




ChunkitApp 2.0, a speech segmentation app

Anna Galkova¹, Alena Konina¹, and Anna Mauranen¹

¹ University of Helsinki

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: Samuel Forbes 

Reviewers:

- [@isdanni](#)
- [@wmvanvliet](#)

Submitted: 01 April 2025

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

ChunkitApp 2.0 is a data collection platform for speech segmentation experiments. It allows researchers to design and run their experiments, as well as pre-process their data. It initially grew out of the Chunking in Language project (2017-2021, University of Helsinki, Mauranen (2012)) and was later featured in our group's research (Vetchinnikova et al. (2017), Vetchinnikova et al. (2022), Dobrego et al. (2023), Vetchinnikova et al. (2023)).

Statement of need

It was designed to present participants with a series of speech extracts accompanied with their transcripts. Participants are then able to segment the orthographic transcript for each extract by tapping an interactive symbol between tokens (less strictly speaking, words) according to experiment instructions. At the back-end, the tapped symbol is recorded as a segment boundary. If participant taps the symbol twice, the boundary is removed both from the interface and the data log. The app was initially designed for collecting the most naturalistic data possible, so we used tablets as a medium for running the experiment. However, it can also run on a laptop or PC with a mouse.

Apart from raw chunking data, as we call it, which includes the position and timestamp of each segment boundary marked by individual participants, the ChunkitApp 2.0 aggregates the data across participants and extracts. Moreover, it can run Monte Carlo simulations to ascertain the degree of randomness in segmentation behaviour. Finally, researchers can include comprehension questions after each extract, collect background data on participants and their feedback, and offer participants an English proficiency test.

The ChunkitApp features two tracks: Design & Run and Fetch & Analyse. Design & Run lets researchers construct their experiment using pre-programmed blocks and generates a link that allows participants to take part in it. If necessary, integration with Prolific is built into the app. The data is collected and saved on the server. Fetch & Analyse is the part of the app where researchers can track data collection and download raw and aggregated data.

The development of the app was funded by the Swedish Cultural Foundation in Finland (grant number 157452). The authors declare that there is no conflict of interest pertaining to the development of the app.

Community guidelines

If you wish to contribute to the ChunkitApp 2.0, you need to fork the initial repository. For publications and talks, see Citations on how to cite.

If you encountered a bug, you can report an issue through Github or seek support at alena.konina@helsinki.fi

Citations

To cite the app, please use the following from the References section (to be updated to a journal reference when published) - Galkova, Konina, and Mauranen (2025)

Acknowledgments

We would like to thank Aleksandra Dobrego, Svetlana Vetchinnikova, Michal Josífko, and Nina Mikušová for their contributions to the initial versions of the app.

References

- The app was modelled after the original ChunkitApp used for collecting data in the CLUMP project. Below you can find a list of publications featuring data collected through the original app.
- Dobrego, A., Konina, A., & Mauranen, A. (2023). Continuous speech segmentation by L1 and L2 speakers of English: The role of syntactic and prosodic cues. *Language Awareness*, 32(3), 487–507.
- Mauranen, A. (2012). Linear unit grammar—a real-time grammar. *The Encyclopedia of Applied Linguistics*, 1–11.
- Vetchinnikova, S., Konina, A., Williams, N., Mikušová, N., & Mauranen, A. (2022). Perceptual chunking of spontaneous speech: Validating a new method with non-native listeners. *Research Methods in Applied Linguistics*, 1(2). <https://doi.org/10.1016/j.rmal.2022.100012>
- Vetchinnikova, S., Konina, A., Williams, N., Mikušová, N., & Mauranen, A. (2023). Chunking up speech in real time: Linguistic predictors and cognitive constraints. *Language and Cognition*, 15(3), 453–479. <https://doi.org/10.1017/langcog.2023.8>
- Vetchinnikova, S., Mauranen, A., & Mikušová, N. (2017). ChunkitApp: Investigating the relevant units of online speech processing. *Proceedings of INTERSPEECH 2017*. <https://doi.org/10.21437/Interspeech.2017>