Word Cloud Meeting: A Visualization System for DHH People in Online Meetings

RYO IIJIMA, University of Tsukuba, Japan
AKIHISA SHITARA, University of Tsukuba, Japan
SAYAN SARCAR, University of Tsukuba, Japan and Birmingham City University, UK
YOICHI OCHIAI, University of Tsukuba, Japan

Fig. 1. Our system for visualizing conversations for DHH people in online meetings. The videos used in the experiment consist of three participants and one caption.

It has become common for deaf and hard of hearing (DHH) people to participate in online meetings using real-time captions. In recent years, the usability of a system that visualizes the ongoing topic in a conference has been confirmed, but it has not been verified in a remote conference that includes DHH people. One possible reason is that visual dispersion occurs when there are multiple sources of visual information. In this study, we introduce "Word Cloud Meeting," a system that generates a separate word cloud for each participant and displays it in the background of each participant's video to visualize who is saying what. We conducted an experiment with seven DHH participants and obtained positive qualitative feedback on the ease of recognizing topic changes. However, when the topic changed in a sequence, it was found to be distracting. Overall, experimental results show that the proposed system is useful for observing topic shifts. Additionally, we discuss the design implications for visualizing topics for DHH people in online meetings.

CCS Concepts: • **Human-centered computing** → *Accessibility design and evaluation methods*; *Accessibility technologies.*

Additional Key Words and Phrases: Visualization, User study, Deaf, Hard of hearing

Authors' addresses: Ryo Iijima, ryoiijima@digitalnature.slis.tsukuba.ac.jp, University of Tsukuba, Tsukuba, Japan; Akihisa Shitara, University of Tsukuba, Tsukuba, Japan; Sayan Sarcar, University of Tsukuba, Japan and Birmingham City University, UK; Yoichi Ochiai, wizard@digitalnature.slis.tsukuba.ac.jp, University of Tsukuba, Tsukuba, Japan.

Manuscript submitted to ACM

ACM Reference Format:
Ryo Iijima, Akihisa Shitara, Sayan Sarcar, and Yoichi Ochiai. 2018. Word Cloud Meeting: A Visualization System for DHH People in Online Meetings. 1, 1 (October 2018), 5 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION AND RELATED WORK

The COVID-19 pandemic has forced us to hold meetings and classes online, and there is a growing need to consider the accessibility of online meetings. Therefore, it has become common to accommodate deaf and hard of hearing (DHH) participants using real-time transcription applications. However, in situations where DHH people receive visual information other than captions, such as information from slides and teachers, they need to juggle between multiple concurrent sources of information.

Visual dispersion is a challenge for DHH people [15]. To address this problem, a notification system for changes in information sources [7], and a system that tracks the presenter and minimizes the visual distance between the caption and the speaker [14] have been proposed. In terms of caption placement, several researchers have investigated caption locations on the screen [13] or in the display area of a head-mounted display [11].

However, additional visual sources of information, if properly designed, can be beneficial to a user. Some researchers have explored ways to visualize topic transitions during meetings to aid users' awareness of the discussion [8]. Others have visualized participant engagement [5] in online meetings. Starting from Vannevar Bush's Memex [6], research on such smart systems has a long history; however, investigations on the benefits of such systems to DHH people are few.

To investigate this issue, we introduce "Word Cloud Meeting," a system that allows users to obtain an overview of a conversation while maintaining a small visual dispersion. The system generates a separate word cloud for each participant to visualize who is saying what. It then displays it in the background of each participant's video in real time. We implemented an initial prototype of Word Cloud Meeting and refined it based on the feedback from a pilot study. Then, we conducted an experiment with seven DHH participants to determine how the frequency of topic shifts affects the perception of Word Cloud Meeting. In this paper, we discuss the implications for the next design iteration and design guidelines for future research based on our analysis of the semi-structured interviews.

2 PILOT STUDY

We co-designed the initial prototype with our co-author Shitara, who has profound hearing loss, and identified the following requirements: Users can 1) understand the ongoing topics in a discussion at a glance, 2) understand who is saying what, and 3) be aware of changes in information. Word cloud was adopted as a visualization method, as it was reported that visualization makes it easier to keep track of topics in face-to-face meetings [8].

Prototype Design: We implemented a web-based application that takes speech as input to generate and update a word cloud. This system first transcribes speech from a microphone into text, and then outputs text data containing nouns, verbs, adjectives, and adverbs extracted by morphological analysis. We used the Web Speech API [4] for transcription and kuromoji.js [3], an open-source JavaScript library, for morphological analysis. Next, the system generates a word cloud from the text data and renders it using D3.js [2]. To determine the word layout, we used an existing implementation by Jason [9]. When a certain period of silence is detected, the system considers it as the end of one sentence. Every time someone speaks a sentence, the system performs the above process to update the word cloud.

Study Design: Two DHH participants (one female and one male, age 24 and 46) were included in this pilot study. The study took one hour and consisted of a demographic questionnaire followed by a semi-structured interview. The purpose of the semi-structured interview was to receive feedback on a possible direction for design and functionality, and to determine parameters that remain unknown along with user behavior. We explored the number of words that would be perceived as confusing by asking the participant while changing the number of words displayed in the author's video. This pilot study and all other experiments reported in this paper were approved by the university ethics committee.

Feedback and System Improvement: Both participants thought that visualizing online meetings in real time would reduce their workload because it is tiring to always follow only the captions. Based on this feedback, the following modifications were made to the system:

- To illustrate a change in topic, new words are presented in a darker color, the words fade over time, and eventually disappear if they are not spoken by the participant.
- Because the participants of the pilot study reported of possible confusions due to too many keywords, old words were removed to avoid displaying more than 30 words simultaneously.

3 EXPERIMENT

3.1 Participants

We recruited seven DHH participants (four females and three males) through email and snowball sampling. Participants were 24.5 years old on average (SD=7.42, range 21-25). They were paid ¥860 for the 60-minute study. We confirmed that all participants had normal vision for daily activities and that they were able to watch videos during our study.

3.2 Procedure

The study was conducted remotely because of the COVID-19 pandemic and social distancing restrictions in Japan. The study protocol began with a background questionnaire, followed by a two-part protocol.

Part 1: Video watching. Prior to the experiment, we recorded an online meeting in which three people were talking in three different scenarios with an average duration of 30 seconds using our system (Figure 1). During recording, a virtual camera application (mmhmm [1]) superimposes a person on a word cloud generated by our system, which is then streamed to an online meeting. The difference between the three scenarios is the frequency of the topic changes. In Scenario 1, the participants stayed on a single topic, whereas in Scenario 2, the topic changed naturally. In Scenario 3, to examine a user's reaction to a sequential change in the topic, the topic was changed with each speaker taking their turn, resulting in a script with four topic changes. Because all the participants were in their 20s, the topics in each scenario were designed to be familiar to young people (e.g., music, exercises). To simulate a situation of watching a real-time transcript while having an online meeting, we designed a specific space on the screen to display the transcript. The participants watched three videos in a random order. We assumed a situation in which four people, including a participant, had an online meeting, and the participant was observing the conversation of the other three. After watching each video, participants answered seven questions, including 5-point Likert scale questions used to collect subjective responses about captions from DHH participants in a prior work [12] and NASA-TLX [10].

Part 2: Semi-structured interviews. We asked semi-structured questions about the experience of watching the video and delved into the reasons for the answers given in Part 1.

3.3 Result

Figure 2 and Figure 3 show the participants' responses to the 5-point Likert questions, and NASA-TLX, respectively. The results were derived using the Kruskal-Wallis test, and we did not find any statistical differences between scenarios. P1, P3, P4, P6 and P7 stated that they were able to read the shifts in topics from the word cloud. The results of the semi-structured interviews were open-coded and organized into the following three main themes:

Customizability. P3 suggested the ability to mute words that were not important to him. P1 and P6 wanted to see only nouns and verbs because they consciously followed nouns and verbs in their daily conversations, whereas P3

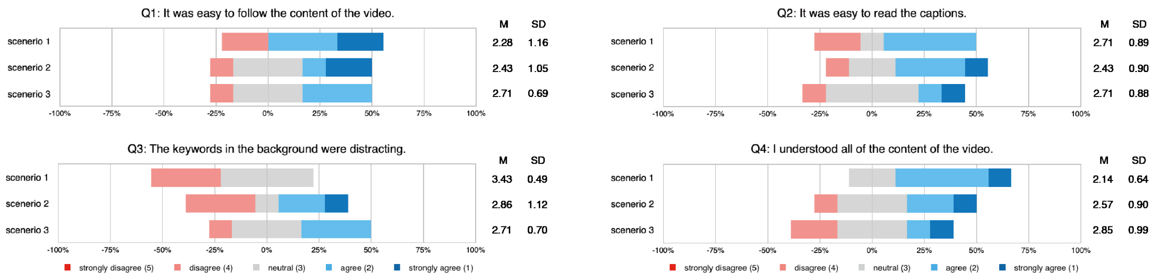

Fig. 2. Distributions of participants' responses on the usability of the visualizations.

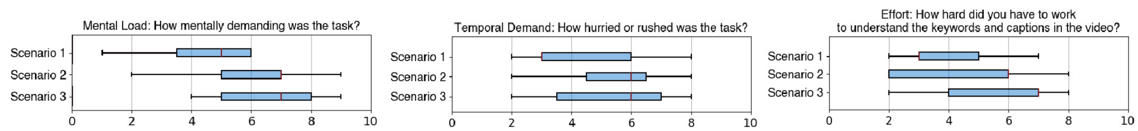

Fig. 3. Summary of participant's rating on NASA-TLX.

wanted only nouns. Furthermore, P1, P3 and P6 wanted to change the color of the word cloud to make it easier for them to see, and P2 and P6 wanted to switch the word cloud display on and off according to their needs.

Situation where the system is useful. P1, P3, P7 said that they would like to use our system in casual conversations with multiple people. P2 and P6 said that the word cloud helps them to passively listen to conversations, as in our experimental setup. P5 mentioned that he would like to use the proposed system in all kinds of situations, but if he had to choose, he would use it in multi-person conversations because it makes it easier to see who said what. P1, P3, and P5 worry that the current system, which uses real-time speech recognition, may frequently make errors in conversations with many technical terms, such as in a university class. In addition, slides are often shared on screen in class; thus, the compatibility with these slides should also be considered.

Optimal Placement. It is recommended to not hide the visuals behind the participants' faces, and that a word's position should be used to indicate how old or new the word is. Four participants were concerned about the word being hidden behind a person, making it partially invisible. It was suggested that the person be placed on the left side of the video and the word cloud on the right (P5, P7), or that the words be placed around the person (P3, P5). In addition, all participants except P6 wanted the words to be aligned horizontally only. Many participants asked that the freshness of the word be expressed by its layout, not by its color (P3, P5, P6, P7).

4 DISCUSSION AND CONCLUSION

Participants' subjective responses show that our system is useful for observing shifts in topics. Meanwhile, no significant difference was found between scenarios in all the questions. There was however a trend that frequent topic shifts make the system distracting (Q3 in the Likert questions). Frequent switching of topics may cause confusion in the word cloud, as words with small relevance get mixed. P6:"*It was easier to grasp the topic when keywords that could be easily connected to each other were presented, but it was confusing when unrelated keywords appeared simultaneously.*"

The purpose of this study was to investigate how a user's perception of the word cloud varies depending on how the topic changes. Therefore, a comparative experiment with existing technologies is the subject for a future study.

It should also be noted that the current system design requires the cooperation of all participants. P3 said that if the word cloud is distracting for participants who can hear, it would be difficult to ask for help. Therefore, further research is needed to determine how word clouds affect people without hearing issues.

REFERENCES

[1] [n.d.]. Clear, compelling communication for everyone | mmhmm. https://www.mmhmm.app/. (Accessed on 06/23/2021).

[2] [n.d.]. D3.js - Data-Driven Documents. https://d3js.org/. (Accessed on 06/23/2021).

[3] [n.d.]. takuyaa/kuromoji.js: JavaScript implementation of Japanese morphological analyzer. https://github.com/takuyaa/kuromoji.js/. (Accessed on 06/23/2021).

[4] [n.d.]. Web Speech API - Web API | MDN. https://developer.mozilla.org/ja/docs/Web/API/Web_Speech_API. (Accessed on 06/23/2021).

[5] Bon Adriel Aseniero, Marios Constantinides, Sagar Joglekar, Ke Zhou, and Daniele Quercia. 2020. MeetCues: Supporting Online Meetings Experience. arXiv:2010.06259 [cs.HC]

[6] Vannevar Bush. 1996. As We May Think. *Interactions* 3, 2 (March 1996), 35–46. https://doi.org/10.1145/227181.227186

[7] Anna C. Cavender, Jeffrey P. Bigham, and Richard E. Ladner. 2009. ClassInFocus: Enabling Improved Visual Attention Strategies for Deaf and Hard of Hearing Students. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, Pennsylvania, USA) *(Assets '09)*. Association for Computing Machinery, New York, NY, USA, 67–74. https://doi.org/10.1145/1639642.1639656

[8] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. 2019. TalkTraces: Real-Time Capture and Visualization of Verbal Content in Meetings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300807

[9] Jason Davis. [n.d.]. jasondavies/d3-cloud: Create word clouds in JavaScript. https://github.com/jasondavies/d3-cloud. (Accessed on 06/23/2021).

[10] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. https://doi.org/10.1016/S0166-4115(08)62386-9

[11] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People Who Are Deaf and Hard of Hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3234695.3236362

[12] Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2019. Evaluating the Benefit of Highlighting Key Words in Captions for People Who Are Deaf or Hard of Hearing. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 43–55. https://doi.org/10.1145/3308561.3353781

[13] Kuno Kurzhals, Fabian Göbel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. 2020. A View on the Viewer: Gaze-Adaptive Captions for Videos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376266

[14] Raja S. Kushalnagar, Gary W. Behm, Aaron W. Kelstone, and Shareef Ali. 2015. Tracked Speech-To-Text Display: Enhancing Accessibility and Readability of Real-Time Speech-To-Text. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 223–230. https://doi.org/10.1145/2700648.2809843

[15] Harry G. Lang. 2002. Higher Education for Deaf Students: Research Priorities in the New Millennium. *The Journal of Deaf Studies and Deaf Education* 7, 4 (10 2002), 267–280. https://doi.org/10.1093/deafed/7.4.267 arXiv:https://academic.oup.com/jdsde/article-pdf/7/4/267/9835869/267.pdf