Bubble Visualization Overlay in Online Communication for Increased Speed Awareness and Better Turn Taking

Reiya Horii, Yurike Chandra, Kai Kunze, Kouta Minamizawa
Graduate School of Media Design, Keio University, Yokohama, Japan
{rhorii018, yurikemch, kai, kouta}@kmd.keio.ac.jp

ABSTRACT
In this paper, we explore the use of a real-time speech visualization overlay to help native English speakers (NES) reflect on their speech speed and allow them to understand how English as a Foreign Language speakers (EFLS) perceive their speech. Our visual system generates a sequence of bubbles based on the speaking speed overlaid close to the user's mouth; the faster the speaking speed, the denser the bubble sequence, and vice versa. The results suggest that the presence of the speech visualization helps NES to understand their speech speed, and subsequently, it helps EFLS to feel comfortable speaking during the online group discussion.

Author Keywords
Cross-cultural Communication, Speech Visualization, Self-Reflection, Behavioral/Perception Change

CCS Concepts
•**Human-centered computing** → **Interaction design;** *Empirical studies in interaction design ;* Systems and tools for interaction design;

INTRODUCTION
Understanding an English discussion in an online setting comes with its challenges for English as Foreign Language speakers (EFLS). For instance, when native English speakers (NES) speak too fast [8], EFLS would not be able to understand the conversation. While many past works have aimed to mitigate this problem, most of them focused on either (1) helping EFLS to improve their English conversation skills [4, 7], or (2) notifying NES of how they perform in the conversation. For instance, Duan et al. created a speedometer that informs NES of their speaking speed to help them slow down [3]. Another study used visualization to show conversation activity (e.g., turn-takings or utterance length) among speakers [6, 2, 5, 1]. Compared to these works, we focus on a real-time feedback overlaid in the video of the speakers rather than showing a separate visual gauge [2].

UIST '20 Adjunct, October 20–23, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7515-3/20/10.
https://doi.org/10.1145/3379350.3416185

Building on previous research, we aim to design a speech visualization system that helps NES to understand how EFLS perceive their speaking (Fig. 1).

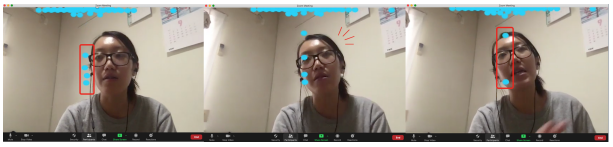

Figure 1. Speech Speed Visualization System

Our main contributions are (1) the development of speech speed visualization system, (2) the evaluation of the effect of perceiving this visualization to NES's speaking behavior, and (3) discussions on other observed benefits of this concept.

CONCEPT DESIGN
Speed is one of the most common factors that hinder EFLS to comprehend what NES speak. Furthermore, due to NES's familiarity with the way English works, they may unconsciously speak faster [2]. Hence, our concept is a visualization system that would help NES realize their speaking speed (Fig. 1).

Visual Design
Speaking speed can be calculated by dividing the total number of words by the number of minutes the speech took. In other words, faster speed will have more information conveyed in a shorter amount of time. Reflecting such a relationship between speed and information, we decided to visualize information density as an indicator of NES's speaking speed (Fig. 2).

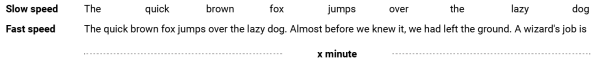

Figure 2. Correlation between speed and information density

We then used a sequence of bubbles to represent the information density. When NES speaks at a fast speed, the sequence of bubbles will become denser, and vice versa (Fig. 3).

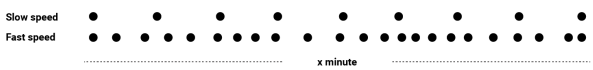

Figure 3. A sequence of bubbles represents information density

Technical Implementation
We used TouchDesigner, a visual programming language [1], to create our visualization system. TouchDesigner would take in

[1] https://derivative.ca/product

the sound input from the computer's microphone and calculate the frequency. When the sound reached a threshold, it would generate a bubble through a particle node. We then connected TouchDesigner output to NDI Virtual Input [2], which streamed the visuals to an online video conference software, Zoom [3] (Fig. 4).

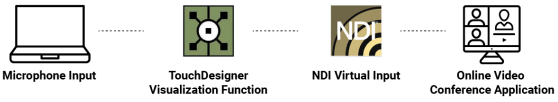

Figure 4. Components of Speech Speed Visualization System

USER STUDY

This study aimed to test our concept and see how visualizing speech speed affects NES's actual speaking speed. We used a within-subjects study design to compare an online discussion experience with two experimental conditions: (1) with visualization and (2) without visualization. We hypothesized that if the visualization helps NES understand their speaking speed, they would speak slower during the session with visualization than the session without visualization. In total, 13 adult participants (4 Japanese - EFLS; 9 English Speakers - NES) took part in this study. Participants were briefed on the experiment and signed a consent form.

Procedure

We divided our participants into 4 groups with 1 Japanese participant each and 2 to 3 English speaking participants. We invited them to join a Zoom meeting with the speech bubble visualization installed on the English speaking participant's video (See Fig. 5).

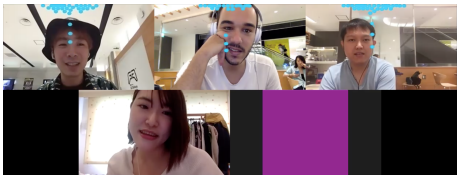

Figure 5. Online Discussion Setting

Then, they went through 2 discussions with two different conditions (10 minutes each). A 5-minute break was given in between the discussions (See Fig. 6).

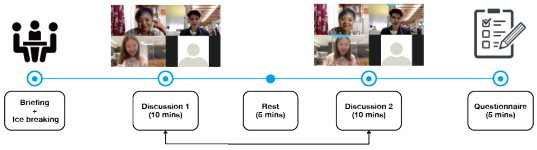

Figure 6. Experiment Procedure

The appearance of the conditions in the discussion was randomized. Before the study began, participants were told that the generated bubble represented their speaking speed. They were also given time to familiarize themselves with one another through a 5-minute ice-breaking session. The selected topics were open-ended questions that allowed participants to speak easily without the need to have prior knowledge. At the end of the experiment, participants were asked to fill up some questionnaires regarding the experience.

[2]https://ndi.tv/tools/
[3]https://zoom.us/

RESULTS

Data Analysis

This study had two different conditions (visualization and no visualization) and one dependent variable (speech speed in words per minute). A Friedman test was carried out to see if there was any difference in NES's speech speed based on the visualization (See Fig. 7).

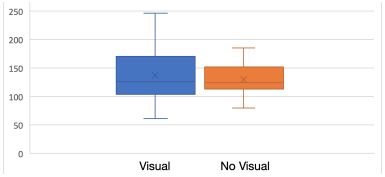

Figure 7. Mean and Standard Deviation of Speech Speed

There was no significant difference in the speaking speed ($X^2(2) = 0.1111$, p = 0.73888) depending on the visualization.

Qualitative Analysis

While there is no significant difference (based on the data analysis) in the speech speed, 5 out of 9 NESs mentioned that they were aware of their speech speed because of the visualization. From the recorded video, participants can be seen prompting a question, "What do you think of this?" and giving speaking chances to others during visuals sessions. This gesture suggested that visualization improved the turn-taking in a discussion and promoted equal participation rate among participants. 6 out of 9 participants also mentioned that it was more comfortable to discuss with visuals, implying that the system can lighten the mood for first-time meetings. Finally, the visualization also helped both NES and EFLS to pay more attention to the speakers.

DISCUSSION AND FUTURE WORK

The qualitative evaluation suggested that the visualization helped most NES to understand their speech speed. However, they were still unable to modify their speaking rate, which caused the Friedman test to show no significant difference in the speech speed for both conditions. A similar observation was found in the previous research when using speech speedometer [3], which implied that NESs require cues (e.g., timing to slow down, ideal speech rate) to be able to adjust their speaking behavior. Moreover, in the interview, many NESs mentioned that when they talk, they tend to look at the listeners' screen to see if they understand. This behavior further highlights the importance of listeners' feedback (e.g., head-nodding, confused face) in enabling speakers to change their speech rate. Based on these findings, we hope to integrate listeners' responses in our next iteration of the speech visualization system. Additionally, we would like to implement psycho-physiological measurements such as head-nodding or eye-blink rate to gain a comprehensive assessment of the participants' level of understanding during the discussion.

REFERENCES

[1] Mark Billinghurst, Kunal Gupta, Masai Katsutoshi, Youngho Lee, Gun Lee, Kai Kunze, and Maki Sugimoto. 2016. Is it in your eyes? Explorations in using gaze cues for remote collaboration. In *Collaboration Meets Interactive Spaces*. Springer, 177–199.

[2] Joan Morris DiMicco, Katherine J. Hollenbach, and Walter Bender. 2006. Using Visualizations to Review a Group's Interaction Dynamics. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. Association for Computing Machinery, New York, NY, USA, 706–711. DOI: http://dx.doi.org/10.1145/1125451.1125594

[3] Wen Duan, Naomi Yamashita, and Susan R. Fussell. 2019. Increasing Native Speakers' Awareness of the Need to Slow Down in Multilingual Conversations Using a Real-Time Speech Speedometer. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 171 (Nov. 2019), 25 pages. DOI: http://dx.doi.org/10.1145/3359273

[4] Shiro Kitayama and Kazushi Nishimoto. 2017. DAFlingual:A Supporting Method to Improve English Conver-sation Skill by Applying Delayed Auditory Feedback. *IPSJInteraction* (2017), 894–896. http://hdl.handle.net/10119/15113

[5] Katsutoshi Masai, Kai Kunze, Maki Sugimoto, and Mark Billinghurst. 2016. Empathy glasses. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1257–1263.

[6] Pooja Mathur and Karrie Karahalios. 2009. Visualizing Remote Voice Conversations. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*. Association for Computing Machinery, New York, NY, USA, 4675–4680. DOI: http://dx.doi.org/10.1145/1520340.1520719

[7] Takeshi Nishida. 2014. Promoting Intercultural Awareness through Native-to-Foreign Speech Accent Conversion. In *Proceedings of the 5th ACM International Conference on Collaboration across Boundaries: Culture, Distance Technology (CABS '14)*. Association for Computing Machinery, New York, NY, USA, 83–86. DOI: http://dx.doi.org/10.1145/2631488.2634058

[8] Yohtaro Takano and Akiko Noda. 1993. A Temporary Decline of Thinking Ability During Foreign Language Processing. *Journal of Cross-Cultural Psychology* 24, 4 (1993), 445–462. DOI: http://dx.doi.org/10.1177/0022022193244005