Notification, Wake-Up, and Feedback of Conversational Natural User Interface for the Deaf and Hard of Hearing

Takashi Kato*, Akihisa Shitara†, Nobuko Kato*, and Yuhki Shiraishi*
*Tsukuba University of Technology, Japan
Email: {a203101,nobuko,yuhkis}@a.tsukuba-tech.ac.jp
†University of Tsukuba, Japan
Email: theta-akihisa@digitalnature.slis.tsukuba.ac.jp

***Abstract*—Most voice-based conversational natural user interfaces (NUIs), such as Amazon Alexa and Google Assistant, rely on speech input and output, posing an accessibility barrier for the deaf and hard of hearing (DHH). For example, DHH users may not be aware of notifications from the system, may not receive response information, and the system may have difficulty recognizing their wake-words. In designing a conversational NUI for DHH users, we consider that simply replacing speech information with sign language information does not suffice to create an accessible, comfortable user experience. In this study, we conducted an experiment with 12 DHH users to determine whether luminous notifications and text display methods showing sign language in place of the standard text output were effective, as well as whether gazing was effective as a wake-up method. The second experiment was conducted with 24 DHH users to identify better wake-up and feedback presentation methods. We propose conversational NUI guidelines for DHH users based on the results of these experiments. We examined accessibility options for DHH users at each step of the conversation with the voice user interface (VUI), and expect this work to serve as a basis for future conversational NUI design.**

Keywords*—*Deaf; hard of hearing; hearing impaired; sign language; accessibility.

I. INTRODUCTION

In this study, to propose design guidelines for a conversational natural user interfaces (NUIs) for dead and hard of hearing (DHH) users, we examined accessibility methods at each step of a conversation with voice user interfaces (VUIs). This study builds on and extends our previous work [1], in which we first introduced a conversational NUI for DHH users.

User interfaces (UIs) are a necessary medium for passing information between computers and humans. Research on NUIs designed to enable natural and intuitive operation by humans has progressed in recent years. Applications include touch panels, gesture control systems, and VUIs. For greater convenience and ease of use, “conversational” interaction with users is required to enable intuitive operation and mental support [2]. Advances in speech recognition, speech synthesis, and natural language processing have enabled humans to interact naturally with UIs [3] However, it remains difficult for DHH users to use and converse with VUI systems due to their inability to receive speech information as audio and high word error rate (WER). WER is a criterion used to evaluate speech recognition technology, which indicates the probability of words that could not be heard over the total number of words uttered by a human. The WER of a speech recognition system developed by Google averaged less than five percent in 2017 [4]. However, Bigham and others reported a WER for DHH users of 40% [5]. Abraham showed that current speech recognition technology is not yet usable by DHH users [6]. Moreover, he also reported that low WER could be overcome if more speech data from DHH individuals could be collected. However, collecting large amounts of speech data from DHH individuals would require considerable time and expense. As a result of the widespread use of Amazon Alexa and Google Assistant, voice control is becoming a ubiquitous interface technology. As this trend continues, an increasing need to address accessibility issues for DHH users in this technology is evident.

Accessibility studies on conversational user interfaces (CUIs) for DHH users have reported that sign language is more suitable than gestures and text as an alternative input method to replace speech [7]. It has also been reported that the use of sign language is preferable to touchscreens as an input method [8], and DHH users are interested in interacting with a system in sign language [9]. As a result, researchers in human-computer interaction (HCI) have begun to consider user interfaces that can interact with sign language [10], and the design and construction of sign language interfaces have been recognized as a notable research topic [11]. In recent years, researchers have begun to evaluate technologies for sign language recognition, generation, and translation from an interdisciplinary perspective [12], and this topic was addressed at a workshop on UIs [11]. Various calls to action were presented, including “develop user interface guidelines for sign language systems”. However, basic research on how sign language users interact with computational systems remains relatively rare in the relevant literature. Although many systems have been developed for sign language users, most studies have focused on evaluating the systems themselves and did not outline the interaction principles between the user and system. As a result, each team developing a new system must design the interface almost from scratch, without the benefit of general design guidelines based on research. Therefore, design guidelines for NUIs should be developed, especially sign language-based NUIs, which have been the subject of considerable research.

In this study, we investigate whether graphical user interface (GUI) and VUI design guidelines can be used as a reference when developing guidelines for conversational NUI from the perspective of DHH users. The main GUI design guidelines include “10 Usability Heuristics for User Interface Design” by Nielsen [13] [14], “Seven user-centered design principles” by Norman [15], and “Eight Golden Rules of Interface Design” by Shneiderman [16]. GUIs have long used heuristics as a key to implementing successful designs and avoiding usability problems. GUI design guidelines cannot be directly applied to VUIs [17]. VUIs and NUIs, including sign language conversations, use touchless interaction and natural language to interact with users. Therefore, it is challenging to devise NUI guidelines based on GUI design guidelines, including for sign language. Next, Alexa [18] and Google Assistant [19] are mainly considered in terms of VUI design guidelines. It has been reported as essential to establish a foundation of VUI principles to guide future designers in developing spoken conversation systems [20]. In addition, guidelines for VUIs are designed to assume speech-based interactions. The author believes that simply replacing speech information, mainly used when conversing with VUIs, with sign language information is not sufficient to create a comfortable user experience for DHH individuals. Therefore, to create conversational NUI guidelines from the perspective of DHH users, we consider alternative access methods for DHH users at each step of a conversation with a VUI. Based on the above, we propose guidelines for conversational NUI that are most suitable for DHH users based on the existing conversation steps with VUI.

From the conversational process between a listener and a conversational NUI (referred to herein as “the system”), we clarify the elements necessary to create a comfortable user experience for DHH users. Figure 1 shows the conversion process. At the beginning of the conversation, the user must invoke the system with a wake-word. A wake-word is a simple spoken phrase that triggers the AI assistant to accept speech input, such as “Alexa.” for Alexa, “OK, Google.” for Google Assistant, and “Hey, Siri” for Siri. The system then detects the user’s wake-up command and provides feedback that the system is ready to accept commands. For example, Alexa displays a blue bar at the bottom of the screen, and Google Assistant displays a bar at the top to provide feedback to the user’s command. The user then commands the system to access weather, news, alarms, etc. The system then executes the task according to the user’s commands. Other conversational situations include “alarm notifications” and “video calls” in situations where the system calls the user. We believe that simply substituting sign language for voice input is not sufficient to create a comfortable user experience for DHH users who mainly use visual information. Therefore, in addition to sign language, more optimal means of notification, information transmission, wake-up, and feedback presentation for DHH users must be identified.

First, we consider the notification method. Hearing persons can obtain audio information from the system without constantly looking at the system. For DHH users who cannot acquire audio information, it may be challenging to catch a response from the system when they are not looking at the display, even if they use a device with a display (Echo Show or Nest Hub). That is, the best output method of the system

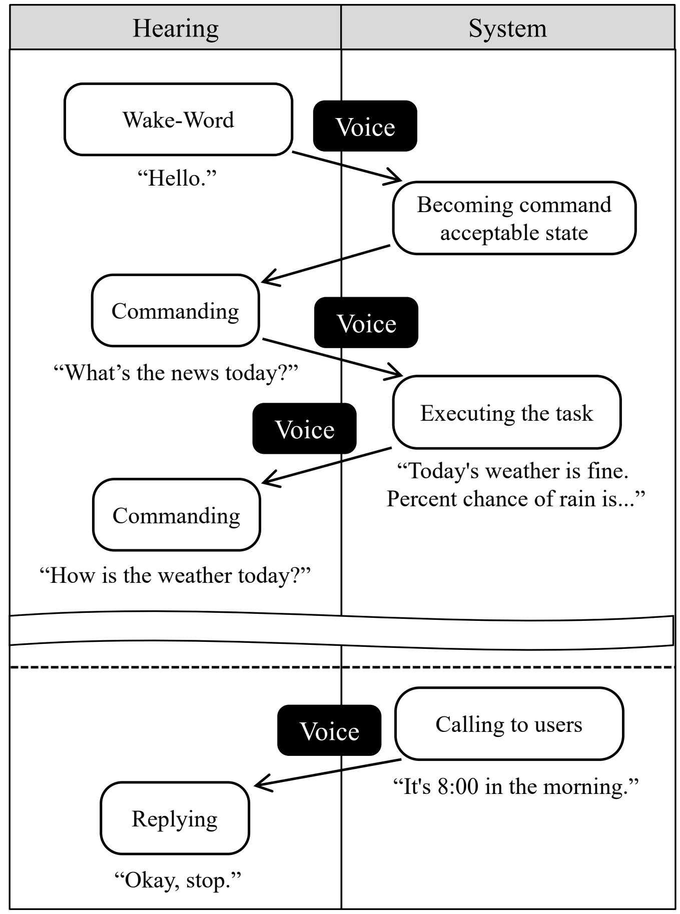

Figure 1. Activity of the conversation process between the Hearing and the system.

for DHH users should be investigated to identify an alternative to voice output provided by CUIs. By contrast, luminous notifications are familiar in Deaf culture. For example, DHH users use intercoms, alarms, and fire alarms with luminous notification functions in their daily lives [21]. In addition, a luminous device that transmits the direction of the sound source of the surrounding alarms to DHH users with light has been developed [22]. Figure 2 shows that we investigated whether luminous notifications could improve usability for DHH users.

Next, we consider the method of information transmission. Subtitling has recently become an accessibility feature for CUIs with displays [23]. Even if DHH users use this feature, there is a concern that the user experience may be worse if the system outputs subtitles, because an interaction will mainly be in sign language if sign language input from the user is enabled. With devices such as Alexa entering common usage in homes, there have been reports of increased hands-free interaction with devices placed in the kitchen or living room [24]. Therefore, it is conceivable that people may be more likely to interact with computational agents while doing other things. Therefore, DHH users should also be able to capture responses from CUIs while doing other things.

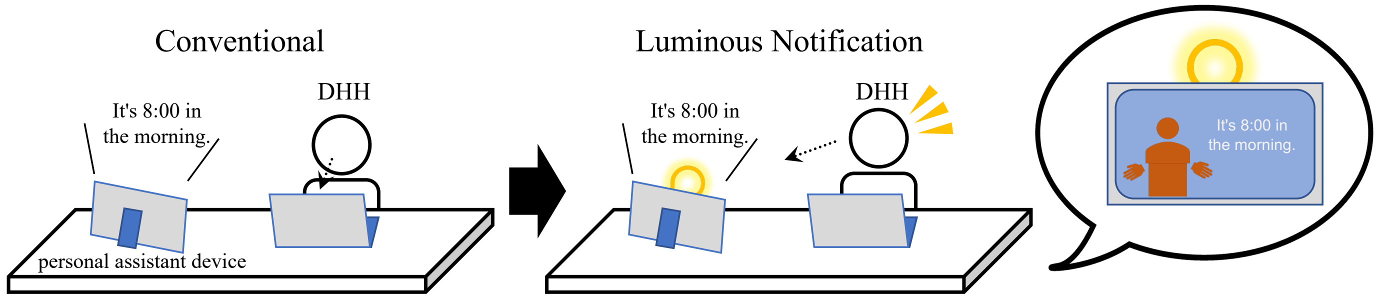

Figure 2. Hypothesis in light notification.

Here, clarifying whether the system's sign language or text output method affects the user experience of DHH users based on the difference between sign language flowing and text shown for a given time is essential. A study was conducted on the impact of sign language interpretation, subtitling, and the two together on the support of deaf students in secondary and higher education settings [25]. The study reported that the performance of the DHH students was significantly higher when only subtitles were used compared to the other two conditions. One study evaluated subtitles and sign language's combined effect in understanding television content [26]. Here, it was reported that providing subtitles and sign language was very helpful in improving the accessibility of TV content to more DHH individuals than providing each of the two alone. In addition, a study investigated whether incorporating sign language video into text-based web pages improved accessibility for DHH users [27]. The authors reported that information presented through sign language videos increased the interest of DHH users on the Web. However, all these studies were conducted for one-way media.

Conversational NUIs with two-way interaction require input from the user, and clarifying the preferences of users for sign language or text display methods is essential in such cases. Moreover, whereas designers must translate speech into the speaker's language in the case of television and the Web, CUIs are transmitted by a computer, so the language can be adjusted to suit the recipient. In other words, the designer should consider output methods using sign language or text along with users preferences in terms of attributes. Therefore, we investigate the necessity of sign language and subtitles for DHH use3rs under the condition of parallel work when CUIs provide both.

Then, we consider the wake-up method. DHH individuals need to make eye contact with a person they intend to talk to in order to start a conversation with them, and they do this by tapping their shoulder or waving [28] [29]. The preferred wake-up techniques of DHH users in descending order of preference include the use of the ASL sign name of the device, waving in the direction of the device, clapping, using a remote control, using a phone app, and fingerspelling the English name of the device [30]. Here, user preference for the sign name exceeded that of waving owing to the concern that the system would recognize unintended waving as a wake-up command. However, the use of eye contact, which is essential in starting a conversation in interpersonal communication with DHH individuals, has not been examined. "Eye contact" evokes and facilitates others' behavior and is involved in the initiation and progression of conversational interaction. We believe that gaze allows for a natural interaction without an explicit wake-up and increases user satisfaction. Based on the above, the possibility of using gazing in a system for DHH users should be considered, and the optimal wake-up method should be selected from among signing a name, waving, and eye gazing.

Finally, we consider how feedback is presented. In contrast to those with hearing, interpersonal communication with a DHH individual requires their attention when calling out to them. This suggests that hearing people need not confirm the feedback presented by the system compared to DHH individuals. It is becoming more common for listeners to enter commands in succession after a wake-word when operating such systems by voice. Hearing people assume that they are constantly being listened to by others, whereas DHH individuals typically believe they are only being listened to when others are looking. As for how Alexa devices present feedback, Echo Show 5 shows a blue bar, and Echo Show 10 shows a shaking head motion. It remains unclear whether these feedback methods will improve the user experience for DHH users. Therefore, investigate feedback presentation for DHH users should be investigated to consider better feedback methods.

To create a comfortable user experience for DHH users, this study examined the following research questions, and we propose guidelines based on the results.

- RQ1: Does the light-based response of the CUI improve usability for DHH users?
- RQ2: What is the best sign language/text display method for CUI for DHH users?
- RQ3: Is gaze tracking an effective method of waking up a CUI for DHH users?
- RQ4: Would DHH users prefer to see an indication from the system that it is ready to accept commands before giving them?
- RQ5: What might be a better way for DHH users to wake up a system?
- RQ6: Is there a better way for the system to indicate that

it is ready to accept a command when it detects DHH users' wake-up commands?

We surveyed students at the Tsukuba University of Technology [33] to investigate these research questions. The Tsukuba University of Technology is a university for the visually and hearing impaired. The hearing level of the DHH students is generally 60dB or higher in both ears. Tsukuba University of Technology students include not only deaf but also hard of hearing students.

This study contributes empirical knowledge regarding the preferences and concerns of DHH users regarding features such as notification, information transmission, wake-up, and feedback presentation, aiming to provide useful guidance for future system designers.

The remainder of this study is organized as follows. In Section II, we describe related studies on smart speakers and gesture interfaces and the Wizard of Oz method. In Section III, we investigate whether a sign language conversation system using luminous notification and gaze tracking can improve usability for DHH users to address RQ1–RQ3. In Section IV, we examine wake-up and feedback methods for DHH users to address RQ4–RQ6. In Section V, we propose design guidelines for conversational NUIs for DHH users based on our investigation of these six research questions. In Section VI, we describe the limitations of this work and suggest some avenues for future research. Finally, in Section VII, we provide some concluding remarks.

II. RELATED WORK

To demonstrate the effectiveness of sign language-based conversational NUIs, we focus on two other NUIs that allow human-computer conversations: smart speakers and gesture interfaces. Then, we describe the Wizard of Oz method used to construct our experimental environment.

A. Diffusion Status and Challenges of Smart Speaker

In the latter half of the 2010s, with the improvement of voice recognition performance and AI technology, various companies released smart speakers to be used in the home [34]. Conversational AI agents that play music, news, and weather forecasts in response to natural language questions from the user have become widely popular. Using these devices, a user can perform a task by speaking a wake word into a nearby microphone or speaker unit and then continuing with a question or request. The reasons for the widespread use of smart speakers include improvements in voice recognition and natural conversation technology as well as their hands-free operation, which allows users to avoid interrupting other activities [35]. According to a Canalys survey, worldwide shipments of smart speakers were expected to reach 320 million units in 2020 and 640 million units by 2024 [36].

However, smart speakers are limited in that they cannot be used in offices or public spaces and are difficult to understand in noisy environments. The former is the case because hearing is a passive and unconscious stimulus compared to vision [37]. Therefore, as long as others are nearby, the sound of a voice may be heard in any direction. Therefore, based on sign language conversation, NUIs are expected to solve the privacy problem in offices and public spaces. The latter does not require speech recognition, and thus does not cause recognition problems in noisy environments.

B. Application Examples and Challenges of Gestural Interfaces

Gestural interfaces use visual and bodily functions such as arms, fingers, and facial expressions. Examples of applications include Motion Sense on the Google Pixel 4 and several such methods to control drones. Motion Sense allows users to operate their phones by holding their hands over them without directly touching them, such as moving forward or backward in list of tracks playing music or stopping incoming calls or alarms. It is also equipped with a motion-sensing function that lights up the screen when the user places their hand near the phone. In combination with face recognition, users can quickly unlock the phone.

In this context, there are three challenges with gestural interfaces. The first is the "cognitive load" [39] [40] [41]. As gesture input is not linguistic, it is necessary to memorize commands, as in a command-line interface (CLI). Although there is a possibility that the object or situation can be suggested to some extent compared by the CLI, the more it relies on gesture operations, the more gestures the user must remember. Therefore, the gestures used as commands tend to be inconsistent. The second is "distinction of intention" [42]. One of the selling points of gesture manipulation is the intuitive and natural behavior that humans often perform. However, this intuition and naturalness overlap with our daily behavior. The problem arises when a gesture is mistakenly recognized as a command and executed without distinguishing between the actual operation and the everyday behavior. The third problem is "physical fatigue" [43]. The human body may be burdened by the movements used to perform gestures, such as moving the arm up and down, left and right, or raising the arm for a long time.

Therefore, NUIs based on sign language conversations are expected to overcome the challenges of gesture UIs, such as reduced expressive power and increased memory load, because they use natural languages for interaction. It is also suggested that for native signers, conversing in sign language is independent of the physical fatigue issues inherent in gestural interfaces.

C. Wizard of Oz

The recognition rate of a real-life continuous sign language recognition system developed in 2019 was 39.6% [44]. Therefore, it was impossible to conduct experiments incorporating sign language recognition technology to interact with a user interface using sign language. We used the Wizard of Oz method is a solution to this problem [45] [46]. With this method, a human referred to as a wizard pretends to act as a computational system and interacts with the user. In the Wizard of Oz method, even if the entire system is not yet

TABLE I
LINGUISTIC MODALITIES OF HEARING AND DEAF INTERACTIONS WITH CUIS. "*" :RQ1,"**" :RQ2,"***" :RQ3

Type / User	Conversation		Call	
	System→User	User → System	System→ User	User →System
Hearing	Voice		Voice	
DHH	Sign Language/Text **		Luminous *	Eye Gaze ***

"*" :RQ1, "**" :RQ2, "***" :RQ3.

complete, the wizard can complement the undeveloped parts of the system and allow it to seem to perform these functions for the users. In this study, we use this approach to experiment with the behavior of a sign language recognition system to extract data generated in situations where people are speaking with sign language.

III. SIGN LANGUAGE CONVERSATIONAL USER INTERFACES USING LUMINOUS NOTIFICATION AND GAZE DIRECTION

A. Research Questions

In Section III, we consider the research questions RQ1-3. As shown in Table I, RQ1–RQ3 address the mutual input and output modalities that DHH users may prefer to achieve when interacting with CUIs. DHH users use sign language/text modalities when interacting with CUIs (RQ2). They use the luminous notification modality when calling from CUIs to DHH users (RQ1). In contrast, when DHH users call to (wake-up) a CUI (RQ3), they use the gaze modality. We investigate whether using these mutual input/output modalities in CUIs can improve the user experience of DHH users. To answer RQ1 to RQ3, we constructed a sign language conversation system with optical notifications based on the Wizard of Oz method and conducted an experimental evaluation (Experiment 1).

B. Methodology

1) Participants: Using a mailing list, we solicited the cooperation of DHH students in their 20s to participate in the Experiment 1. 12 students ultimate participated.

We also investigated the characteristics of the participants to analyze the effect of their attributes on the results of the experiment. Specifically, we conducted a preliminary questionnaire survey on the age, gender, and cochlear implant/hearing aid use of the participants to determine whether they use their voices when communicating, whether they mainly use auditory or visual communication, and whether they use both, as well as their experience with learning sign language and their experience using VUIs. Figure 3 shows the results. The age of our participants, 8 males and 4 females, ranged between 20 and 24. We asked the participants to rate their experiences with using VUIs on a 4-point Likert scale (1 = usually, 2 = sometimes, 3 = rarely, and 4 = never). The results showed that the response of one participant was 2, four participants responded with 3, and seven participants

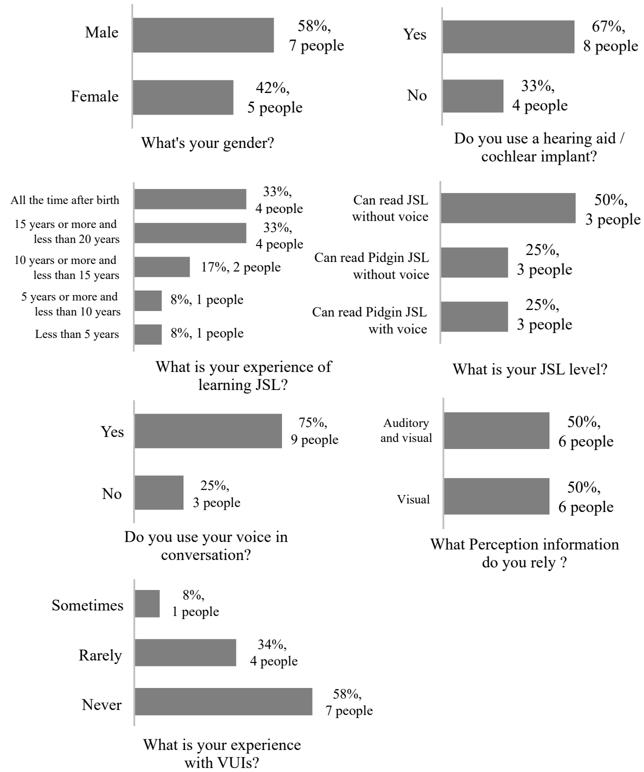

Figure 3. Characteristics of Participants (N=24) in Experiment 1.

responded with the value 4. The majority of the participants commented that "their voices could not be recognized" or that "they lived without speaking verbally." Research mentioning that a minimal number of DHH users use personal-assistant devices [47] indicates a similar trend to that exhibited by the participants in this experiment.

Experiment 2 was approved by the Research Ethics Review of the Tsukuba University of Technology, where the experiment was conducted. The duration of the Experiment 2 was 90 min, and the honorarium paid to the participants was 1,305 yen (approximately $12).

2) System Architecture: The basic configuration of the system constructed in Experiment 1 comprised an iPad, a Meross Smart Wi-Fi LED Bulb (LED Bulb), and a GoPro HERO9 camera. Figure 4 shows the appearance and operation of the system. We set four tasks that the system can perform: "Phone call," "Alarm settings and notifications," "Checking the Weather," and "Checking the News." An Apple iPad simulated Alexa, and the display was created using Microsoft PowerPoint 2019 and combined with the signer's video. To switch screens remotely, the remote function of Keynote was used. We included LED bulbs that could be set to any color (16 million RGB colors) and a blink cycle, and controlled the system remotely from a smartphone app. An LED bulb flashes yellow when the system notifies the user of a "Phone call" or "Alarm notification" and emits a light green when the system provides "Weather" or "News" to the user. In addition, the

Figure 4. System Prototype in Experiment 1.

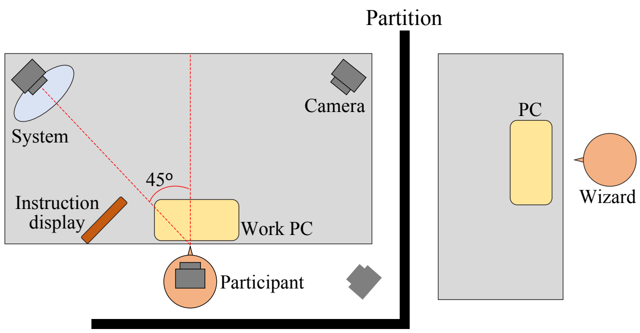

Figure 5. Experiment 1 Setup.

GoPro camera was used to view the sign language input from the user.

3) Procedure: Figure 5 shows the Experiment 1 environment. In the environment of Experiment 1, we assumed that the participants interacted with the system while working on their PCs. Therefore, we placed the system on the left side of the desk in front of the participants at 45 degrees, with a workstation PC in front of them. We tried to make the participants aware of the system response while the system was operating so that they did not have to constantly look at the system. We also aligned the system at the eye level of the participants. The instruction device prompted the participant to issue commands to the system at certain times. We incorporated a program in PsychoPy (v2021.1) [48] to display numbers and/or English letters at random positions on the screen. In addition, the frame rate of the installed camera was 50 FPS.

The critical points for the participants and the experimenter (Wizard) in this environment were as follows.

Participant

1) Owing to the nature of the Wizard of Oz method, the participants assumed that aa computational sign language recognition system was being used, and did not know that a person (wizard) was operating the system.
2) During the experiment, the participant were asked to continuously work on the task of "entering numbers and English letters displayed at random positions on the screen of a workstation PC with the keyboard as they appear."
3) The participants commanded the system using sign language commands for "Setting the alarm," "Checking the weather," and "Checking the news." The participant pressed the button as soon as they noted the end of the description of "Weather" or "News" from the system.
4) During the work, the participants used sign language commands to stop the "Alarm notification" and "Phone call" sent by the system.
5) During the experiment, the users wore a GoPro attached to a head strap mount.

Experimenter (Wizard)

1) Owing to the nature of the Wizard of Oz method, the experimenter was required to avoid letting the participants know that the experimenter is operating the system when performing the sign language recognition system.
2) During the experiment, the experimenter operated the system and the LED bulb.
3) When we asked the participants to conduct a specific task at an arbitrary time, we showed them the content of the task and an example of the command to be performed, and we immediately turned off the screen after confirming that the participants understood the task.

Before the experiment, we explained how to use the system and how the system was designed to behave for each of the four tasks. In addition, to familiarize the participants with command execution using sign language, we asked them participate in a practice session to perform a task equivalent to the real one before the actual experiment was conducted. The participants conducted each of the four tasks once and repeated them twice. To eliminate order effects, the order of the tasks for each participant and the two conditions, "Luminous/Conventional," were counterbalanced.

4) Analysis Method: For a time analysis using video, we applied the ELAN [49] tool.

For RQ1, we used the system usability scale (SUS) [50], a widely applied evaluation index for a quantitative evaluation of usability, to examine the usability of "Luminous" and "Conventional." In addition, we believe that improved usability is also related to awareness. To evaluate the awareness of the notifications from the system, we measured by video the time between the notifications were provided and when the participant noticed and reacted to them. We defined the reaction time for a "Call" as the time between the change in the display screen as the reaction starting point and the users turning their eyes to the screen as the reaction endpoint. However, if the light turned on before the screen changed, the reaction starting point was when the light turned on. We defined the reaction time for a "Response" as the time between the change in the display screen and the user pressing the button. However, if the light turned off before the screen changed, the reaction starting point was defined as the time when the light turned off.

For RQ2, we examined the participants' need for sign language/text. After the experiment, we administered a questionnaire to determine the need for sign language/text using a five-point Likert scale (1 = agree, 2 = agree a little, 3 = neutral, 4 = disagree a little, and 5 = disagree).

For RQ3, we examined whether the participants gazed at the system before giving a command in sign language. For this purpose, we measured the percentage of the total number of times the participants gazed at the system at least once in the 5 s before the sign language command and the time between the start of the gazing and sign language using video. For the data to be analyzed, there was a scene during the experiment in which the system responded to an "Alarm notification" or "Phone call," and the user issued a command to stop the system. The users looked at the response screen before making a sign language command, we did not collect analysis data to investigate whether they gazed at the screen before the sign language command. The three tasks for which the user actively gave a sign language command were used as data for analysis, i.e., "Setting the alarm," "Checking the weather," and "Checking the news."

C. Results

1) System Usability Scales: Figure 6 shows the results of the SUS investigated after the experiment. The mean SUS value of "Luminous" was 80.67 (SD 7.62), and that of "Conventional" was 68.96 (SD 14.6). As a result of the Wilcoxon signed-rank test, "Luminous" was found to be significantly higher ($p < .05$).

2) Reaction Time: Figure 7 shows the results of the reaction time. The mean reaction time to "Alarm notification" and "Phone call" of "Luminous" was 0.91 s (SD 0.35), and the mean value of "Conventional" was 1.19 s (SD 0.57). The Wilcoxon signed-rank test showed that the reaction time was significantly shorter for "Luminous" ($p < 0.01$). The mean reaction time to the end of "Weather" and "News" for "Luminous" was 1.37 s (SD 0.50) s, and the mean value of "Conventional" was 1.91 s (SD 1.22). The Wilcoxon signed-rank test showed no significant differences between "Luminous" and "Conventional" ($p > 0.05$).

3) Necessity of Sign Language/Text: Figure 8 shows the results of a 5-point Likert scale used to assess the need for sign language and text for the 12 participants, respectively. In terms of sign language, we found the following. 1: "Agree" was reported by four participants. 2: "Agree slightly" was reported by two participants. 3: "Neutral" was reported by three participants. 4: "Disagree slightly" was reported by two participants. 5: "Disagree" was reported by one participant. There were three participants who did not need sign language (4,5), and their sign language experience was, in order of shortest to longest, three years (1st), five years (2nd), and fifteen years (5th). By contrast, for text, "1 = Agree" was reported by nine of the participants, and "2 = Agree slightly" was reported by three of the participants.

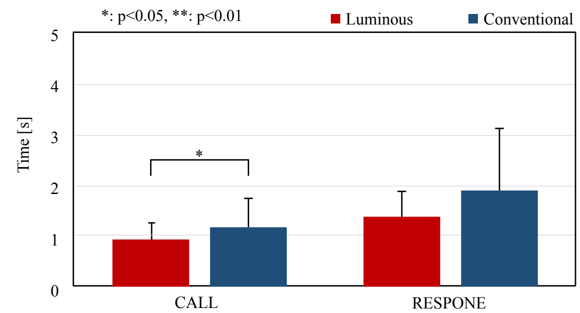

Figure 7. Reaction time for each feedback from the system.

4) Gaze Direction: During the experiment, a pattern occurred in which the experimenter turned off the screen of the instructional device late, indicating the task to be performed, and the participant gave a sign language command while reading. We removed these data from our analysis because they were unsuitable for examining whether the participants were gazing at the system. The participants (N = 12) input sign language commands into the system 69 times: 23 times for "Setting the alarm," 24 times for "Checking the weather," and 22 times for "Checking the news." Table II lists the percentage of the total number of times the participants gazed at the system at least once during the 5 s before the sign language command and the average time from the start of gazing to the start of the sign language, as well as the standard deviation and minimum and maximum values. A high percentage of the total number of users gazed at the system before using the sign language commands.

Three participants, P3, P8, and P9, waved before using sign language. These three participants had experience using

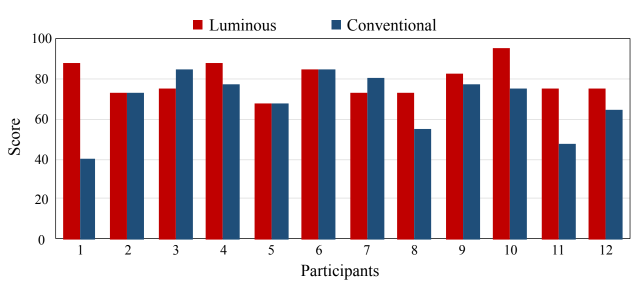

Figure 6. SUS score for each participant (N = 12).

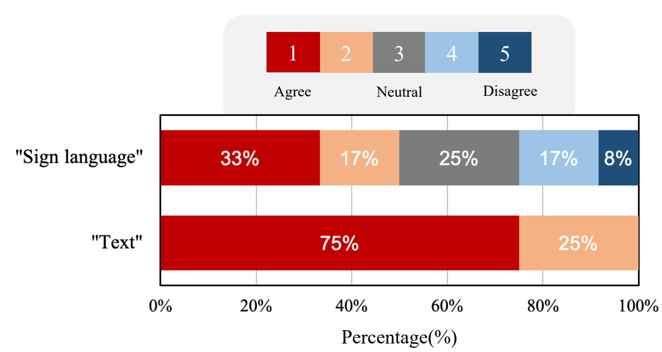

Figure 8. Necessity of "Sign language" / "Text".

TABLE II
PERCENTAGE OF EYE GAZE, MEAN AND STANDARD DEVIATION OF TIME OF EYE GAZE

Task	*Percentage* (%)	*Mean±SD* (*s*)	*Min* (*s*)	*Max* (*s*)
Alarm	100	0.76±0.61	0.20	3.18
Weather	100	0.43±0.23	0.10	1.08
News	86.4	0.59±0.44	0.20	2.08
Total	93.4	0.59±0.47	0.10	3.18

VUIs and knew that they should use a waking command. The interviews also revealed that they thought it was necessary to take explicit action before talking to the system during this experiment.

D. Discussion

1) RQ1: Efficacy of Luminous Notification: The results described in Section III-C1 suggest that luminous notification improved usability for DHH users in noticing notifications from the system. In addition, the reaction times to “Alarm notification” and “Phone call” were significantly shorter when using a luminous notification, suggesting that it is easier to notice such notifications from the system.

Participants commented, “I am familiar with luminous notification methods, such as the intercom system in my house, which notifies me by light, so it would be more impressive to add light to the system as well. I can notice the light notification even when I am concentrating on my work.” However, there were also comments such as “I feel uncomfortable with the luminous notification because I live my life relying on sound. Therefore, the system may not be suitable for people who use their daily hearing functions.

From Figure 6, we can see that the usability of P3 and P7 decreased with a luminous notification. The participants commented that they did not feel the need to use a luminous notification because they only noticed the change in the system screen. This may have occurred because there were cases in which DHH users could respond to conventional methods [51]. In this experimental environment, the system was placed on the left side of the desk in front of the user at a 45-degree angle within their peripheral vision. During this experiment, we placed the system within the peripheral vision of the front of the user, and thus some of the subjects noticed changes in the screen without looking at the system.

In contrast, there were no significant differences in the reaction time to the end of the “Weather” and “News” responses when using the luminous notification, as described in Section III-C2. However, some of the participants commented positively that “it was convenient to know when the response ended without having to look at the system.” By contrast, others commented negatively that “luminous notification was not necessary for the information (weather and news) that I wanted,” and “the light was too bright.” As a result, we found that the usability of the system could be improved by reducing the light exposure and improving the luminous notification method, although the noticeability remained the same.

A participant commented that it would be preferable to increase the brightness of the display, as in ON AIR, instead of directly informing the user with LED bulbs. For a luminous notification, we used LED bulbs, which were initially used as lighting fixtures. Therefore, a way to change or vary the brightness of the display directly should be considered instead of using external LEDs.

2) RQ2:How to Display Sign Language/Text Suitably: From Section III-C3, it may be observed that all of the participants needed to display text regardless of the user attributes. By contrast, the necessity for the sign language display varied from participant to participant. In addition, it may be observed that those who had not signed for a long time tended not to believe that sign language was necessary.

The participants who did not need sign language commented that they did not understand sign language and had trouble processing information when both sign language and text were output simultaneously. By contrast, the participants who needed to use sign language commented that the sign language display made it easier for them to remember the system responses. Some of the participants commented, “When I look at the task screen while working, the remaining text is better than the flow of the sign language.”

For hearing users, interaction with VUIs has the advantage of being eyes-free [52]. Therefore, users frequently interact with conversational NUIs while performing other tasks. However, in the case of DHH users, the advantage of eyes-free interaction is lost because they cannot acquire audio information and instead gaze at the screen. To complement this, we anticipate that DHH users would need text information that they can recognize, even if they look away for a moment. One possible solution to this problem is to stop the sign language flow when the user looks away and resume when the user returns their gaze to the screen.

3) RQ3:Efficacy of Gaze Direction: Section III-C4 shows that the participants tended to gaze at the screen before speaking in sign language.

During this experiment, we did not provide instructions on how to wake-up the device. Nevertheless, the participants naturally gazed at the system with a high probability.

By contrast, 3 of the 12 participants did not gaze at the system but waved instead. When DHH users use waving as a wake-up method, there is a concern that signs made while talking to another person may be recognized as waving at unexpected times, such as during a “Phone call” or “Alarm notification.” In addition, we believe that gazing is a more natural manner of interacting than waving each time a command is used. These results suggest that directing one’s gaze to a device may be considered a compelling wake-up method.

When Alexa waits for a response from the user, there is a time limit of 8.0 s [53]. From Table II, the maximum time between gazing at the system and the start of sign language was 3.18 s. That is, when DHH users used gazing as a wake-up method, they could use commands within the system’s waiting time.

IV. CONSIDERATION OF WAKE-UP METHOD AND FEEDBACK METHOD

A. Research Questions

In Section IV, we consider research questions RQ4–RQ6, which cover the input and output methods for DHH users at the beginning of a conversation with the conversational NUIs. First, we investigated the need for feedback on the system for DHH users (RQ4). Then, we investigated the best method to wake up the system from DHH users (RQ5). We investigated the best way to present the feedback to the user to indicate that the system has detected the user's wake-up command and is ready to accept commands (RQ6). To answer RQ4–RQ6, we constructed a conversational system that presented a variety of feedback based on the Wizard of Oz method and conducted an evaluation experiment (Experiment 2).

B. METHODOLOGY

1) Participants: Using a mailing list, we solicited the cooperation of DHH students in their 20s to participate in Experiment 2. 24 students participated.

We also investigated the characteristics of the participants to analyze the effect of their attributes on the results of the experiment. Specifically, we conducted a preliminary questionnaire survey on the age, gender, and cochlear implant/hearing aid use of the participants to determine whether they used their voices when communicating, whether they used both, and their experience of learning sign language, their identity, and their experience using VUIs. Figure 9 shows the results. The age of our participants, 14 males and 10 females, ranged between 20 and 23. We asked the participants to rate their experience of using VUIs on a 5-point Likert scale (1 = everyday, 2 = several times a week, 3 = several times a month, 4 = less than once a month, and 5 = never). As a result, the most significant number of 20 participants answered "5:Never," accounting for 83% of the total. Research mentioning that a minimal number of DHH users use personal-assistant devices [47] has indicated a similar trend to that of the participants in this experiment.

Experiment 2 was approved by the Research Ethics Review of the Tsukuba University of Technology, where the experiment was conducted. The duration of Experiment 2 was 90 min, and the honorarium paid to the participants was 1,500 yen (approximately $13).

2) Experimental conditions: **Wake-up**
We compared the gaze-based method validated in Section IV with other wake-up conditions and identify the preferences of DHH users.
I1: Eye Gaze
In the reports of studies on wake-up commands for DHH users, signing a name was rated as the highest condition. However, the authors did not consider gaze direction in their comparison. Therefore, in Experiment 2, we investigate whether adding gaze tracking to the wake-up method changed the preferences of DHH users.
I2: Sign-name
Assuming that the system installed was Alexa, and referring to the previous research, the method of signing a name involved using an "A" handshape used to draw an "X."
I3: Waving
A wake-up wave is a left-right-forth motion in which the user holds their palm toward the system. According to reports on wake-up research [30], the evaluation of the wake-up was lowered due to the concern that the system would recognize unintentional waving as a wake-up gesture. Therefore, we thought we could gain new insights by comparing the results with the gaze-based system, in which the user's intention to wake up the system is clearer.

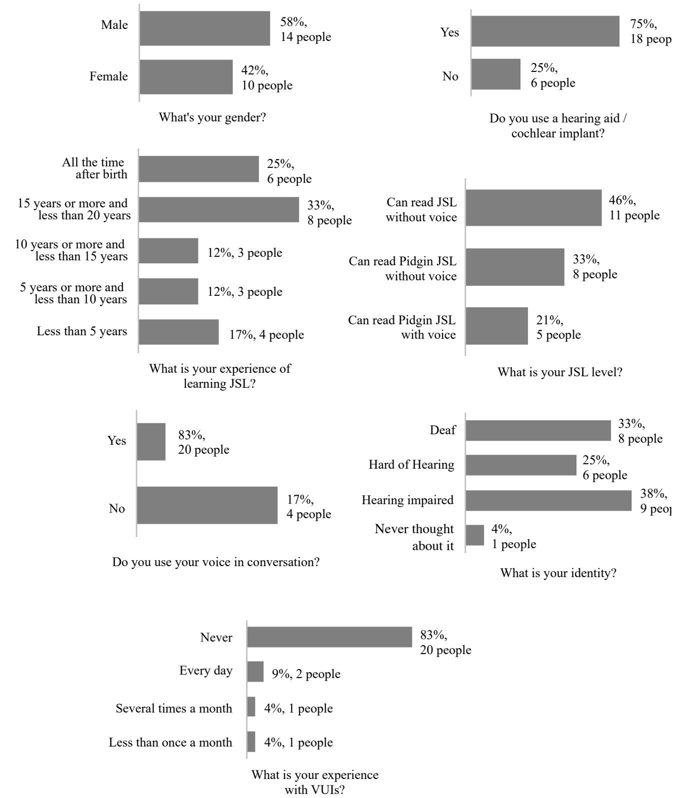

Figure 9. Characteristics of Participants (N=24) in Experiment 2.

Feedback method
O1: Blue bar and low-intensity
The Echo Show recognizes a wake word and displays a blue bar at the bottom of the screen. The screen's brightness is lowered to make the blue bar stand out to show that the device is ready to process the user's request [54]. We added a baseline condition to explore whether these conventional response presentation methods are desirable for DHH users.
O2: Sign language added to O1
As DHH users mainly use sign language in their conversations, we believe it would be optimal for DHH users to have the system respond in sign language. For this reason, we added a sign language response.
O3: Shaking head motion added to O1
To investigate whether a head-shaking motion, like the Echo

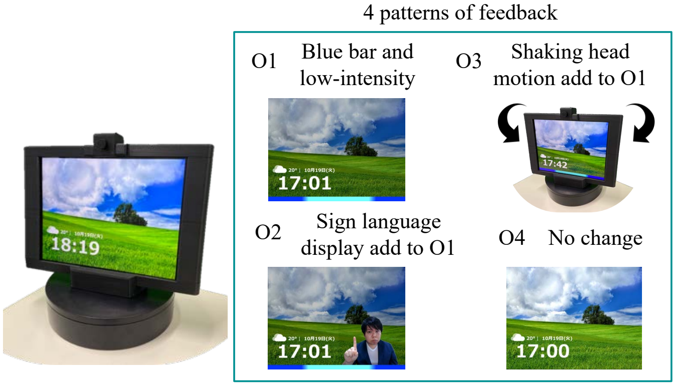

Figure 10. System Prototype in Experiment 2.

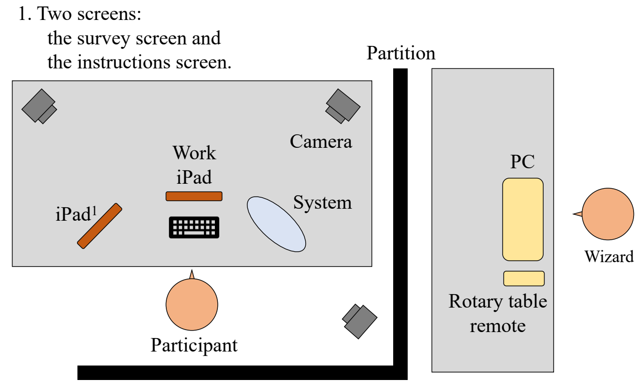

Figure 11. Experiment 2 Setup.

Show10 [55], could improve usability by DHH users, we added a response with a head-shaking motion.
O4: No change
To investigate the need for feedback for DHH users of RQ4, we added a condition of not presenting feedback, indicating no change to the system, in contrast to O1–O3.

3) System Architecture: The basic configuration of the system constructed in Experiment 2 comprised an iPad and an electric rotary table. Figure 10 shows the appearance and operation of the system. The screen size of the iPad was 10.2 inches, and that of the Echo Show10 was 10.1 inches, which was almost identical. For the system's design, Echo Show 10 was simulated, the display was created with Microsoft PowerPoint 2019, and the cover around the iPad display was created with a FlashForge Adventure 3 FFA-103 3D printer. In addition, the remote function of Keynote was used to switch the screen remotely. For the sign language display, we synthesized a sign language movie called "What's up?". To add the function of the head-shaking motion, we used an electric rotary table with a diameter of 20 cm and a load capacity of 20 kg manufactured by BAOSHISHAN. A remote controller was used to remotely control the rotation speed and direction angle of the table.

4) Procedure: Figure 11 shows the Experiment 2 environment. In the Experiment 2 environment, we assumed that the participants interacted with the system while working. Therefore, we placed the system on the right side of the desk in front of the participants at a 45-degree angle, the work iPad in front of them, and the iPad with the questionnaire and instruction screens split on the left side of the desk in front of the participants at a 45-degree angle. Question on user satisfaction for both the wake-up and feedback conditions, the ranking of satisfaction for each, and the need for feedback were created in Forms using a 7-point Likert scale. The instruction screen was created in Keynote to show the wake-up procedure to be performed by the participant and the feedback to be provided by the system. In addition, the frame rate of the installed camera was 50 fps.

The critical points for the participants and the experimenter (Wizard) in this environment are summarized as follows.

Participant

1) Owing to the nature of the Wizard of Oz method, the participants assumed that a sign language recognition system was in use, and did not know that a person (wizard) was operating the system.
2) During the experiment, the participants were asked to continuously work on the task of "entering numbers and English letters displayed at random positions on the screen of a workstation PC with the keyboard as they appear."
3) The participants confirmed the wake-up procedure and the feedback provided by the system on the instruction screen.
4) After the participant typed about three words, they woke up the system, and as soon as the system feedback was presented, they selected one of the commands, "Check weather" or "Check news", and commanded the system in sign language.
5) After confirming that the system had finished answering the "weather" and "news" questions, the user completed the "satisfaction with both the wake-up method and the system feedback" questionnaire.

Experimenter (Wizard)

1) Owing to the nature of the Wizard of Oz method, the experimenter was required to avoid letting the participants know that the experimenter was operating the system when performing the sign language recognition system.
2) During the experiment, the experimenter controlled the system.
3) When we asked the participants to conduct a specific task at an arbitrary time, we showed them the content of the task and an example of the command to be performed, and we immediately turned off the screen after confirming that the participants understood the task.

Before the experiment, we explained to the participants how to use the system and how the system behaved. In addition, to familiarize the participants with wake-up and command execution using sign language, we provided a practice session in which they performed a task equivalent to the real one before the actual experiment was conducted. Participants performed each of the 12 conditions once, for 12 repetitions, combining the three wake-up conditions and the four feedback conditions. The order of the 12 conditions was determined using a Latin square design to eliminate order effects.

5) Analysis Method: To investigate the optimal combination conditions from the three conditions of the wake-up method (I1–I3) and the four conditions of the feedback method (O1–O4), we used a questionnaire with a Likert scale of 7 levels of satisfaction (3. very satisfied, 2. satisfied, 1. slightly satisfied, 0. neither satisfied nor dissatisfied, -1. slightly dissatisfied, -2. dissatisfied, -3. very dissatisfied). We asked the participants to respond to a questionnaire based on the Likert scale. From the response data (N=24), better wake-up conditions and feedback conditions were clarified. After completing the 12 conditions, the participants were asked to rank their satisfaction with the wake-up method (I1–I3) and the feedback method (O1–O4). The reasons for their satisfaction were investigated in an interview. To evaluate the effectiveness of the system in conversations with the participants, we conducted a video analysis of their behavior using the ELAN [49] tool. The minimum video measurement time was 0.02 s.

1) Time between the beginning and end of the wake-up action
2) Duration of gaze during waving
3) Time from the end of the wake-up operation to the beginning of command input

Then, to investigate the necessity for the presentation of feedback in RQ4, we asked the participants to respond to a questionnaire with seven levels of necessity (1. strongly agree, 2. agree, 3. agree a little, 4. neutral, 5. disagree a little, 6. disagree, 7. strongly disagree). The questionnaire was administered using a 7-point Likert scale. To investigate the effectiveness of gazing in the wake-up method of RQ3, we asked the participants whether they would like to perform I1 (eye gaze) together with I2 (sign name) and I3 (waving) in the wake-up method. We categorized the participants' responses into three patterns (1. like, 2. limited like, 3. dislike).

C. Results

1) Effect of the feedback: Figure 12 shows a summary of the mean and standard deviation of satisfaction of each feedback method (O1–O4). The mean satisfaction values were 1.90 (SD 1.16) in O1, 1.67 (SD 1.21) in O2, 1.67 (SD 1.55) in O3, and -0.18 (SD 1.64) in O4. Multiple comparisons using the Tukey's test were conducted to determine whether there was a significant difference in satisfaction between the four levels (O1–O4). The results showed that the level of satisfaction in O4 (no change) was significantly lower than that in the other feedback presentations (O1–O3, change) ($p < 0.01$).

Figure 13 shows a summary of the mean and standard deviation of the time from the end of wake-up to the beginning of command input for each feedback presentation condition (O1–O4). The times of the mean and standard deviation were 2.60 s (SD 1.06) for O1, 2.92 s (SD 1.31) for O2,

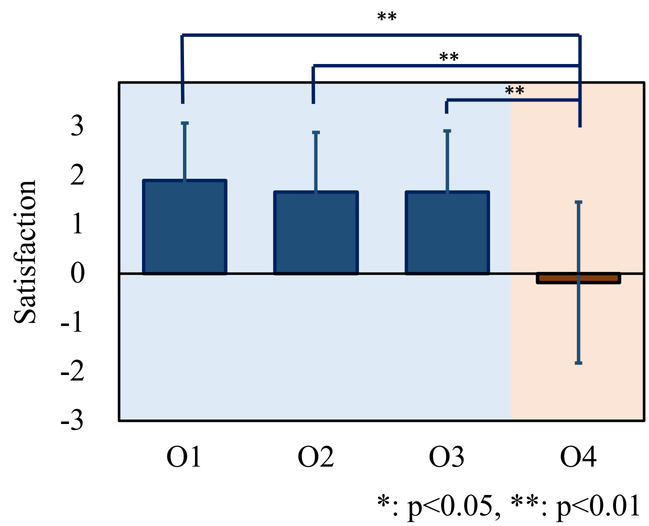

Figure 12. Mean and standard deviation of individual satisfaction with feedback.

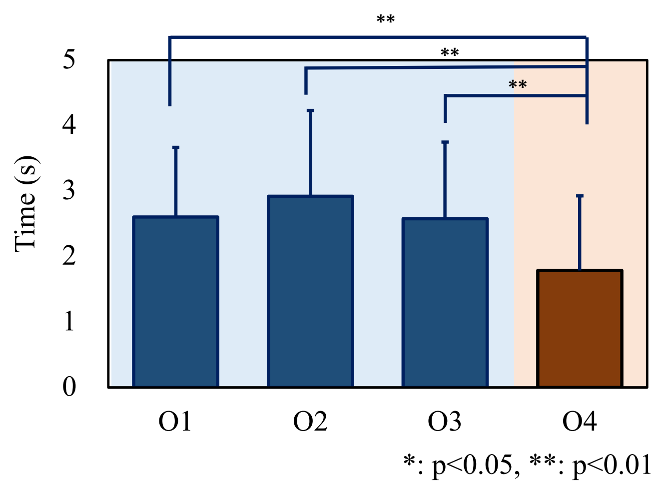

Figure 13. Mean and standard deviation of time from the end of wake-up to the beginning of command input for each feedback.

2.57 s (SD 1.17) for O3, and 1.79 s (SD 1.14) for O4. We conducted multiple comparisons using Tukey's test to check for a significant difference in time between the four levels (O1–O4). The results showed that the time in O4 (no change) was significantly shorter than the time in the other feedback presentation (O1–O3, with change) ($p < 0.01$).

2) Satisfaction: Figure 14 shows a summary of the mean and standard deviation of the satisfaction for each of the nine conditions combining the wake-up and the feedback methods. The highest mean satisfaction of the input and output methods was 2.13 (SD 1.08) for the combination of I3 (waving) and O1 (blue bar and low-intensity). The lowest was 1.21 (SD 1.41)

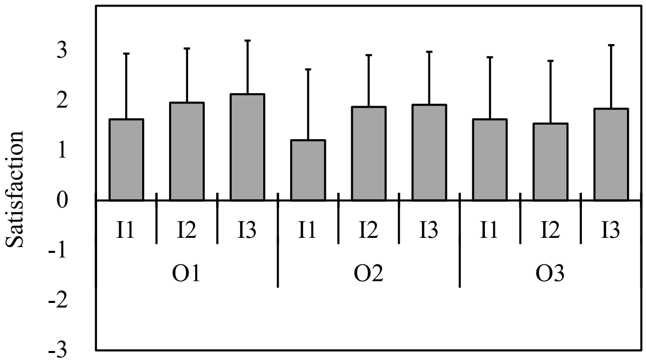

Figure 14. Satisfaction for each 12 conditions(N=24).

for the combination of I1 (eye gaze) and O2 (sign language add to O1) (SD 1.41).

We conducted a two-way analysis of variance (ANOVA) using the wake-up method (I1–I3) and the feedback method (O1–O3) as factors. The results showed that the main effect for satisfaction with the wake-up method was significant ($p < 0.05$), but the main effect for satisfaction with the feedback method was not significant ($p > 0.05$). In addition, the interaction between the satisfaction of the wake-up method and the feedback method was not significant ($p > 0.05$). For the wake-up method for which the main effect of satisfaction was significant, the mean satisfaction values were 1.49 (SD 1.32) in I1 (gazing), 1.79 (SD 1.13) in I2 (signing a name), and 1.96 (SD 1.13) in I3 (waving). Multiple comparisons of the Tukey's test were conducted to examine the difference in the mean satisfaction values among the three levels of the wake-up method. The results showed that the satisfaction level was significantly higher for I3 (waving) than I1 (gazing) ($p < 0.05$).

We conducted a two-way ANOVA of satisfaction using the wake-up method (I1–I3) and the feedback method (O1–O3) as factors for each characteristic. Table III shows a summary of the significance probabilities of the wake-up method, feedback method, and interaction for each characteristic. Four attributes were analyzed: "Whether the user was using hearing aids or cochlear implants," "Japanese sign language (JSL) level," "Whether the user was using voice in the conversation," and "Identity." The interaction was not significant for all characteristics ($p > 0.05$). The main effect characteristic for "other using aids/cochlear implants" was the wake-up method for participants who answered "no." These participants had the characteristic of "not relying on auditory information." The mean satisfaction values were 0.78 (SD 1.40) in I1 (gaze), 1.83 (SD 0.99) in I2 (sign name), and 2.06 (SD 1.00) in I3 (waving). We conducted multiple comparisons of the Tukey's test to examine the difference in the mean satisfaction values among the three levels of the wake-up method. As a result, we found that the satisfaction level of I2 (sign name) was significantly higher than that of I1 (gazing) ($p < 0.05$), and the satisfaction level of I3 (waving) was significantly

TABLE III
RESULTS OF ANALYSIS OF VARIANCE FOR TWO FACTORS OF SATISFACTION WITH WAKE-UP METHOD AND FEEDBACK PRESENTATION METHOD AS INDEPENDENT VARIABLES BY CHARACTERISTICS.

		Significance probability (p-value)		
Attributes	Characteristics	Wake-up (I1-I3)	Feedback (O1-O3)	Interaction
Whether using aids/cochlear implants	Yes (N=14)	0.63	0.81	0.67
	No (N=10)	4.45×10^{-3} **	0.22	0.98
Reading JSL level	JSL without voice (N=11)	1.20×10^{-3} **	0.12	0.82
	Pidgin JSL without voice (N=8)	0.80	0.68	0.92
	Pidgin JSL with voice (N=5)	0.71	0.50	0.83
Whether using voice in the conversation	Using voice (N=20)	0.38	0.63	0.61
	Not at all (N=4)	6.43×10^{-3} **	0.33	0.87
Identity	Deaf (N=8)	1.82×10^{-2} *	1.82×10^{-2} *	0.91
	Hard of hearing (N=6)	0.64	0.82	0.95
	Hearing impaired (N=9)	0.53	0.79	0.90

*: $p < 0.05$, **: $p < 0.01$

higher than that of I1 (gaze) ($p < 0.01$). For the "JSL level," the characteristic that showed the main effect was the wake-up method for participants who answered, "I can read JSL without voice." These participants had a "high sign language level" characteristic. The mean satisfaction scores were 1.18 (SD 1.38) in I1 (gazing), 1.94 (SD 0.97) in I2 (signing a name), and 2.21 (SD 1.02) in I3 (waving). Multiple comparisons of the Tukey's test were conducted to examine the difference in the mean satisfaction values among the three levels of the wake-up method. As a result, we found that the satisfaction level was significantly higher for I2 (sign name) than for I1 (gaze) ($p < 0.05$), and the satisfaction level was significantly higher for I3 (waving) than for I1 (gaze) ($p < 0.01$). Concerning "Whether the user was using voice in conversation," the characteristic that showed the main effect for participants who answered "No voice," was the wake-up method. This participant had the characteristic of "not using voice information." The mean satisfaction scores were 0.25 (SD 1.22) for I1 (gazing), 1.50 (SD 0.90) for I2 (sign name), and 1.67 (SD 0.98) for I3 (waving). To examine the difference in the mean level of satisfaction among the three groups of the wake-up method, multiple comparisons using the Tukey's test were conducted. The results showed that the satisfaction level was significantly higher for I2 (signing name) and I3 (waving) than for I1 (gazing) ($p < 0.05$). For the "Identity," the main effect characteristics were the wake-up method and the feedback method within the participants who answered "Deaf." These participants were raised in a school for the

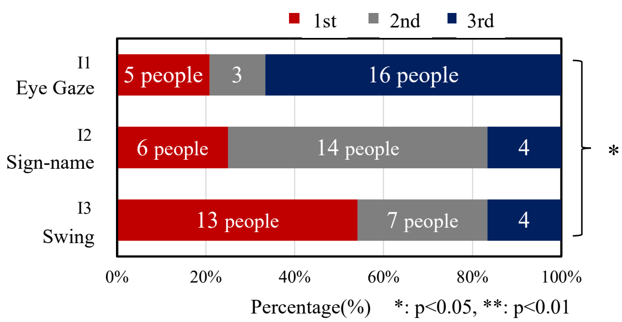

Figure 15. Ranking result of wake-up(N=24).

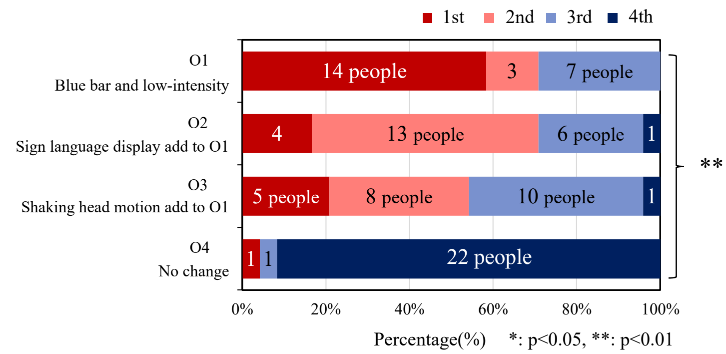

Figure 16. Ranking result of feedback (N=24).

Deaf, came from a Deaf family, and belonged to the Deaf community. For the wake-up method, the mean satisfaction level was 1.00 (SD 1.38) for I1 (gaze), 1.67 (SD 1.17) for I2 (sign name), and 1.96 (SD 1.00) for I3 (waving). Multiple comparisons of the Tukey's test were conducted to examine the differences in the mean satisfaction values among the three levels of the wake-up method. The results showed that the satisfaction level of I3 (waving) was significantly higher than that of I1 (gazing) ($p < 0.05$). For the feedback methods, the mean satisfaction values were 2.08 (SD 1.10) for O1 (blue bar and low-intensity), 1.42 (SD 1.10) for O2 (sign language added to O1), and 1.13 (SD 1.36) for O3 (head-shaking motion added to O1). Multiple comparisons of the Tukey's test were performed to examine the differences in the mean values of satisfaction among the three levels of feedback methods. The results showed that O1 (blue bar and low-intensity) was significantly more satisfactory than O3 (head-shaking motion added to O1) ($p < 0.05$).

3) Ranking: **Ranking results by overall participants** Figure 15 shows the results of the ranking done by the participants (N=24) for each wake-up condition (I1–I3). The mean rank was 2.46 (SD 0.82) for I1 (gazing), 1.92 (SD 0.64) for I2 (signing name), and 1.63 (SD 0.75) for I3 (waving). From these results, the Friedman test was used to investigate whether the mean rank varied with different wake-up methods, and the results showed a significant difference ($p < 0.05$).

For I1 (gazing), the participants who ranked it first mainly commented, "It was an easy calling action because it was hands-free," "It was the same feeling as using face recognition on a smartphone," "I was concerned that the system would recognize the user when they glanced at the system unintentionally. Therefore, it might be better to design the system to respond only after two seconds." In contrast, participants who ranked it third mainly commented, "I felt uncomfortable when I gazed at the screen when calling out," "I thought it would be better to call out using hands from the beginning if I gave a command in sign language afterward because I was unsure if the system would recognize me if I simply gazed at it. So it might be better to set a two-second rule to let the system react to gaze."

For I2 (sign name), the participants who ranked it first mainly commented, "I thought that the system would easily recognize the name-signing operation because it is a system-specific name," "Because it is a system-specific name and the system can easily recognize the name-signing operation, I felt comfortable with signing the name," "I feel closer to the system because it uses names." In contrast, the participants who ranked it third mainly commented, "I felt uncomfortable because I rarely call people by their names in my daily life," "I am not used to sign language, so I am not used to calling people by their sign name," "Compared to gazing and waving, I think that sign names are challenging to recognize."

For I3 (waving), the participants who ranked it first mainly said, "Because I can perform the action immediately," "Because it is the same action as calling out in daily conversation," "It is easy to call out even for the first time." In contrast, the participant who ranked it third commented, "Because the target of the name is not so clear compared to a sign name," "Because there was a possibility that my friends around me would misunderstand if I called out to the system with waving."

Figure 16 shows the results of the ranking done by the participants (N=24) for each feedback condition (O1–O4). The mean rankings were 1.71 (SD 0.91) in O1 (blue bar and low-intensity), 2.17 (SD 0.76) in O2 (sign language add to O1), 2.29 (SD 0.86) in O3 (shaking head motion add to O1), and 3.83 (SD 0.64) in O4 (no change) condition. (SD 0.64) in O4 (No change). We investigated the mean rank change with different feedback methods using the Friedman test from these results. We found a significant difference ($p < 0.01$).

For O1 (blue bar and low-intensity), the participants who ranked it first mainly commented, "I could immediately see that the system responded to my call," "The response was similar to Siri on my smartphone, so it was easy to get used to," "I felt that the other reactions were too much, and the blue bar was enough for me." In addition, there was a low evaluation comment, "It was too simple and difficult to understand compared to the sign language display and the head-shaking motion."

For O2 (sign language added to O1), the participants who ranked it first mainly commented, "I wanted the other person to respond in sign language because I spoke to them in sign

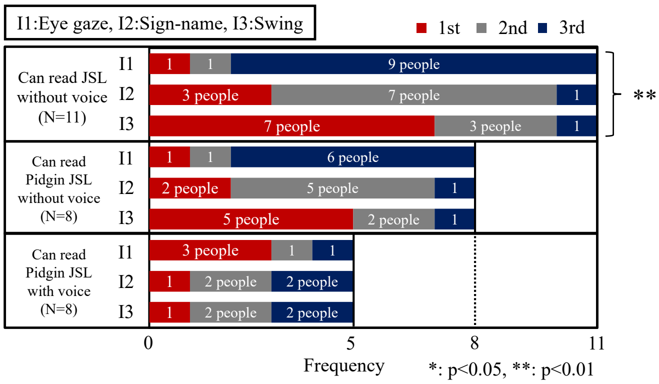

Figure 17. Ranking results for Wake-Up divided by JSL level attributes.

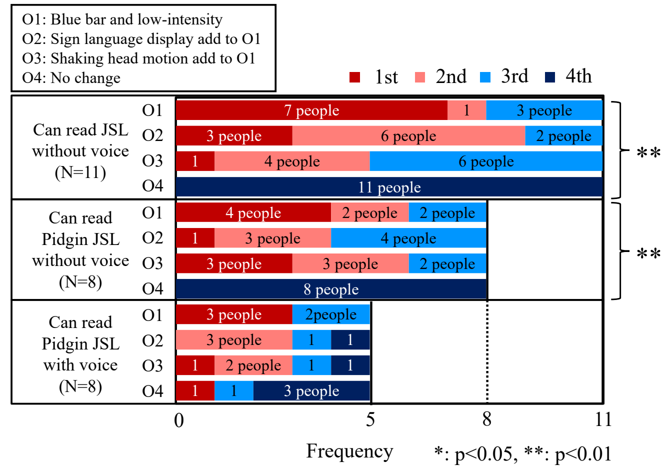

Figure 18. Ranking results for feedback divided by JSL level attributes.

language," "I felt relieved that the other person responded in sign language" "It was easy to understand their facial expressions and what they were asking." In contrast, there were comments such as "It would have been better if the system had used an avatar instead of a human," "I felt that it was sufficient to display text information instead of sign language," "It would be better if the system displayed the signer as an avatar instead of a human being, and if the signer was displayed from the default state, such as the home screen, instead of appearing only when I wake-up."

For O3 (shaking head motion added to O1), the participants who ranked it first mainly commented, "I could see that the system worked and responded clearly, and it gave me a cute impression," "It was straightforward to understand that the system responded to the shaking head." In contrast, there were comments on improving the head-shaking motion, such as, "I felt uncomfortable when the system turned to the outside. To wake-up the system without gazing, we would have preferred vertical to horizontal waving," "It was good in terms of visibility, but I was concerned that it would become difficult to see if more text information was added," "The head-shaking motion takes a little time, so the evaluation is lowered."

For O4 (no change), the participants who ranked it first commented, "No change was better because I can say what I want to say without pausing from the calling motion." In contrast, there were many comments with low evaluation, such as "When should I issue a command?," "I felt uneasy because I did not know when to give the command," "I want visible changes."

Ranking on sign language level characteristics Figure 17 summarizes the ranking data of the wake-up method for the sign language level attributes of the participants. We performed the Friedman test, and then we analyzed whether there was a significant difference in the level of satisfaction for each of the wake-up conditions (I1–I3) for each attribute. As a result, the only participant characteristic that showed a significant difference in satisfaction was the participants who answered "I can read JSL without voice" ($p < 0.01$). There was no significant difference in the level of satisfaction among the participants who answered "I can read Pidgin JSL without voice," "I can read Pidgin JSL with voice"($p > 0.05$).

Figure 17 summarizes the ranking data of the feedback method for the sign language level attributes of the participants. We performed the Friedman test, and then we analyzed whether there was a significant difference in the level of satisfaction for each feedback condition (O1–O4) for each attribute. As a result, the participant characteristics that showed significant differences in satisfaction were those who answered "I can read JSL without voice" ($p < 0.01$) and those who answered "I can read Pidgin JSL without voice" ($p < 0.01$). There was no significant difference in the level of satisfaction among the participants who answered "I can read Pidgin JSL with voice" ($p > 0.05$).

4) Participant behavior: **Time between the beginning and end of the wake-up action** We conducted multiple comparisons using the Tukey's test to see significant differences in the mean values of waving in each feedback condition (O1–O3). As a result, no significant difference was found ($p > 0.05$). Table IV summarizes the characteristics that reduced the time of waving compared to other attributes for each attribute of the participants. We analyzed four attributes: "whether the participant had hearing aids or cochlear implants," "JSL level,"

TABLE IV
PARTICIPANT CHARACTERISTICS FOR WHICH THE TIME OF WAVING WAS SIGNIFICANTLY SHORTER THAN THE OTHER CHARACTERISTICS.

Do a hearing aid/cochlear implant?	What is your JSL level?	Do you use your voice in conversation?	What is your identity?
Yes (N=14)	Can read JSL without voice * (N=11)	Yes (N=20)	Deaf * (N=8)
No (N=10)	Can read Pidgin JSL without voice (N=8)	No * (N=4)	Hard of Hearing (N=6)
	Can read Pidgin JSL with voice (N=5)		Hearing impaired (N=9)

* : Characteristics with significantly shorter swing time compared to other characteristics (p < 0.05)

TABLE V
COMPONENT RATIO AND DIFFERENCE IN START TIME FOR EACH OF EYE GAZE AND WAVING

First behavior	Component ratio %	Difference in start time (s)		
		Mean (SD)	Min	Max
Eye Gaze	81.9	0.37 (0.35)	0.04	1.60
Waving	15.3	0.30 (0.32)	0.04	1.12
Same	2.8	n/a	n/a	n/a

TABLE VI
TIME OF SIMULTANEOUS EYE GAZE AND WAVING BEHAVIOR

Behavior	Time (s)		
	Mean (SD)	Min	Max
Eye gaze and waving	1.07 (0.52)	0.20	2.46

"whether participant voiced in conversation," and "Identity." To compare whether there was a significant difference between the levels, we performed Tukey's test of multiple comparisons for "JSL level" and "Identity" (three levels), and Welch's t-test for multiple comparisons for "whether the participants used hearing aids or cochlear implants" and "whether the participants used voice in conversation" (two levels). We found no significant difference in time for the former. For the attribute "JSL level," the time required for waving was 0.99 s (SD 0.35) for participants who answered "I can read JSL without voice" and 0.99 s (SD 0.35) for participants who answered "I can read Pidgin JSL with voice." We observed that the time taken for the waving was significantly shorter at higher sign language levels ($p < 0.01$). For the attribute "Using voice in the conversation," the time for waving was 1.17 s (SD 0.52) for participants who answered "Not at all" and 0.87 s (SD 0.41) for those who answered "Using voice." The time required for waving was significantly shorter for those who did not use their voices ($p < 0.05$). For the "Identity" attribute, the waving time was 0.93 s (SD 0.35) for participants who answered "Deaf" and 1.26 s (SD 0.61) for participants who answered "hearing-impaired." The time was significantly shorter for participants with a "Deaf" identity than for those with a "hearing-impaired" identity ($p < 0.01$).

Eye gaze in waving Table V shows the ratio of gaze initiation and waving first, the time from gaze initiation to the beginning of the waving motion, and the time from starting a waving gesture to beginning to gaze at the display. For the remaining 2.8% of the data, the difference between the starting times of the two behaviors were within 0.02 s. Table VI shows the times when both gazing and waving were performed.
We conducted multiple comparisons of the Tukey's test determine whether the difference in feedback methods affected the time spent performing both gazing and waving (O1–O3). The results showed that there were no significant differences between the two methods.

The time from the end of the wake-up to the beginning of command input Figure 19 shows the mean values and standard deviations of the time from the end of the wake-up to the start of command input for each of the nine conditions combining the wake-up method and the feedback method. We

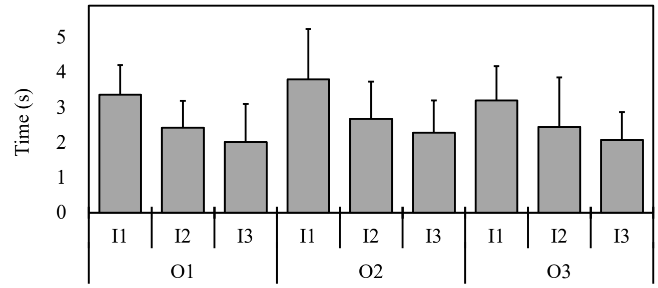

Figure 19. Mean of time from the end of wake-up to the start of command input for each 9 conditions.

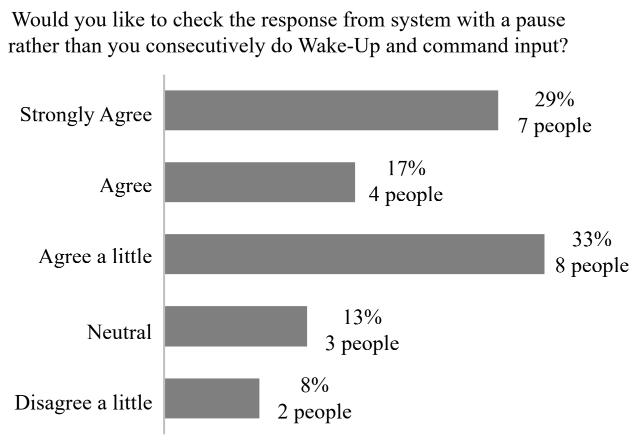

Figure 20. Necessity for feedback.

conducted a two-way ANOVA with the wake-up method (I1–I3) and feedback presentation method (O1–O3) as factors. The results showed that the main effect for the time of the wake-up method was significant ($p < 0.01$), but the main effect for the time of the feedback method was not significant ($p > 0.05$). In addition, the interaction of time for each of the wake-up and feedback methods was not significant ($p > 0.05$). For the wake-up method for which the main effect of time was significant, the mean time values were 3.45 s (SD 1.28) in I1 (eye gaze), 2.52 s (SD 1.20) in I2 (sign name), and 2.12 s (SD 1.20) in I3 (waving). To examine the difference in the mean satisfaction values among the three levels of the wake-up method, we conducted multiple comparisons using the Tukey's test. I3 (waving) was significantly shorter ($p < 0.01$) than I2 (sign name) ($p < 0.05$).

5) Subjective Evaluation: **Necessity for feedback presentation** Figure 20 shows the results of whether the participant preferred to check the response from the system with a pause rather than consecutively performing wake-up and command input. Participants who answered "1. strongly Agree" mainly said, "I can say what I want to say without anxiety if I can check system's response," "I was concerned that the system would have difficulty recognizing my sign language if I did not pause to check the system's response," "When I use sign

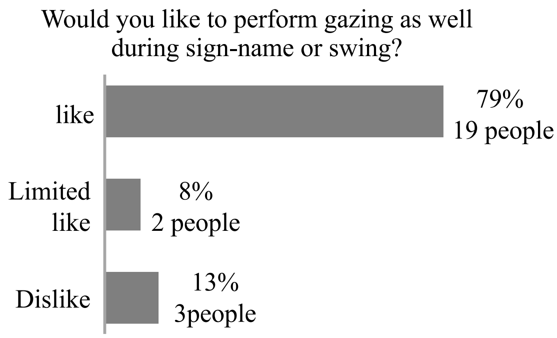

Figure 21. Necessity for gazing during sign name and waving.

language in my daily life, it is natural for me to pause every time I speak, and I would like to do so even if the other party is a system."

Participants who answered "2. agree" mainly commented, "I would like to input commands after I have confirmed that the system responds to my wake-up gesture," "I wanted to check the system's response, and I was concerned that if I didn't pause, it would be difficult for the system to recognize my sign language," "I don't think the pause is so necessary between hearing people. However, I want to be sure that the machine will respond to me before I say what I want."

Participants who answered "3. agree a little" mainly commented, "With voice control devices such as Alexa, I think I can talk without pauses. However, I would prefer to give the wake-up command and see the response from the system, as well as feedback before issuing further commands," "When I am cooking, I don't think I need a pause," "I believe that the system's recognizing of my sign language would be more stable if there was a pause. I thought that if the reading accuracy was as good as a human's, it might not need a pause."

Participants who answered "4. neutral" mainly commented, "It depends on the accuracy of the system's sign language recognition. If the system can recognize the sign language properly, it is not necessary to pause," "I am busy, and I want to send commands immediately after wake-up, considering situations where I have to multi-task."

Participants who answered "5. disagree a little" mainly commented, "Even in interpersonal communication, there is no time between the wake-up and the request. Rather, as the system can see me from the beginning, it should be able to notice my call immediately," "I am impatient, and I wanted to say what I wanted to say right after I woke up the system."

Necessity for gazing when performing wake-up with the hands Figure 21 shows the results of the question as to whether the participants preferred to perform gazing as well during wake-up movements using their hands, such as sign name and waving. Participants who answered "1. like" mainly commented, "I am concerned about the possibility of being mistakenly recognized when I wake-up the system without gazing at it me when talking with others, so I would like to gaze at the system as well," "I'm not used to calling out to people without looking at them because, in interpersonal communication in daily life, we call out with our eyes," "I tend to look at people with my eyes when I talk to them, so I feel safer when I add gazing," "By adding gazing, I can bidirectionally know that the target of the call is the system," "I want to add gazing so that I can check the system's response."

Participants who answered "2. limited like" mainly commented, "I thought that we don't need to gaze at the target of the call because the target of the call can be clearly identified just by signing the name, which is a system-specific way of calling," "If waving is used to call out to the people around me, I would like to gaze at them as well to clarify the target of the call," "If you are busy at work, you may be able to call out without gazing."

Participants who answered "3. dislike" mainly commented, "I don't think gazing is necessary with the possibility of signing the name and waving. If the system can read my sign language, then there is no problem," "I don't feel like calling out while looking at the screen when I envision using the system after coming home. I don't want to be tied to gazing at the screen. I want to look at the screen when the system responds, not to wake it up."

D. Discussion

1) RQ4:Need for feedback: Usability is defined in ISO 9241-11, a standard of the International Organization for Standardization, as "the degree of effectiveness, efficiency, and user satisfaction in which a given user uses a product to achieve a specified goal under specified conditions of use" [56]. From Figure 12, it may be observed that the participants' satisfaction was significantly higher when the system provided feedback, indicating that the usability of the system in terms of "degree of user satisfaction" was improved. However, Figure 13 shows that the time between the call to the system and the command's input was significantly shorter when the system did not provide feedback than when it did do so, indicating that the usability improved in terms of "efficiency". From Figure 16, it may be observed that the ranking of satisfaction was lower when feedback was not provided. In addition, as there were many comments about feeling uneasy, safety could not be ensured, and therefore, the usability decreased in terms of "effectiveness". This suggests that it is more important for usability improvements to obtain a proper response from the system than to speed up the command input in the conversation between a DHH user and the system. In addition, Figure 20 shows that DHH users tended to prefer to check the response from the system by leaving some time between the wake-up action and the command input. These results suggest that the system must provide DHH users with feedback for their wake-up actions.

2) RQ5:Optimal wake-up method: From Figure 14 and Figure 15, it is clear all DHH users preferred wake-up with waving to eye gaze or signing a name. From Section IV-C3, a comment was made by a participant, "waving is often used for calling out in interpersonal communication, so it was easy." This is because it has been reported that Deaf people mainly wave their hands when they try to contact a physically distant person [29]. From these findings, we can say that the wake-up of the system partner showed behaviors similar to Deaf people's conversations. From Section IV-C3, the participants commented that the waving was immediately transferable. From Figure 19, the time between the end of the wake-up action and the start of the command input was significantly shorter for the waving than for the other wake-up conditions. This suggests that waving is practical for a smooth transition to the command input.

Table V shows that 97.2% of the time, the participants were gazing when they performed the waving. Figure 21 shows that even when participants performed wake-up actions using their hands, such as sign name and waving, they tended to perform them in conjunction with gazing. These results suggest that using gaze and waving is an optimal wake-up method for DHH users.

From Table III, participants whose physical characteristics and identity were "Deaf", such as "not relying on auditory information," "high sign language level," and "not using audio information," were significantly more satisfied with waving than with the other wake-up methods. Figure 17 shows that they greatly preferred waving in the satisfaction ranking. In addition, from Table IV, the time required for waving was significantly shorter for participants whose physical characteristics and identities were "Deaf", such as "high sign language level" and "not relying on auditory information," than for the participants with other characteristics. Deaf people have been learning sign language since they were young and are used to communicating with their hands, including sign language, because they belong to the Deaf community. It has been reported that DHH individuals have better visual senses and are more aware of their surroundings than ordinary people [57]. Therefore, if the person you call out to is Deaf, the other person is more likely to notice even if the waving time is short. This suggests that participants who have the physical characteristics of the Deaf are more likely to be accustomed to such situations and therefore prefer hand movements such as waving.

In contrast, from Table III, participants whose physical characteristics and identity were Hard of hearing (hearing impaired), such as "relying on auditory information," "low sign language level," and "using audio information," did not show a preference for waving compared to other wake-up methods. From Table IV, the time required for waving was significantly longer than that of participants whose physical characteristics and identity were "Deaf". This suggests that DHH participants are resistant to using waving as a wake-up movement because such movements are time-consuming. Figure 17 shows that three of the five participants (66% of the total) with a low sign language level preferred gazing. This suggests that hearing-impaired users may prefer not to use their hands, and simply use gaze instead.

3) RQ6:Optimal feedback method: From Figure 14, there was no way to present the system's feedback that resulted in a significantly higher level of satisfaction among all DHH users. However, it is clear from Figure 16 that for all DHH users, the blue bar display was preferred over other feedback.

Regarding the characteristics of DHH users, Figure 18 shows that participants whose identity was "Deaf" were significantly more satisfied with the blue bar display only than with the system's head-shaking motion, and participants with a higher sign language level especially preferred the blue bar display only. In Section IV-C3, there were comments from participants such as "Blue bar display alone is enough," so we suggest that the short time it took for the system to show a response improved the usability of the system.

However, the blue bar display was asymmetrical in that the participants' input method was sign language, but the output method was text or the blue bar display. The blue bar display is asymmetrical from the text and blue bar display. Figure 18 shows that no participants with a low sign language level ranked the sign language display first. Still, many participants with a high sign language level ranked the sign language display as their first or second preference. As for why the satisfaction level of the sign language display was lower than that of the blue bar-only display, based on the participants' comments on the sign language display in the Section IV-C3, it is possible that improving the specifications of the sign language display, such as "displaying the signer as an avatar" or "displaying the signer from the default state, such as the home screen, so that it responds to the user's input," could improve the satisfaction level. These results suggest that the feedback of the system needs to be further studied.

V. DESIGN GUIDELINES FOR A CONVERSATIONAL NATURAL USER INTERFACE FOR THE DEAF AND HARD OF HEARING USERS

From Section III and Section IV, based on the experimental investigation of the research questions, Figure 22 shows the style of the process that DHH users should realize when conversing with a conversational NUI. We propose five design guidelines for conversational NUI for DHH users.

1. Allow wake-up to be performed on a directed gaze
   For DHH users, it is natural to gaze at the system and then enter commands (RQ3). However, it should be noted that the preferred wake-up method differed depending on user characteristics (RQ5). For users who are Deaf, waving can be used as an option to personalize the system. In this case, the designer can use the Table VI data as a reference to create a recognition system for waving and eye gaze.
2. Provide feedback for wake-up
   DHH users prefer confirmation before entering commands to the system after wake-up (RQ4). To make DHH users feel secure, the system should provide feedback such as a sign language display.
3. Command input should be in sign language
   There is some research on alternative input methods to speech for DHH command input, such as sign language [7] [8]. As DHH users are interested in interacting with the system using sign language [9],

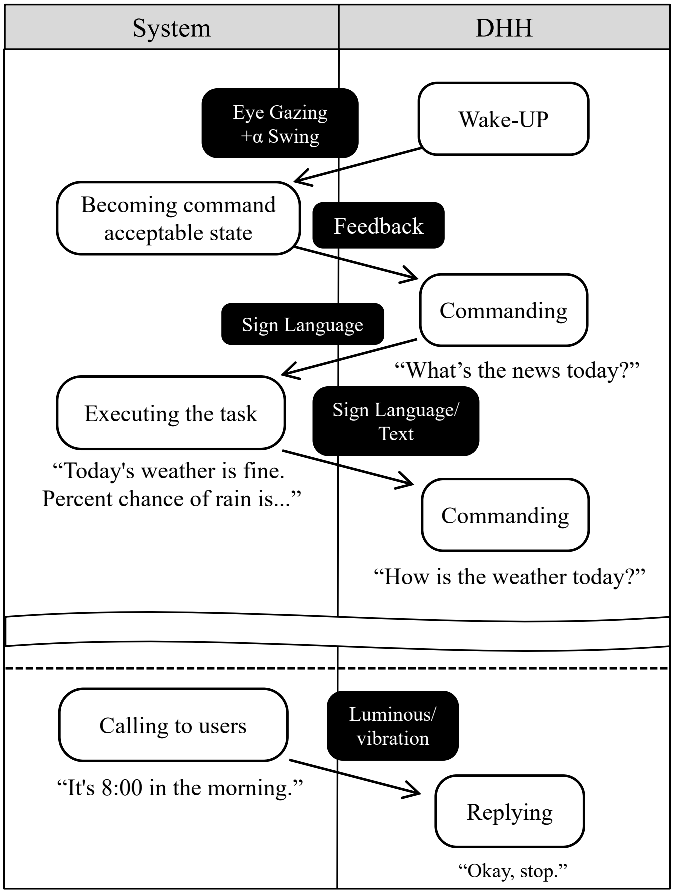

Figure 22. Activity of the conversation process between the DHH user and the system.

they should be able to provide input to the system using sign language.

4. System outputs sign language and text

   Text output is the best form of interaction with the system in multi-tasking. For users who mainly use sign language input, sign language is the best output mode for the system (RQ2). However, the sign language output may not be satisfactory for users who are not accustomed to sign language, so the display method needs to be personalized.

5. Use illumination or vibration to notify the user

   There are times when the DHH user is not looking at the system, so they are called by it and do not notice. To avoid this, the system outputs a luminous notification (RQ1). In addition, the use of vibrations for notifications was considered following the guidelines for mobile applications (applications that run directly on devices such as smartphones and tablets), which were developed with the DHH user experience in mind [58]. Therefore, we can expect vibration-based notification methods in conversational NUIs.

VI. LIMITATIONS AND FUTURE WORK

In this study, the age range of the participants was low (20-24), with all participants being university students, and the sample size was small, with a maximum of 24 participants. Therefore, it was impossible to investigate the preferences and behavior patterns of a wide range of age groups. It has been reported that there are differences in preferences between younger and older people for CUIs equipped with AI assistants [59]. Therefore, evaluation experiments should be conducted with a more diverse range of people in future research.

In addition, 92% of the participants in Experiment 1 and 83% of the participants in Experiment 2 had little or no experience using VUIs. In other words, as the participants were not used to the system, the system's behavior was not very predictable, and their evaluation of the system may change with more regular use. Therefore, an evaluation experiment should be conducted after participants are entirely accustomed to using the system. In addition, issues and comfort levels in daily life should be explored, and fieldwork over extended periods of time should be considered.

In this study, we incorporated a luminous notification as a means of responding to DHH users. However, some participants commented, "I think it would be easier to notice if there was a notification method using vibration as well as light." In the future, we intended to conduct an experiment that includes a vibration notification. In addition, because we placed the system in front of the participants in this experiment, we need to find a way to make them aware of the notifications from behind.

Avatar display functions must also be extended for sign language display and the user face-tracking function for system rotation motion. It is also necessary to investigate usability by participants.

In addition, we examined the system's sign language/text display method and the method of presenting feedback for wake-up, but we did not review feedback for failure to recognize a user's command. When interacting with the system using sign language, experiments should be conducted on possible innovations such as making the signer on the system appear to be asking a question.

When designing a CUI, conventional approaches often consider only a one-time task [52]. However, human conversation rarely involve only a single exchange. Therefore, in the future, it is desirable to design CUIs that allow conversations to continue further. Here, a study reported that in Deaf interpersonal communication, whether a conversation is interrupted is mainly due to the end of mutual gaze [29]. Therefore, it is expected that the gazing modality can be used to determine how to end a conversation. Therefore, it is necessary to conduct experiments on conversational termination method in the future as well.

VII. CONCLUSION

In this study, we have proposed design guidelines for conversational NUIs for DHH users. We have investigated optimal accessibility methods for DHH users at each step of a conversation with VUIs. To this end, we conducted two experiments. In Experiment 1, we asked for responses from DHH users (N=12) to investigate whether a sign language conversation system using luminous notification and gazing could improve usability. In Experiment 2, we collected responses from DHH users (N=24) to investigate optimal wake-up and feedback presentation method.

The main empirical contributions of this work are summarized as follows.

(1) We have provided evidence showing that output with luminous notifications increased DHH users' satisfaction.
(2) We have also demonstrated the necessity of sign language/text output for DHH users.
(3) We have provided evidence that gaze can serve as a natural wake-up method for DHH users, but some users prefer waving.
(4) We have also provided evidence of a high need for DHH users to be provided feedback on wake-up.
(5) Finally, we have developed guidelines for conversational NUIs best suited for DHH users.

This study serves as a design guideline for future conversational NUIs to improve accessibility for DHH users.

References

[1] T. Kato, A. Shitara, N. Kato, and Y. Shiraishi, "Sign Language Conversational User Interfaces Using Luminous Notification and Eye Gaze for the Deaf and Hard of Hearing," Proceedings of ACHI'21: The 14th International Conference on Advances in Computer-Human Interactions, pp. 30–36, Nice, France, 2021.

[2] R. Byron, N. Clifford, "The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places," Bibliovault OAI Repository, the University of Chicago Press, pp. 18–36, 1996.

[3] H. Candello et al., "CUI@CHI: Mapping Grand Challenges for the Conversational User Interface Community," CHI EA '20: In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–8, New York, USA, April 2020.

[4] TECH DRIVERS, (February 15, 2022), "Making sense of Google CEO Sundar Pichai's plan to move every direction at once," https://www.cnbc.com/2017/05/18/google-ceo-sundar-pichai-machine-learning-big-data.html

[5] J. P. Bigham, R Kushalnagar, T. K. Huang, J. P. Flores, S. Savage, "On How Deaf People Might Use Speech to Control Devices," ASSETS '17: In Proceedings of ASSETS '17: The 19th International ACM SIGACCESS Conference on Computers and Accessibility, New York, USA, pp. 383–384, October 2017.

[6] A. Glasser, "Automatic Speech Recognition Services: Deaf and Hard-of-Hearing Usability," In Extended Abstracts of CHI EA '19: The 2019 CHI Conference on Human Factors in Computing Systems, No. SRC06, New York, USA, pp. 1–6, May 2019.

[7] G. Evan, K. Raja S., R. Jason, V. Christian, W. Brittany, "Accessibility of voice-activated agents for people who are deaf or hard of hearing," In Proceedings of CSUN '19: The34th Annual Assistive Technology Conference Scientific/Research, Vol. 7, pp. 144–156, San Diego, 2019.

[8] W. Gilmore et al., "Alexa, Can You See Me?" Making Individual Personal Assistants for the Home Accessible to Deaf Consumers," Proceedings of ASSETS '19: The 35th Annual Assistive Technology Conference Scientific/Research, Vol. 8, pp. 16–31, San Diego, USA, 2020.

[9] A. Glasser, V. Mande, M. Huenerfauth, "Understanding deaf and hard-of-hearing users' interest in sign-language interaction with personal-assistant devices," In Proceedings of W4A '21: The 18th International Web for All Conference, Association for Computing Machinery, pp. 1–11, New York, USA, 2021.

[10] A. Glasser, V. Mande, M. Huenerfauth, "Accessibility for Deaf and Hard of Hearing Users: Sign Language Conversational User Interfaces," In Proceedings of CUI '20: The 2nd Conference on Conversational User Interfaces, New York, USA, No. 55, pp. 1–3, July 2020.

[11] D. Bragg et.al., "Sign Language Interfaces: Discussing the Field's Biggest Challenges," In Extended Abstracts of CHI EA '20: The 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–5, New York, USA, April 2020.

[12] D. Bragg et.al., "Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective," In Proceedings of ASSETS '19: The 21st International ACM SIGACCESS Conference on Computers and Accessibility, pp. 16–31, New York, USA, October 2019.

[13] Nielsen Norman Group, (February 15, 2022), "10 Usability Heuristics for User Interface Design," https://www.nngroup.com/articles/ten-usability-heuristics/

[14] J. Nielsen, "Enhancing the explanatory power of usability heuristics," In Proceedings of CHI '94: The SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, pp. 152–158, New York, USA, 1994.

[15] D. Norman, "The design of everyday things," Basic Books, 1988.

[16] B. Shneiderman, C. Plaisant, "Designing the User Interface: Strategies for Effective Human-Computer Interaction," Addison Wesley, 2010.

[17] C. Murad, C. Munteanu, L. Clark, B. R. Cowan, "Design guidelines for hands-free speech interaction," In Proceedings of MobileHCI '18: The 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Association for Computing Machinery, pp. 269–276, New York, USA, 2018.

[18] Alexa, (February 15, 2022), "Design Principles," https://developer.amazon.com/en-US/alexa/alexa-haus/design-principles

[19] Google Assistant, (February 15, 2022), "Conversational Actions, Guides Design guidelines," https://developers.google.com/assistant/interactivecanvas/design

[20] C. Murad, C. Munteanu, "'I don't know what you're talking about, HALexa': the case for voice user interface guidelines," In Proceedings of CUI '19: The 1st International Conference on Conversational User Interfaces, Association for Computing Machinery, No. 9, pp. 1–3, New York, USA, 2019.

[21] J. Berke, (February 15, 2022), "Assistive Listening Devices for the Deaf and HOH," https://www.verywellhealth.com/assistive-listening-devices-1046105

[22] A. Matsuda, M. Sugaya, H. Nakamura, " Luminous device for the deaf and hard of hearing people. In Proceedings of the second international conference on Human-agent interaction," In Proceedings of HAI '14: The second international conference on Human-agent interaction, pp. 201–204, October 2014.

[23] Amazon, (February 15, 2022), "Hearing Communicate and stay connected with Alexa," https://www.amazon.com/b/ref=ods_afe_hop_hp?node=21213721011

[24] F. Bentley et al., "Understanding the Long-Term Use of Smart Speaker Assistants," In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 2, No. 3, pp 1–24, September 2018.

[25] M. Marschark et al., "Benefits of Sign Language Interpreting and Text Alternatives for Deaf Students' Classroom Learning," The Journal of Deaf Studies and Deaf Education, Vol. 11, pp. 421–437, Fall 2006.

[26] S. Deepika, R. Rangasayee, "The Combined Effect of Captioning and Sign Language in Understanding Television Content in Deaf," Journal of Communication Disorders, Deaf Studies & Hearing Aids, Vol. 6, No. 1, pp. 1–7, 2018.

[27] M, Debevc, P. Kosec, A. Holzinger, "Improving multimodal web accessibility for deaf people: sign language interpreter module," Multimed Tools Appl Vol. 54, pp. 181–199, April 2010.

[28] SignGenius, (February 15, 2022), "Do's & Don'ts - Getting Attention in the Deaf Community," https://www.signgenius.com/info-do's&don'ts.shtml

[29] U. Bartnikowska, "Significance of touch and eye contact in the Polish Deaf community during conversations in Polish Sign Language: ethnographic observations," Hrvatska Revija za Rehabilitacijska Istraživanja, Vol. 53, pp. 175–185, 2017.

[30] V. Mande, A. Glasser, B. Dingman, M Huenerfauth, "Deaf Users' Preferences Among wake-up Approaches during Sign-Language Interaction with Personal Assistant Devices," In Extended Abstracts of CHI EA '21: The 2021 CHI Conference on Human Factors in Computing Systems, Vol. 370, pp 1–6, May 2021.

[31] C. Heath, K. Nicholls, Body Movement and Speech in Medical Interaction, Studies in Emotion and Social Interaction, Cambridge University Press, 1986.

[32] A. M. Lieberman, "Attention-getting skills of deaf children using American Sign Language in a preschool classroom," Applied psycholinguistics, Vol. 36, No. 4, pp. 855–873, July 2016.

[33] National University Corporation, Tsukuba University of Technology, (February 15, 2022), "About," https://www.tsukuba-tech.ac.jp/english/index.html

[34] I. Lopatovska et al., "Talk to me: Exploring user interactions with the Amazon Alexa," Journal of Librarianship and Information Science, Vol. 51, No. 4, pp. 984–997, 2019.

[35] S. Ahire, A. Priegnitz, O. Önbas, M, Rohs, W. Nejdl, "How Compatible is Alexa with Dual Tasking? — Towards Intelligent Personal Assistants for Dual-Task Situations," In Proceedings of HAI '21: The 9th International Conference on Human-Agent Interaction, Association for Computing Machinery, pp. 103–111, New York, USA, 2021.

[36] Canalys, (February 15, 2022), "Global smart speaker market 2021 forecast," https://canalys.com/newsroom/canalys-global-smart-speaker-market-2021-forecast

[37] National Institutes of Health (US); Biological Sciences Curriculum Study, "NIH Curriculum Supplement Series," Bethesda (MD): Information about Hearing, Communication, and Understanding, 2007.

[38] Pixel Phone Help, (February 15, 2022), "Control your Pixel without touching it," https://support.google.com/pixelphone/answer/9517454?hl=en

[39] F. Donna, "The Use of waving as Self-Generated Cues for Recall of Verbally Associated Targets," The American journal of psychology, Vol. 115, No. 1, pp. 1–20, 2002.

[40] F. Donna, R. E. Guttentag, "The Effects of Restricting waving Production on Lexical Retrieval and Free Recall," The American Journal of Psychology, Vol. 115, No. 1, pp. 1–20, 2002.

[41] M. Henschke, T. Gedeon, R. Jones, "Touchless Gestural Interaction with Wizard-of-Oz: Analysing User Behaviour," In Proceedings of OzCHI '15: The Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, Association for Computing Machinery, pp. 207–211, New York, USA, 2015.

[42] J. Schwarz, C. C. Marais, T Leyvand, S. E. Hudson, J. Mankoff, "Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces," In Proceedings of CHI '14: The SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, pp. 3443–3452, New York, USA, 2014.

[43] W. Yee, "Potential Limitations of Multi-touch Gesture Vocabulary: Differentiation, Adoption, Fatigue," In Proceedings of the 13th International Conference on Human-Computer Interaction, Part II: Novel Interaction Methods and Techniques. Springer-Verlag, pp. 291–300, Berlin, Heidelberg, July 2009.

[44] R. Cui, H. Liu, C. Zhang, "A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training," in IEEE Transactions on Multimedia, Vol. 21, No. 7, pp. 1880–1891, July 2019.

[45] N. Crook, (February 15, 2022), "Wizard of Oz testing – a method of testing a system that does not yet exist," https://www.simpleusability.com/inspiration/2018/08/wizard-of-oz-testing-a-method-of-testing-a-system-that-does-not-yet-exist/

[46] N. Fraser, N. Gilbert, "Simulating speech systems. Computer Speech & Language," Vol. 5, pp. 81–99, January 1991.

[47] A. Pradhan, K. Mehta, L. Findlater, ""Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities," In Proceedings of CHI '18: The 2018 CHI Conference on Human Factors in Computing Systems, No. 459, pp. 1–13, New York, USA, 2018.

[48] J. Peirce, J. R. Gray, "Simpson, S. et al. PsychoPy2: Experiments in behavior made easy," Behavior Research Methods, Vol. 51, pp. 195–203, February 2019.

[49] The Language Archive, (June 5, 2021), "ELAN," https://archive.mpi.nl/tla/elan

[50] J. Brooke, "SUS: A quick and dirty usability scale," Usability Eval. Ind.. 189. pp. 1–7. November 1995.

[51] D. Bavelier et al., "Visual attention to the periphery is enhanced in congenitally deaf individuals," The Journal of neuroscience : the official journal of the Society for Neuroscience, Vol. 20, No. 17, pp. 1–8. September 2000.

[52] C. Pearl, "Designing Voice User Interfaces: Principles of Conversational Experiences," O'Reilly Media, 2017.

[53] Developer documentationamazon alexa, (February 15, 2022), "Alexa Design Guide. Be Avaiable," https://developer.amazon.com/en-GB/docs/alexa/alexa-design/available.html

[54] Amazon, (February 15, 2022), "What Do the Lights on Your Echo Device Mean?," https://www.amazon.com/gp/help/customer/display.html?nodeId=GKLDRFT7FP4FZE56

[55] Amazon, (February 15, 2022), "Echo Show 10 (3rd Gen) — HD smart display with motion and Alexa — Charcoal," https://www.amazon.com/echo-show-10/dp/B07VHZ41L8

[56] Online Browsing Platform(OBP), (February 15, 2022), "ISO 9241-11:2018(en)" https://www.iso.org/obp/ui/\#iso:std:iso:9241:-210:ed-2:v1:en

[57] D. Bavelier, W.G. Matthew, P. C. Hauser, "Do deaf individuals see better?," Trends in Cognitive Sciences, Vol. 10, No. 11, pp. 512–518, 2006.

[58] R. P. Schefer, M. S. Bezerra, L. A. M. Zaina, "Supporting the Development of Social Networking Mobile Apps for Deaf Users: Guidelines Based on User Experience Issues," In Proceedings of DSAI 2018: The 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, Association for Computing Machinery, pp. 278—285, New York, USA, 2018.

[59] D. Gollasch, G. Weber, "Age-Related Differences in Preferences for Using Voice Assistants," In Proceedings of MuC '21: In Mensch und Computer 2021, Association for Computing Machinery, pp. 156–167, New York, USA, 2021.