A Multimodal Corpus of Expert Gaze and Behavior during Phonetic Segmentation Tasks

Arif Khan\textsuperscript{1,3}, Ingmar Steiner\textsuperscript{1,2}, Yusuke Sugano\textsuperscript{4}, Andreas Bulling\textsuperscript{1,5}, Ross Macdonald\textsuperscript{6}

\textsuperscript{1}Multimodal Computing and Interaction, Saarland University
\textsuperscript{2}DFKI Language Technology Lab, Saarbrücken
\textsuperscript{3}Saarbrücken Graduate School of Computer Science
\textsuperscript{4}Osaka University, Japan
\textsuperscript{5}Max Planck Institute for Informatics, Saarbrücken, Germany
\textsuperscript{6}University of Manchester, UK

1 Introduction

- Phonetic segmentation is the process of splitting speech into distinct phonetic units.
- Automatic phonetic segmentation should replicate the precision of human segmentation as closely as possible, but segmentation behavior data is scarce.
- This corpus captures human segmentation behavior by recording phonetician’s gaze with an eyetracker, along with other relevant modalities.

2 Data recording

- Segmentation of a 46 s audio recording (48 kHz sampling rate).
- The audio was segmented using the \textit{Praat} software.
- Segmenter gaze was recorded with a Tobii TX300 eyetracker.
- Playback audio, webcam video, and screen contents were also recorded.
- The \textit{Praat} UI state and final manual segmentation were also stored.

3 Gaze on spectrogram

- Spectrogram of the utterance “The North Wind and the Sun”.
- Fixations are rendered as red circles.
- Blue lines represent the manual boundaries placed by the participant.

4 Fixation regions

- Percentage of fixations across sections.

5 Segmentation progress

- Speech segmentation data (s) against session duration (normalized).

6 Conclusion

- Recorded a multimodal corpus of segmentation behavior data from phonetic experts.
- All relevant information sources were recorded, e.g., gaze, playback audio, video, and screen recording.
- The processed data is released under a Creative Commons license and publicly available on GitHub: https://git.io/eyeseg-data.

7 Acknowledgements

- We are grateful to our participants for their time and valuable feedback.
- This project was funded by the German Research Foundation (DFG) under grant EXC 284.