

Readalongs: Automatic alignment of speech and text for Indigenous language audiobooks

David Huggins-Daines¹, Patrick Littell²
Aidan Pine², Eric Joanis²

¹Nuance Communications

²National Research Council of Canada

Abstract

Interactive read-along or sing-along activities, that highlight words as they are spoken and allow students to click on words to hear them aloud, are well liked by both students and teachers, but are highly labour intensive to create manually, and require experience with specialized software such as ELAN or Audacity. This process can mostly be automated using speech recognition technology, but even the most widely spoken Indigenous languages in Canada are still considered extremely “low-resource” from the point of view of speech technology and remain unsupported due to the lack of data for training statistical models. There are also important considerations of privacy and cultural sovereignty associated with the collection of such data. Our open-source ReadAlong Studio combines a custom G2P engine, the PanPhon (Mortensen et al., 2016) phonetic distance library (for automatically determining cross-linguistic phoneme mappings), and the lightweight PocketSphinx (Huggins-Daines et al., 2006) speech recognition library to allow linguists and educators to create a text/speech alignment system for a new language without the usual requirement of labeled training data. ReadAlong Studio currently supports about 22 languages, including Anishinaabemowin, Atikamekw and East Cree. Adding a new language is the work of only a few hours, depending on the complexity of the language’s orthography. ReadAlong Studio can already create high-quality interactive webpages, MP4 movies, and EPUB documents, and can also export in ELAN, TextGrid (PRAAT), and subtitle formats. A user-friendly interface, with the goal of making a read-along as easy to make as a PowerPoint presentation, is currently in development.

Keywords— speech technology, phonology, education

References

- [1] David R. Mortensen, Patrick Littell, Akash Bharadwaj, Kartik Goyal, Chris Dyer, Lori Levin. PanPhon: A Resource for Mapping IPA Segments to Articulatory Feature Vectors. *Proceedings of COLING 2016*, Osaka, Japan.
- [2] David Huggins-Daines, Mohit Kumar, Arthur Chan, Alan W Black, Mosur Ravishankar, Alex I. Rudnicky. PocketSphinx: A Free, Real-Time Continuous Speech Recognition System for Handheld Devices. *Proceedings of ICASSP 2006*, Toulouse, France.