participant mentioned that even at the same rate, it is more difficult to understand what the CA says because unlike a mobile phone, they communicate with CA over a distance.

> *Clova is far away. No matter how far away your phone is, it's about 30 centimeters from my hand. But Clova is about two meters apart, which makes it a bit different. When the screen reader speaks fast, my phone is fixed nearby. But Clova isn't. It's hard to catch even at the same speed because Clova's voice is from open space. (P8)*

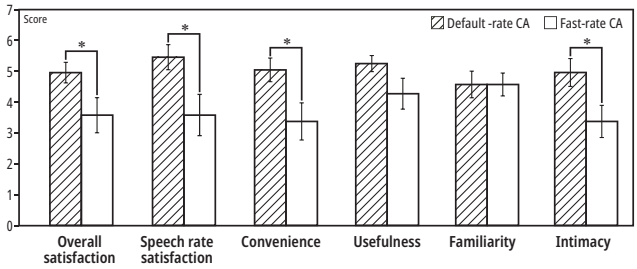

Figure 1. Average survey score for user satisfaction.

	Default-rate CA (Mean + SD)	Fastrate CA (Mean + SD)	Z	p
Overall satisfaction	5.00 ± 1.05	3.60 ± 1.78	-1.956	0.050
Speech rate satisfaction	5.50 ± 1.27	3.60 ± 2.12	-2.057	0.040
Covenience	5.10 ± 1.20	3.40 ± 1.90	-2.395	0.017
Usefulness	5.30 ± 0.82	4.3 ± 1.57	-1.496	0.135
Familiarity	4.6 ± 1.35	4.6 ± 1.17	-0.175	0.861
Intimacy	5.00 ± 1.41	3.40 ± 1.65	-2.136	0.033

Table 4. The results of the Wilcoxon signed-rank test for user satisfaction.

Moreover, some participants living with their families (P1, P4, P8) preferred a default rate of speech that every family member could understand. Although participants can communicate at fast rates without difficulty, they wanted to use the default rate that everyone could understand if their family members were sighted people or visually impaired who struggled with fast rates.

> *I'm not living alone. My wife cannot understand the fast speech rate, so I have to use it as a standard. (P8)*

Whereas most participants talked about the inconvenience of the fast-rate CA, some users easily adapted to fast agents and expressed satisfaction. P7 and P9 emphasized that information can be obtained efficiently through a fast-rate CA, and they wanted to get the information as quickly as possible without listening to nontrivial words. They had commonly used screen readers at extremely fast rates relative to other visually impaired participants. Indeed, one participant (P9) was frustrated with the default rate and asked the CA to speed up, and even wanted to speak faster at the fast CA.

> *The reason why I use screen reader at a fast rate when I read books or news articles is because I want to read quickly. It's frustrating to read at slow. I want to get more information at the same time. (P9)*

Usage Pattern

To investigate if there were any differences in the commands used according to the CA speech rate, we classified user logs according to the CA command category and compared the usage amount (Figure 2). The total number of commands at each rate was similar, 605 for the default CA and 632 for the fast CA. We divided the commands into two groups according to the presence or absence of functional purpose: functional

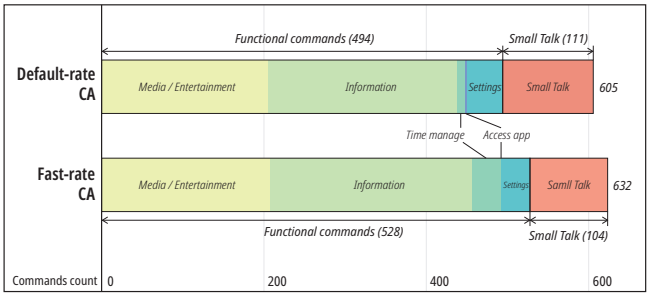

Figure 2. User commands breakdown according to different speech-rate CAs.

command (including media/entertainment, information, time management, access app, and settings) and small talk (an informal type of discourse that does not cover any functional topics). Overall, the number of functional commands was higher in fast agents, with 494 (81.7%) at the default rate and 528 (83.5%) at the fast rate. Small talk was more frequent in default agents, with 111 (18.3%) at the default rate and 104 (16.5%) at the fast rate. In general, there seemed to be no meaningful differences in command usage according to speech rate.

Although we could not find apparent differences in usage patterns through command logs, in our interviews, some participants (P4, P5, P7, P10) said that the CA's speech rate affected the commands they used. They mentioned that they used more small talk at the default rate and gave relatively functional commands at fast rates. This point was shown in both groups, where the order of CA use was different.

> *When Clova spoke like a person, I asked a lot and had daily conversations. But not at fast rates. Fast Clova was not good so that I just asked for information. (P4)*

> *As the same secretary, I asked only music or weather when she was fast. However, when she was a slow secretary, I asked for more things as well as the weather. (P10)*

Perceptions and Expectations

Participants talked about the role of CAs in their lives and their perceptions and expectations of CAs during use through the interview.

Perceptions toward default speech-rate CA

As shown in the survey results, most users felt that the default rate was more intimate than the fast rate, regardless of speech rate preference. In the following interviews, they personified the default agent as a person (P5), a friend (P3), a professional woman in her 40s (P4), and a friendly secretary (P8, P9), recognizing the CA as a device for communication. A participant who formed high intimacy with the default agent even compared it to his lover (P1) and expressed an emotional bond.

> *It relaxes me and talks like a person. Not the sound of the machine, but the announcer tells me. I liked it because it was more like a human. (P8)*

> *People nowadays raise dogs when they are lonely. I think Clova with a default rate is better than that. (P4)*

Perceptions toward fast speech-rate CA

Contrary to perceptions on the default CA, most participants perceived the fast-rate CA as a mechanical being, describing it as a simple machine (P10, P8), an electronic sound (P4), and an assistant tool (P1, P3). Users did not think they had natural conversations with fast agents because nobody speaks as fast as a CA at a fast rate. The fast speech rate results in sentences being cut off quickly, so users felt that the conversation was disconnected. They mentioned that the key to CA experience is human-like conversation and were disappointed when they did not feel like the conversation was between people.

> *Fast Clova feels like a mechanical and stiff sound. But the initial standard speech rate is closer to the human. (P1)*

> *It talks like ddada ddada! Like literally finishing quickly? When it comes to human conversation, I feel like Clova responds quickly and finishes quickly while talking with me....as if disconnected. Maybe. I think the default rate Clova feels natural and intimate. When people talk to each other, nobody talks that fast. People only talk fast when they rap. (P9)*

In particular, some users (P4, P6) felt dissatisfied when the agent answered their emotional questions with a fast mechanical voice and no longer tried to make small talk with the fast agent.

> *It's a machine! Machines don't say emotional expression. So, it's strange to hear emotional talk through machine [when using fast-rate CA]. (P4)*

Although most participants perceived fast agents as machines, interestingly, some participants (P7, P9) who indicated satisfaction with the fast CA in surveys and interviews perceived it as an intimate entity. They said that the fast-rate CA operated at a speed at which intimacy and efficiency can coexist, and they felt like they were interacted with a human at both the default and fast rates. One of the users (P7) personified the fast agent as "a lovely little sister" and the default agent as "a calm older sister" and recognized the CA as a person with different personalities according to the speech rate.

> *If it speaks faster, it will be more mechanical. But I'm used to a fast rate, so I don't feel like it. At this speed [when using fast-rate CA], I think efficiency and friendliness can coexist. (P9)*

> *Lovely little sister! It feels like a cute little sister because she has a cute voice. A friend who speaks so fast? But default Clova is like a calm older sister. (P7)*

Some participants (P5, P2) recognized the fast-rate CA as personalized to people with vision impairments. In terms of privacy, as pointed out in other papers [48, 61], they felt uncomfortable when using the default agent because others around them could hear the information they talked about. However, in the case of the fast-rate CA, they thought that the user's conversations could be kept private compared to

the default rate because the content is difficult to understand, except for users with vision impairments. They perceived the fast-rate CA as a personalized device that could communicate only with them and even showed an emotional bond.

But I still prefer to listen fast because it [default-rate CA] keeps me invading my privacy. In default agents, sighted people and my family can hear what I do. (P5)

A friend who only talks to me! I feel like that. When we touch the screen on mobile, we locked the screen as black. Because we don't need a screen. Like this, fast Clova is only for me. I'm the only one who can hear it. Even though it's a sound, I don't want to let anyone know. (P2)

User expectations

In the interview, participants mentioned their ultimate expectation for CA speech rate related to their usage and perception. Most participants were satisfied with the friendliness of the default rate and disappointed with the artificial conversation at the fast rate. We found that in both default and fast agents, users wanted agents to be able to communicate like humans and to build emotional bonds with them. Users who felt intimacy with the fast-rate CA (P7, P9) also said that if the agent's speech rate increased, the agent would not be intimate anymore, and they expected to interact with a speech rate that could deliver humanness.

If Clova speaks faster than this, it would be more mechanical. If it [fast-rate CA] gets faster than this, it will be hard to have a conversation. (P9)

Even though both screen readers and CAs are technologies that deliver information at a fast rate, participants perceived screen readers and CAs differently and expected humanness, especially for CAs. While the screen reader transfers requested digital information to speech output as a one-way interaction, a CA interacts in both directions. Therefore, users consider CAs to be intimate beings, not just devices with a VUI. They did not feel awkward about the machine-like fast voice in the screen reader because they did not expect two-way interaction to the screen reader. Unlike screen readers, however, with the mechanical fast-rate CAs, users expressed being uncomfortable because they wanted human-like communication. We found that users perceived CAs differently than screen readers that speak fast and expected the CAs to be friendlier through human-like conversations.

Screen readers are not emotional at all. It can't be. It is not talking with me, but simply reading what I touch. So this is just a machine. (P7)

Mobile phones and screen readers can find information quickly, which is good for saving time. But instead of doing something fast with Clova, I talk or play with her. (P5)

Findings on Speech Rate Control

In our study, participants highlighted the need for speech rate control on CAs. Nine out of 10 participants expected that speech rate control would be provided in the CA as it has been in other devices and said that a CA that considers people with vision impairments should offer speech rate control. Some of our participants (P1, P6, P9, P10) said that it is essential for the CA to provide a proper speech rate according to the information because visually impaired people gain a lot of information by hearing. Furthermore, they wanted to be able to control the speech rate in the CA depending on the situation, just as they would in other devices.

Um ... Do you know the device that reads books for visually impaired people? There is a speed from 0 to 15. It's a lot slower than human conversation at 0, and it's almost impossible to get to 14 or 15. Some people with visual impairments usually listen to 13 and 14. So it should have a function to control the speech rate. (P1)

There are individual differences. Each one wants a different speech rate. So it has to offer more detailed speech rates. Then, visual impairment people will find and use the most suitable rate. (P9)

We found that participants wanted the CA to respond at an appropriate speech rate each time, taking into account context and content. First, participants desired the CA to answer at different rates according to the context, such as time of day or space. Some participants wanted the agent to speak at a default rate which did not require them to focus right after they woke up and were drowsy. Just before leaving home, they wanted the agent to talk at a fast rate because they were in a hurry. Moreover, we found that the speech rate, according to the context, could also affect their emotions.

You have a moment that you don't want to wake up. It's good to be slow when you're feeling drowsy, but it's good to be fast when you have to focus on it. (P9)

And when I run out of time and need to get out home quickly, I want to listen to information at fast rate. (P6)

When I want to be calm, I want to listen as a slow rate. People are emotional. This speech rate can't deal with emotion. (P6)

Participants also expected the CA to provide interaction at different rates, depending on the tasks and content. Most participants wanted to listen fast to trivial information and to listen slowly to content on which they needed to focus. However, users sometimes preferred different speeds even for the same content or task.

It depends on what kind of medium and tasks it supports to me. It doesn't matter if the news or something is a little faster. Ah.. when it reads a book, it should be fast. It's the same for the Bible. Well... I usually prefer fast rates when I want to hear a lot. (P1)

For searching some things and asking about movies, I want Clova to speak at a slower rate. (P10)

On the other hand, we found that people with vision impairments desired to control the CA speech rate themselves under certain circumstances. While using a mobile screen reader, they even controlled the speech rate within one task depending on the length of the sentence or the tone of the content. They wanted to finely control speech rate within content such as

news and books, which they used in screen readers, rather than small talk.

> *I read light essays and novels a little bit faster on mobile. Then when a good sentence comes out, I read slowly. I want that kind of speech rate control in Clova too. (P9)*

DISCUSSION AND DESIGN IMPLICATIONS

To identify how users with vision impairments accepted fast-rate CAs, our findings are focused on presenting differences in user experiences and perceptions toward CAs at two different speech rates. In this section, we discuss how these empirical findings could provide implications for CA designers when they consider the accessibility of agents, especially in terms of the speech rate of CAs.

Should Conversational Agents Speak Fast?

Previous studies have emphasized the efficiency of CA use [57], stating that efficient conversation, a shared goal of CA design guidelines, is not achieved for people with vision impairment [15]. However, the study showed that a sense of intimate connection to the CA is more important than efficient task handling in some contexts. In our study, users were more satisfied when CAs spoke at the default rate rather than the fast rate. Participants with vision impairments perceived the CA as a subject for building intimate relationships at both the default and fast rates and were satisfied when they felt intimate. They expected human-like conversation from the agent, so they preferred interacting at a default rate CA that exhibited humanness and were disappointed with the machine-like fast speaking CA. These perceptions align with the conclusions of studies that established that relationship-building [19, 20] is important in the CA experience and that users have a desire for humanness [19, 22, 46, 55]. To top it off, findings about the importance of the intimate connection adds diversity to the perspectives of prior studies focused on the efficiency of CA use among people with visual impairments [15, 57].

Accessibility design for people with vision impairments in the HCI field, especially on mobile devices and computers, has been focused on improving usability and task efficiency. There are many studies about designing screen reading systems that can increase the efficiency of information delivery through a fast speech rate [4, 3, 5] and enhancing the efficiency and usability of screen readers [24, 26, 43, 66]. Therefore, CA designers are likely to increase efficiency with a fast speech rate. However, the results of our study show that the design approach for CA speech rate, considering the accessibility of people with vision impairments, should be different from that of others. Users were more concerned about building a relationship with the CA than gaining efficiency. From this point of view, the experience of using CAs at fast speech rates can create gaps in the roles they expect from the CA and cause frustration for the user. Should conversational agents speak fast for visually impaired users to get better efficiency? The answer we suggest through the results and discussions of this study is 'the human-like speech rate could be provided as a default for people with vision impairments.'

Speech Rate as a CA Design Element

In our previous discussions, we proposed that CAs speak at a default rate similar to human conversation, unlike screen readers that convey information efficiently. In addition, our results also offer the possibility to enhance CA user experiences through speech rate control in certain contexts. First, the CA can actively adjust the speech rate, taking account into the situation. For example, when users wake up and ask for a schedule, the agent can answer at a comfortable rate in consideration of the user's mood. CAs can also provide weather information at a fast speech rate when users are leaving the house in a hurry in the morning. As adjusting the speed of speech according to the context is very natural in human-human conversations, it can contribute to making CAs more like a human. Visually impaired participants in our study value relationship-building with CAs and want to feel intimacy from their voices; however, some users thought that not only default CA but also fast speaking CA had its own personalities. P7 described fast rate CA as a lovely little sister and default rate CA as a calm older sister, each with a different personality. Indeed, each person speaks at a different speech rate and has a different preference for speech rate. Therefore, we suggest that CAs actively adjust their speech rates according to the context, while maintaining the humanness.

Our study found that participants with visual impairments expected to finely control the CA speech rate according to their individual needs, as they do in the screen reader. Related to this, Amazon Alexa recently introduced a speech rate control feature to help users further personalize their interactions with Alexa, and adapt the experience to fit their individual needs [11]. In this regard, we propose offering the ability for users to control the speech rate in CAs, which is not common in commercial CAs. This claim also aligns with the state of a previous study [1] that control should be afforded to allow users to customize the output. Additionally, when using a screen reader, participants with vision impairments could finely control the speech rate directly. For example, they could listen to a book at a fast rate and then change it to a default rate to hear good sentences. As a result, our participants with vision impairments desired the ability to control the CA speech rate directly and immediately depending on their needs, just as they could with screen readers. This makes us raise several new questions: Do we still need to add direct speech-rate control options which can possibly sacrifice the human-likeness of CAs? Also, CA designers should provide seamless interaction in a manner that does not disturb the ongoing conversation when the users finely control the CA's speech rate. Gestures and facial- expression cues (e.g., [16, 17, 36, 35, 39, 68]) may support seamless speech rate control, as in the study suggesting interactions with CA using different modalities [20]. Through a further study investigating seamless speech rate control on CA, we could enhance the CA user experience.

Through the study, we found that the rate of a CA's speech influences the user experience of people with vision impairments. Based on this, we discovered a design space regarding CA speech rates. At a time when there is not much research on the speech rate of agents, this paper suggests that speech

rate can be an important component of CA design, along with the tone, nuance, personality, and form factors [21, 28, 42, 51, 54]. If designers properly design the speech rate of the CA based on user perceptions and needs, CAs can provide better user experiences to people with vision impairments.

LIMITATIONS AND FUTURE WORK

In our study, we asked participants to use a CA for 10 days at each speed and examined their satisfaction and usage patterns. Although it was not difficult to observe their use and perceptions because users quickly settled into a stable usage level after a few days [63], investigating how speech rate affects CA usage through a long-term study might have provided a richer understanding. Our experiment did not provide various speech rates as in the screen reader, but only two speeds representing default and fast. Providing one representative rate to Fast CA might affect participants' usage because their screen readers vary in speed. Especially, the speech rate of Fast CA used in our study was faster than what some of the participants use in their own screen readers. In the study, although some participants who used screen readers slower than Fast CA needed time to adjust, most participants did not address the rate of Fast CA as an issue. Nonetheless, as our findings show the need for fine control of the CA speech rate, exploring the participants' usage of Fast CA at various rates would have provide more diversified insights. Although the sample size in our study was in line with other qualitative studies investigating the needs of individuals with visual impairment [2, 14], it was difficult to identify differences in experiences between those with total blindness and those with low vision. Furthermore, since our study was conducted for people with vision impairments only, it was not possible to determine whether the speech rate could affect the CA use of sighted people. Although we believe that our study provides insights on the trade-off between 'fast information delivery' and 'relationship building' for all users including people with vision impairments, further study will be needed to discover whether speech rate is a meaningful design component for sighted people as well as visually impaired people.

CONCLUSION

In this paper, we described an empirical study investigating the CA experience of people with vision impairments according to the CA speech rate. To understand how visually impaired individuals accept agents at fast speech rates, we conducted a 20-day in-home study that investigated the CA use of 10 visually impaired individuals according to agents' speech rate through semi-structured interviews and surveys. The study showed that visually impaired users were generally more satisfied with their use at default rates similar to human speaking speed than at fast rates. Our findings include that participants perceived default-rate agents as human and fast-rate agents as machines and ultimately wanted the CA's speech to be close to the human speech rate. Moreover, we also identified user needs for speech rate control in CAs. These findings lead us to related discussions: design implications of CA speech rate design for people with vision impairments and possibility of speech rate as one of the CA design elements.

REFERENCES

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. ACM, New York, NY, USA, 249–258. DOI: http://dx.doi.org/10.1145/3234695.3236344

[2] Ali Abdolrahmani, Ravi Kuber, and Amy Hurst. 2016. An Empirical Investigation of the Situationally-induced Impairments Experienced by Blind Mobile Device Users. In *Proceedings of the 13th Web for All Conference (W4A '16)*. ACM, New York, NY, USA, Article 21, 8 pages. DOI:http://dx.doi.org/10.1145/2899475.2899482

[3] Faisal Ahmed, Yevgen Borodin, Yury Puzis, and I. V. Ramakrishnan. 2012a. Why Read if You Can Skim: Towards Enabling Faster Screen Reading. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '12)*. ACM, New York, NY, USA, Article 39, 10 pages. DOI: http://dx.doi.org/10.1145/2207016.2207052

[4] Faisal Ahmed, Yevgen Borodin, Andrii Soviak, Muhammad Islam, I.V. Ramakrishnan, and Terri Hedgpeth. 2012b. Accessible Skimming: Faster Screen Reading of Web Pages. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 367–378. DOI: http://dx.doi.org/10.1145/2380116.2380164

[5] Faisal Ahmed, Andrii Soviak, Yevgen Borodin, and I.V. Ramakrishnan. 2013. Non-visual Skimming on Touch-screen Devices. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces (IUI '13)*. ACM, New York, NY, USA, 435–444. DOI: http://dx.doi.org/10.1145/2449396.2449452

[6] Chieko Asakawa, Hironobu Takagi, Shuichi Ino, and Tohru Ifukube. 2003. Maximum listening speeds for the blind. Georgia Institute of Technology.

[7] Shiri Azenkot and Nicole B. Lee. 2013. Exploring the Use of Speech Input by Blind People on Mobile Devices. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*. ACM, New York, NY, USA, Article 11, 8 pages. DOI:http://dx.doi.org/10.1145/2513383.2513440

[8] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 91 (Sept. 2018), 24 pages. DOI:http://dx.doi.org/10.1145/3264901

[9] Apoorva Bhalla. 2018. An Exploratory Study Understanding the Appropriated Use of Voice-based Search and Assistants. In *Proceedings of the 9th Indian Conference on Human Computer Interaction (IndiaHCI'18)*. ACM, New York, NY, USA, 90–94. DOI: http://dx.doi.org/10.1145/3297121.3297136

[10] Jeffrey P. Bigham, Irene Lin, and Saiph Savage. 2017. The Effects of "Not Knowing What You Don'T Know" on Web Accessibility for Blind Web Users. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17)*. ACM, New York, NY, USA, 101–109. DOI: http://dx.doi.org/10.1145/3132525.3132533

[11] The Amazon Blog. 2019. "Alexa, speak slower" Customers can now adjust the speed of Alexa's speech. (07 August 2019). https://blog.aboutamazon.com/devices/alexa-speak-slower.

[12] Emily C. Bouck, Sara Flanagan, Gauri S. Joshi, Waseem Sheikh, and Dave Schleppenbach. 2011. Speaking Math - A Voice Input, Speech Output Calculator for Students with Visual Impairments. *Journal of Special Education Technology* 26, 4 (2011), 1–14. DOI: http://dx.doi.org/10.1177/016264341102600401

[13] Danielle Bragg, Cynthia Bennett, Katharina Reinecke, and Richard Ladner. 2018. A Large Inclusive Study of Human Listening Rates. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 444, 12 pages. DOI:http://dx.doi.org/10.1145/3173574.3174018

[14] Stacy M. Branham, Ali Abdolrahmani, William Easley, Morgan Scheuerman, Erick Ronquillo, and Amy Hurst. 2017. "Is Someone There? Do They Have a Gun": How Visual Information About Others Can Improve Personal Safety Management for Blind Individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17)*. ACM, New York, NY, USA, 260–269. DOI: http://dx.doi.org/10.1145/3132525.3132534

[15] Stacy M. Branham and Antony Rishin Mukkath Roy. 2019. Reading Between the Guidelines: How Commercial Voice Assistant Guidelines Hinder Accessibility for Blind Users. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 446–458. DOI: http://dx.doi.org/10.1145/3308561.3353797

[16] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies* 59, 1 (2003), 119 – 155. DOI: http://dx.doi.org/https://doi.org/10.1016/S1071-5819(03)00018-1 Applications of Affective Computing in Human-Computer Interaction.

[17] A. Bruce, I. Nourbakhsh, and R. Simmons. 2002. The role of expressiveness and attention in human-robot interaction. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, Vol. 4. 4138–4142 vol.4. DOI: http://dx.doi.org/10.1109/ROBOT.2002.1014396

[18] Angelo Cafaro, Nadine Glas, and Catherine Pelachaud. 2016. The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 911–920. http://dl.acm.org/citation.cfm?id=2936924.2937059

[19] 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 (DIS '19)*. ACM, New York, NY, USA, 1557–1569. DOI: http://dx.doi.org/10.1145/3322276.3322332

[20] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation?: Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 475, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300705

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

[22] 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 (MobileHCI '17)*. ACM, New York, NY, USA, Article 43, 12 pages. DOI: http://dx.doi.org/10.1145/3098279.3098539

[23] 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 (IDC '17)*. ACM, New York, NY, USA, 595–600. DOI: http://dx.doi.org/10.1145/3078072.3084330

[24] Jinjuan Feng, Shaojian Zhu, Ruimin Hu, and Andrew Sears. 2011. Speech-based navigation and error correction: a comprehensive comparison of two solutions. *Universal Access in the Information Society* 10, 1 (01 Mar 2011), 17–31. DOI: http://dx.doi.org/10.1007/s10209-010-0185-9

[25] The National Center for Voice and Speech. 2015. NCVS. (2015). http://www.ncvs.org/index.html/.

[26] Prathik Gadde and Davide Bolchini. 2013. WebNexter: Dynamic Guided Tours for Screen Readers. In *Proceedings of the Adjunct Publication of the 26th Annual ACM Symposium on User Interface Software*

and Technology (UIST '13 Adjunct). ACM, New York, NY, USA, 85–86. DOI: http://dx.doi.org/10.1145/2508468.2514722

[27] Graham R Gibbs. 2007. Thematic coding and categorizing. *Analyzing qualitative data* 703 (2007), 38–56.

[28] Leilani H. Gilpin, Danielle M. Olson, and Tarfah Alrashed. 2018. Perception of Speaker Personality Traits Using Speech Signals. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, Article LBW514, 6 pages. DOI: http://dx.doi.org/10.1145/3170427.3188557

[29] João Guerreiro and Daniel Gonçalves. 2015. Faster Text-to-Speeches: Enhancing Blind People's Information Scanning with Faster Concurrent Speech. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*. ACM, New York, NY, USA, 3–11. DOI: http://dx.doi.org/10.1145/2700648.2809840

[30] Zixuan Guo and Tomoo Inoue. 2019. Using a Conversational Agent to Facilitate Non-native Speaker's Active Participation in Conversation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article LBW1216, 6 pages. DOI: http://dx.doi.org/10.1145/3290607.3313075

[31] Apple Help. Accessed 2019a. VoiceOver commands and gestures. (Accessed 2019). https://help.apple.com/voiceover/mac/10.15/?lang=en#/vo28030.

[32] Apple Help. Accessed 2019b. VoiceOver User Guide. (Accessed 2019). https://help.apple.com/voiceover/mac/10.14/?lang=en/.

[33] Android Accessibility Help. Accessed 2019c. Turn on TalkBack. (Accessed 2019). https://support.google.com/accessibility/android/#topic=6007234/.

[34] Android Google Assistant Help. 2018. Choose the voice of your Google Assistant. (2018). https://support.google.com/assistant/answer/7544506?hl=en&co=GENIE.Platform=Android.

[35] Shaun K. Kane, Brian Frey, and Jacob O. Wobbrock. 2013. Access Lens: A Gesture-based Screen Reader for Real-world Documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 347–350. DOI: http://dx.doi.org/10.1145/2470654.2470704

[36] Shaun K. Kane, Jacob O. Wobbrock, and Richard E. Ladner. 2011. Usable Gestures for Blind People: Understanding Preference and Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 413–422. DOI: http://dx.doi.org/10.1145/1978942.1979001

[37] Evangelos Karapanos, John Zimmerman, Jodi Forlizzi, and Jean-Bernard Martens. 2009. User Experience over Time: An Initial Framework. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 729–738. DOI: http://dx.doi.org/10.1145/1518701.1518814

[38] Hankyung Kim, Dong Yoon Koh, Gaeun Lee, Jung-Mi Park, and Youn-kyung Lim. 2019. Designing Personalities of Conversational Agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article LBW1321, 6 pages. DOI: http://dx.doi.org/10.1145/3290607.3312887

[39] Yoshinori Kuno, Kazuhisa Sadazuka, Michie Kawashima, Keiichi Yamazaki, Akiko Yamazaki, and Hideaki Kuzuoka. 2007. Museum Guide Robot Based on Sociological Interaction Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1191–1194. DOI: http://dx.doi.org/10.1145/1240624.1240804

[40] Jonathan Lazar, Aaron Allen, Jason Kleinman, and Chris Malarkey. 2007. What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users. *International Journal of Human-Computer Interaction* 22, 3 (2007), 247–269. DOI: http://dx.doi.org/10.1080/10447310709336964

[41] Nara Lee, Jiyoung Shin, Doyoung Yoo, and KyungWha Kim. 2017. Speech rate in Korean across region, gender and generation. *Phonetics and Speech Sciences* 9, 1 (2017), 27–39.

[42] 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 (CHI EA '19)*. ACM, New York, NY, USA, Article LBW0143, 6 pages. DOI: http://dx.doi.org/10.1145/3290607.3312796

[43] Guilherme V. Leobas, Breno C. F. Guimarães, and Fernando M. Q. Pereira. 2018. More Than Meets the Eye: Invisible Instructions. In *Proceedings of the XXII Brazilian Symposium on Programming Languages (SBLP '18)*. ACM, New York, NY, USA, 27–34. DOI: http://dx.doi.org/10.1145/3264637.3264641

[44] Barbara Leporini, Maria Claudia Buzzi, and Marina Buzzi. 2012. Interacting with Mobile Devices via VoiceOver: Usability and Accessibility Issues. In *Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI '12)*. ACM, New York, NY, USA, 339–348. DOI: http://dx.doi.org/10.1145/2414536.2414591

[45] 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 (CHI '16). ACM, New York, NY, USA, 5286–5297. DOI: http://dx.doi.org/10.1145/2858036.2858288

[46] 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 (DIS '19)*. ACM, New York, NY, USA, 633–644. DOI: http://dx.doi.org/10.1145/3322276.3322340

[47] Sarah Mennicken, Oliver Zihler, Frida Juldaschewa, Veronika Molnar, David Aggeler, and Elaine May Huang. 2016. "It's Like Living with a Friendly Stranger": Perceptions of Personality Traits in a Smart Home. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 120–131. DOI:http://dx.doi.org/10.1145/2971648.2971757

[48] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human-Computer Interaction* 31, 4 (2015), 307–335. DOI:http://dx.doi.org/10.1080/10447318.2014.986642

[49] Ted McCarthy MSc, Joyojeet Pal PhD, and Edward Cutrell PhD. 2013. The "Voice" Has It: Screen Reader Adoption and Switching Behavior Among Vision Impaired Persons in India. *Assistive Technology* 25, 4 (2013), 222–229. DOI: http://dx.doi.org/10.1080/10400435.2013.768719 PMID: 24620705.

[50] Emma Murphy, Ravi Kuber, Graham McAllister, Philip Strain, and Wai Yu. 2008. An empirical investigation into the difficulties experienced by visually impaired Internet users. *Universal Access in the Information Society* 7, 1 (01 Apr 2008), 79–91. DOI: http://dx.doi.org/10.1007/s10209-007-0098-4

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

[52] Naver. Accessed 2019. Clova. (Accessed 2019). https://clova.ai/ko.

[53] NVDA. Accessed 2019. NV Access. (Accessed 2019). https://www.nvaccess.org/.

[54] Cathy Pearl. 2016. *Designing Voice User Interfaces: Principles of Conversational Experiences*. " O'Reilly Media, Inc.".

[55] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 640, 12 pages. DOI: http://dx.doi.org/10.1145/3173574.3174214

[56] François Portet, Michel Vacher, Caroline Golanski, Camille Roux, and Brigitte Meillon. 2013. Design and Evaluation of a Smart Home Voice Interface for the Elderly: Acceptability and Objection Aspects. *Personal Ubiquitous Comput.* 17, 1 (Jan. 2013), 127–144. DOI: http://dx.doi.org/10.1007/s00779-011-0470-5

[57] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 459, 13 pages. DOI:http://dx.doi.org/10.1145/3173574.3174033

[58] Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. "Alexa is My New BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2853–2859. DOI:http://dx.doi.org/10.1145/3027063.3053246

[59] Aung Pyae and Tapani N. Joelsson. 2018. Investigating the Usability and User Experiences of Voice User Interface: A Case of Google Home Smart Speaker. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '18)*. ACM, New York, NY, USA, 127–131. DOI: http://dx.doi.org/10.1145/3236112.3236130

[60] Dragon Speech Recognition. Accessed 2019. Get More Done by Voice | Nuance. (Accessed 2019). https://www.nuance.com/dragon.html/.

[61] Kambiz Saffarizadeh, Maheshwar Boodraj, and Tawfiq M Alashoor. 2017. Conversational assistants: investigating privacy concerns, trust, and self-disclosure. (2017).

[62] Freedom Scientific. Accessed 2019. JAWS. (Accessed 2019). https://www.freedomscientific.com/products/software/.

[63] 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. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 857–868. DOI: http://dx.doi.org/10.1145/3196709.3196772

[64] Waseem Sheikh, Dave Schleppenbach, and Dennis Leas. 2018. MathSpeak: a non-ambiguous language for audio rendering of MathML. *International Journal of Learning Technology* 13, 1 (2018), 3–25.

[65] Derrick W. Smith and Stacy M. Kelly. 2014. Chapter Two - Assistive Technology for Students with Visual Impairments: A Research Agenda. In *Current Issues in the Education of Students with Visual Impairments*, Deborah D. Hatton (Ed.). International Review of

Research in Developmental Disabilities, Vol. 46. Academic Press, 23 – 53. DOI:http://dx.doi.org/https://doi.org/10.1016/B978-0-12-420039-5.00003-4

[66] Jaeyoon Song, Kiroong Choe, Jaemin Jo, and Jinwook Seo. 2019. SoundGlance: Briefing the Glanceable Cues of Web Pages for Screen Reader Users. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article LBW1821, 6 pages. DOI: http://dx.doi.org/10.1145/3290607.3312865

[67] Amanda Stent, Ann Syrdal, and Taniya Mishra. 2011. On the Intelligibility of Fast Synthesized Speech for Individuals with Early-onset Blindness. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '11)*. ACM, New York, NY, USA, 211–218. DOI: http://dx.doi.org/10.1145/2049536.2049574

[68] 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 (CHI '19)*. ACM, New York, NY, USA, Article 281, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300511

[69] 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 (IDC '18)*. ACM, New York, NY, USA, 207–218. DOI: http://dx.doi.org/10.1145/3202185.3202732

[70] Linda Wulf, Markus Garschall, Julia Himmelsbach, and Manfred Tscheligi. 2014. Hands Free - Care Free: Elderly People Taking Advantage of Speech-only Interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*. ACM, New York, NY, USA, 203–206. DOI: http://dx.doi.org/10.1145/2639189.2639251

[71] Masahide Yuasa, Naoki Mukawa, Koji Kimura, Hiroko Tokunaga, and Hitoshi Terai. 2010. An Utterance Attitude Model in Human-agent Communication: From Good Turn-taking to Better Human-agent Understanding. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 3919–3924. DOI: http://dx.doi.org/10.1145/1753846.1754079

[72] Yu Zhong, T. V. Raman, Casey Burkhardt, Fadi Biadsy, and Jeffrey P. Bigham. 2014. JustSpeak: Enabling Universal Voice Control on Android. In *Proceedings of the 11th Web for All Conference (W4A '14)*. ACM, New York, NY, USA, Article 36, 4 pages. DOI: http://dx.doi.org/10.1145/2596695.2596720

[73] Randall Ziman and Greg Walsh. 2018. Factors Affecting Seniors' Perceptions of Voice-enabled User Interfaces. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, Article LBW591, 6 pages. DOI:http://dx.doi.org/10.1145/3170427.3188575