Objective Evaluation Method of Reading Captioning using HMD for Deaf and Hard-of-Hearing

Nobuko Kato* and Hiroya Kitamura*
Faculty of Industrial Technology, Tsukuba University of Technology, Tsukuba, Ibaraki, Japan

Keywords: Assistive Technology, Communication Support, Closed Caption, Accessibility, Person with Deafness or Hearing Impairment.

Abstract: In this basic study on the presentation of captions to deaf and hard-of-hearing people using augmented reality technology, we propose an objective evaluation method of reading captioning using speed of keystroke. In an experiment in which various types of information were presented in the peripheral vision, the speed of keystrokes varied according to the complexity of the information presented, suggesting that this speed can be used as an objective evaluation. In the experiment where captioning with different contents and simplicity were presented, there was a significant difference in the speed of keystroke depending on the contents.

1 INTRODUCTION

In recent years, museums have become facilities that play an important role in science education. For museums to provide a positive and worthwhile experience to people who are deaf and hard-of-hearing (DHH), they must include some type of support to supplement auditory information, such as creating learning content (Constantinou, 2016) and preparing guided tours for such people (Namatame, 2019).

Advances in research investigations focused on museum education using augmented reality (AR) technology to add information to the real world (Gonzalez Vargas, 2020) and the development of speech recognition technologies (Shadiev, 2014) have triggered research on the presentation of closed captions using head-mounted displays (HMDs) (Olwal, 2020). This method is expected to make it possible for people with hearing impairment to participate in real-time events by presenting the results of speech recognition as captions even in situations where a sign language interpreter is not present.

Earlier studies have shown that reading captions while listening to audio and watching images is a labour-intensive task (Diaz-Cintas, 2007). Further, compared to the task of listening to audio while viewing images that provide visual information, simultaneously viewing images and reading text are believed to create mental overload in recipients. In an experiment conducted on people who are DHH where the result of speech recognition was presented as captions on HMDs during a guided tour of a museum, it was suggested that although the subjective evaluation of captioning in the questionnaire was very high, there might not have been sufficient time to view the exhibits (Kato, 2020).

The spread of AR technology to communication support anywhere and anytime in one's field of vision, using a transparent HMD, is the first step toward the realization of a dream. In an online survey with 201 DHH participants, 70% of participants were very or extremely interested in making full-captioning of conversations available on wearable devices (Findlater, 2019). On the other hand, from the standpoint of safety and comprehension of the presented content, it is necessary to objectively evaluate the amount of cognitive load that is placed on DHH people during the presentation of captions and the extent of influence they have on the task being performed.

There are two ways to measure cognitive load in subtitling: subjective and objective evaluations (Brünken, 2003). Subjective evaluation includes self-reported invested mental effort or self-reported stress level. Brain activity measures (e.g., fMRI) or gaze measurement are used for objective evaluation, but

* http://www.tsukuba-tech.ac.jp/

Kato, N. and Kitamura, H.
Objective Evaluation Method of Reading Captioning using HMD for Deaf and Hard-of-Hearing.
DOI: 10.5220/0010477605730578
In *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU 2021) - Volume 1*, pages 573-578
ISBN: 978-989-758-502-9
Copyright © 2021 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

this measurement and analysis are not always simple. A simple and objective evaluation method is necessary to clarify the appropriate captioning for DHH people by varying various factors.

Hence, in this study, we propose an objective evaluation method using the speed of keystrokes and examine its effectiveness.

2 RELATED WORK

To quantitatively evaluate the effect of AR on work speed, an earlier study examined the method of presenting information in the peripheral field of an AR environment (Ishiguro and Rekimoto, 2011). This study evaluated the effect of the complexity of the display position and contents on work speed and obtained the objective results of the effect of presenting information in the central or peripheral vision on the people's work speed.

On the other hand, a study that analyzed the visual information-processing characteristics of people who are DHH suggested that an enhancement of visual attention to peripheral visual space in deaf individuals (Bavelier, 2000; Bosworth, 2002). Therefore, it is necessary to confirm whether the keystroke method used for the evaluation of presenting information on AR glasses can also be used for people who are DHH.

In the experiment to measure the cognitive load of deaf students when using the online-learning materials with or without captioning, participants answered the questions on a seven-point scale, and there was no significant difference in cognitive load (Yoon, 2011); therefore, objective evaluation is expected. Thus, it is necessary to develop a simple objective evaluation method for AR glasses to present complex information containing mixed phonetic and ideographic characters (such as Japanese captions) to people who are DHH.

3 EXPERIMENT METHOD

In this study, we will examine the following two points in order to study the appropriate captioning for people who are DHH.

- Is the speed of keystrokes while reading captioning appropriate as an objective evaluation index?
- What are the factors that affect the reading of captioning, such as content and simplicity?

(a) example of symbol (b) example of icon (c) example of three-letters alphabet

Figure 1: Examples of symbols and icons used in Experiment 1.

Accordingly, two experiments were conducted: one in which several types of information were presented in the peripheral vision and keystrokes were performed, and the other in which captions were presented on the assumption that people who are DHH would use AR glasses when visiting a museum.

3.1 Experiment 1: Viewing Information While Keying

In order to confirm whether the speed of keystrokes can be used as an objective evaluation of the reading captioning by people who are DHH, we conducted an experiment in which information is presented in the peripheral vision while keystrokes are being made. To compare the results of people who are DHH with those of participants with no hearing impairment, we used the same experimental method as in the earlier study, that is, research participants input keys while displaying the symbols and letters in the central and peripheral visions (Ishiguro and Rekimoto, 2011).

3.1.1 Method of Experiment 1A

We installed a monitor in front of the experimenter (instead of an HMD for AR glasses) as the experimental environment. The display screen was placed approximately 60° from the participants' field of view. The participants were required to input the numbers (0–3) displayed in the centre of the screen as a task using a keyboard. The screen presented five different types of content:

- **[No Display]** Do not display additional information.
- **[Symbol In Centre]** Display a symbol (a circle, cross, triangle, or rectangle) in the centre of the screen(Figure 1(a)).
- **[Symbol In Peripheral Vision]** Display a symbol (circle, cross, triangle, or rectangle) in the peripheral vision at a viewing angle of 28°.

Table 1: Results of experiment 1A: changes in average task processing time according to displayed content.

Displayed content	No display	Symbol in centre	Symbol in peripheral vision	Icon in peripheral vision	Three-letters alphabet in peripheral vision
Average task processing time (second)	0.60	0.76	0.69	0.69	0.81
Standard deviation of task processing time	0.09	0.15	0.18	0.16	0.15
Error rate [%]	2.4	2.6	3.8	3.6	4.4

- **[Icon In Peripheral Vision]** Display an icon, such as a mail mark or a music symbol, in the peripheral vision of the participants (Figure 1(b)).
- **[Alphabet In Peripheral Vision**] Display three letters of the English alphabet selected randomly in the peripheral vison of the participants (Figure 1(c)).

Each symbol or icon was presented at random intervals of 2–3 seconds for 3 seconds. Each participant was required to sign the identified symbol or finger alphabet to confirm that they had seen the symbol. Each experiment took approximately 3 minutes, with 10 DHH participants who are university students in their 20s.

3.1.2 Method of Experiment 1B

In Experiment 1B, study participants viewed a background image and the corresponding caption while performing the task of entering the number presented in the centre of the screen using a keyboard.

Table 2: Results of experiment 1B.

Average task processing time (second)	1.25
Standard deviation of task processing time	0.44
Error rate [%]	3.0

Captioning was displayed in three lines at the bottom of the screen with no more than 23 characters per line. In one experiment, the captioning was presented for approximately 3 minutes, and 8 university students who are DHH participated in this experiment.

3.2 Method of Experiment 2

In order to investigate the factors that affect the reading of captioning, we conducted an experiment using captioning with different contents and simplicity.

As in Experiment 1B, participants in Experiment 2 viewed a background image and several types of captions while performing the task of entering the number using a keyboard. The screen presented the following two types of closed captions of explanatory notes of the museum exhibition:

- Closed captions that presented the original text as it was.
- Closed captions that presented simplified text; the number of characters was set to approximately 80% of the original text.

The experiment was performed twice, with each participant experiencing two types of captioning with different contents (A and B) and two types of captioning with different simplicity, original text and that of the simplified text, presented in a random order. A comprehension test was conducted before and after the experiment, and the each test scored out of 15.

Captioning was displayed in three lines at the bottom of the screen with no more than 23 characters per line. In one experiment, the captioning was presented for approximately 3 minutes, with 12 DHH participants who are university students in their 20s.

Table 3: Number of keystrokes per minute.

Displayed content	No display	Symbol in centre	Symbol in peripheral vision	Icon in peripheral vision	Three-letters alphabet in peripheral vision	Captioning
Average number of keystrokes	94.7	75.8	87.0	84.5	73.3	48.0
Percentage compared to [No Display]	100%	80%	92%	89%	77%	51%

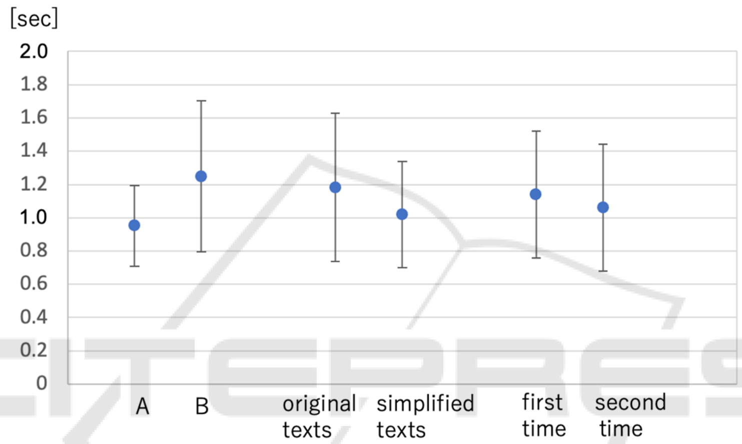

Figure 2: Average task processing time for key inputs.

4 RESULTS OF EXPERIMENT

4.1 Result of Experiment 1A

Table 1 depicts the results of Experiment 1A and reveals the average task processing time and standard deviation of the processing time for keystrokes with five different displays, as well as the percentage of incorrect keystrokes. Table 1 indicates the following:

- The average processing time is longer when there is a symbol in the centre of the screen than when there is a symbol in the peripheral vison of the participant.
- The processing time tends to be longer when there are three-letters of the alphabet in the participant's peripheral vision than when there are symbols in the peripheral vision.

These tendencies were similar to the results of hearing experiments reported by earlier studies (Ishiguro and Rekimoto, 2011).

The results of a t-test on the task processing time revealed a significant difference between no display and symbol in the centre and no display and letters of the alphabet in the participant's peripheral vison, at $p<0.01$.

In addition, the task processing time for letters of the alphabet in the peripheral vision of the participant was significantly longer than that for icons ($p<0.01$). For people who are DHH as well as those with no hearing impairment, reading the alphabet in the peripheral vision significantly slowed down the task processing time ($p<0.01$).

4.2 Result of Experiment 1B

Table 2 shows the results of Experiment 1B. Comparing the average processing time for three letters of the alphabet in the periphery in Experiment 1A (Table 1) and captioning presentation in Experiment 1B (Table 2), we found that the processing time for captioning presentation was significantly longer ($p<0.05$).

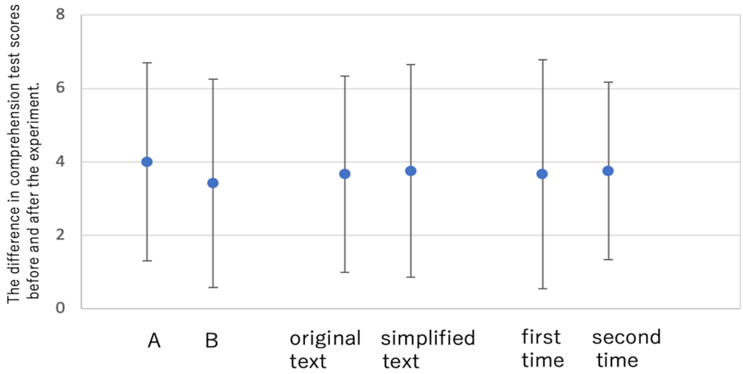

Figure 3: Results of comprehension test conducted before and after the experiment.

Table 3 lists the number of keystrokes for each display content. The depicted percentages indicate the ratio of the number of keystrokes compared to the case where no information was displayed in the participant's peripheral vision.

4.3 Result of Experiment 2

Figure 2 depicts the average processing time for the key input while participants were reading captions. We analyzed the average processing time from the following three perspectives:

- The average times for contents A and B were 0.95 and 1.25 seconds, respectively ($p<0.05$).
- The average processing time for original text was 1.18 seconds and that for simplified text was 1.02 seconds.
- The average for the first time was 1.14 seconds and that for the second time was 1.06 seconds.

There was a significant difference in processing speed between A and B with different contents.

Figure 3 depicts the average of the difference in comprehension test scores before and after the experiment. The results of the comprehension test show as follows:

- The difference in comprehension test scores was 4.0 points on average for A, and 3.4 points on average for B.
- The difference in comprehension test scores was 3.7 points on average for the original text, and 3.8 points on average for the simplified text.
- The difference in comprehension test scores for the first time was 3.7 points and that for the second time was 3.8 points.

There was no significant difference between the mean scores of the A and B, original and simplified text, the first time and the second time.

5 DISCUSSION

5.1 Validity as an Objective Evaluation Index

The results of Experiment 1 revealed that reading the alphabet in the peripheral vision significantly slowed down the task processing time for both people who are DHH and those with no hearing impairment ($p<0.01$). The results of the experiment showed the same tendency for both people who are DHH and the participants with no hearing impairment, and there was no evidence that people with deafness were superior to others in reading text in the peripheral vision. In other words, the speed of keystrokes varied according to the information presented in the experiment with people who are DHH.

The results of Experiment 1B revealed that the processing time increased when participants read the captioning than when they checked the three-letter alphabet in the peripheral vision. In other words, reading captioning likely reduced work efficiency. The workload was reduced to approximately half the original workload (Table 3). This result is consistent with the results of previous objective evaluation experiments, such as eye gaze measurement.

In Experiment 2, we tried two objective evaluations: a comprehension test and keystrokes. Compared to the comprehension test, the keystrokes were found to be less influenced by other factors such as the presence or absence of prior knowledge. Therefore, keystrokes are considered to be an effective indicator for the evaluation of reading captioning.

5.2 Factors Influencing the Reading Captioning

In Experiment 2, we compared the effects of different caption contents A and B and different simplicity in the original and simplified texts on the work. We found a significant difference in processing time depending on the content, rather than the number of characters, of captions. It has been pointed out in previous studies that the content of captioning has a greater influence on reading than the number of characters or the speed of display; an identical tendency was found in the Japanese language, where phonetic and ideographic characters are mixed.

In previous studies, subjective evaluations such as questionnaires have been used; however, we were able to show an objective evaluation index using a simple method of key-input speed in the present study.

6 CONCLUSIONS

We conducted experiments to clarify the effectiveness of an objective evaluation index for considering what is appropriate captioning for people who are deaf or hard of hearing (DHH), assuming that we assisted such people by presenting captions using AR technology. In an experiment in which symbols, icons, or captioning were presented in the peripheral vision to people who are DHH, the results showed that keystroke speed varied appropriately with information. In other words, the key-input can be used to evaluate the reading of captioning.

In our experiment using captioning with varying content and simplicity, we confirmed that the content affected the reading of captioning using our proposed objective evaluation method.

Compared to the experiment using symbols, the standard deviation of task processing time tended to be larger in the experiment using captioning, indicating large individual differences in reading captioning. It has been pointed out that effectiveness of captions is strongly related to the level of individual reading skills (Lewis, 2001); this point must be clarified in the future.

ACKNOWLEDGEMENTS

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Number JP 18H03660).

REFERENCES

Bavelier, A., Tomann, C., et.al., 2000. Visual attention to the periphery is enhanced in congenitally deaf individuals. *Journal of Neuroscience*, 20(17), RC93.

Bosworth, R.G., Dobkins, K.R., 2002. Visual field asymmetries for motion processing in deaf and hearing signers. *Brain and Cognition*. 49(1), 170-181.

Brünken, R., Plass, J.L., and Leutner, D., 2003. Direct measurement of cognitive load in multimedia learn-ing. *Educational Psychologist*, 38(1), 53–61.

Constantinou, V., Loizides, F.,et.al., 2016. A personal tour of cultural heritage for deaf museum visitors. Progress in cultural heritage: Documentation, preservation, and protection. *EuroMed 2016. Lecture notes in computer science*, 10059. Springer, Cham, 214-221.

Díaz-Cintas, J., Remael, A., 2007. Audiovisual translation: Subtitling. *Manchester & Kinderhook*, St. Jerome.

Findlater, L., Chinh, B., et.al., 2019. Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. *2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, 46, 1–13.

Gonzalez Vargas, J.C., Fabregat, R., Carrillo−Ramos, A., Jove, T., 2020. Survey: Using Augmented Reality to Improve Learning Motivation in Cultural Heritage Studies. *Applied Science*, 10(3), 897.

Ishiguro, Y., Rekimoto, J., 2011. Peripheral vision annotation: Noninterference information presentation method for mobile augmented reality. *AH'11: 2nd Augmented Human International Conference*, (8), 1-5.

Kato, N., Kitamura, M., Namatame, M., et al., 2020. How to make captioning services for deaf and hard of hearing visitors more effective in museums?, *12th International Conference on Education Technology and Computers (ICETC '20)*, 157-160.

Lewis, M., & Jackson, D., 2001. Television Literacy: Comprehension of Program Content Using Closed Captions for the Deaf, *The Journal of Deaf Studies and Deaf Education*, 6(1), 43–53,

Namatame, M., et al., 2019. Can exhibit-explanations in sign language contribute to the accessibility of aquariums?, *HCI International 2019*, 289-294.

Olwal,A., Balke, K., Votintcev,D., et al., 2020. Wearable Subtitles: Augmenting Spoken Communication with Lightweight Eyewear for All-day Captioning, *Annual ACM Symposium on User Interface Software and Technology (UIST '20)*, 1108–1120.

Shadiev, R., Hwang, W., et al., 2014. Review of speech-to-text recognition technology for enhancing learning. *Journal of Educational Technology & Society*, 17(4), 65-84.

Yoon, J. O., & Kim, M., 2011. The effects of captions on deaf students' content comprehension, cognitive load, and motivation in online learning. *American annals of the deaf*, 156(3), 283–289.