

# A Preliminary Study on Understanding Online Meetings Using Emoji-based Captioning for Deaf or Hard of Hearing Users

Kotaro Oomori

The use of real-time captioning systems in online meetings has increased. Speech-to-text is a technology used in real-time captioning systems that takes the user's voice as input and the text conversion as output. This system is particularly useful for communication in groups that include Deaf or Hard of Hearing (DHH) people, where sign language and written communication have been used. In this thesis, two experiments to verify the possibility of using emojis in a real-time captioning system are conducted. The first experiment is a preliminary experiment for measuring the effect of emoji-based captioning system on understanding voice-only online meetings for DHH users. In this study, four DHH participants were recruited. The results show that emoji-based captioning system do not have a significant advantage over the baseline system without emoji. It was also found that the use of emoji-based captioning system was not preferred by DHH users in the scope of this preliminary experiment. The concern of miss-recognition by the automatic emotion recognition leading to discommunication was cited as a reason. The second experiment was compared two emoji-based captioning systems that selected emojis based on speech emotion recognition or selected emojis randomly. In this study, four DHH participants and four hearing participants were recruited. It was the hypothesis of the study that DHH and hearing people can enjoy communicating with each other by inserting emojis even if that are not based on any particular context into the spoken text. As a result, there was no significant difference between the condition using randomly selected emojis and the condition using emojis selected based on speech emotion recognition. Also, the scores on the question "Did you enjoy the communication?" were same in both of conditions. 75 percent of the participants gave the highest score on a five-point Likert scale, and 25 percent of the participants gave the second highest rating. According to the participants' comments, a hearing participant felt that it was better to have a lot of emoji, even if it was not always appropriate. On the other hand, there were also participants who felt that their intentions were misunderstood by the emojis or that the gap between the content of the subtitles and the emojis was uncomfortable.

(Advisor: Yoichi Ochiai, Tatsuki Fushimi)