



WP 1.2 Speech processing and annotation

DELIVERABLES

D:1.2.1 Forced-alignment Service, D:1.2.2
Tranliteration Service for Finnish Interviews

The Data:

Donate Speech

The aim of the Donate Speech campaign is to collect all kinds of spoken Finnish as much as possible. The donated speech is used to develop speech recognition and artificial intelligence to better understand spoken Finnish. Speech can be donated using a computer browser or a mobile app. The already accrued one terabyte of speech data is used as foundation to create improved forced aligners and speech recognition systems.

Forced Alignment

Forced alignment refers to the process by which orthographic transcriptions are aligned to audio recordings to automatically generate phone level segmentation. We aim to provide an improved aligner based on Donate Speech data.

Speech Recognition

The aim is to provide a Finnish language speech recognition system based on Donate Speech data suitable for transcription of interviews. The system is based on speech models created at Aalto University.

Speaker Diarization

Speaker diarization means the automatic detection of different speakers in an audio file. Successful speaker diarization not only recognizes what was spoken but also detects the number of speakers and classifies utterances accordingly. This is relevant for both Alignment and Recognition, in this project especially for interview speech recognition.

CURRENT STATUS

- The raw Donate Speech data has been packaged and divided into testing and training data.
- A forced aligner and speech recognizer has been installed on CSC's Puhti Supercomputer. The system does not yet use Donate Speech data.
- Aalto provides the speech aligner/recognizer as a container which contains all relevant parts
- The installation process is automated and publicly available via GitHub, but not yet documented. It can be used to install the Aligner/Recognizer on a local Linux computer.
- A derived version of the recognizer runs in the CSC's Rahti cloud and can be used via API calls. Access control and billing is not yet supported.

OPEN QUESTIONS

Question 1. How likely will audio data be classed as sensitive data[1]? Interviews in Research might very well touch politics, health, sexual orientation.

Question 2. Sensitive audio data processing is officially not in scope. If most of real world use requires secure processing, how do we handle that?

Question 3. Sustainability: Can we ensure stability of APIs beyond the project?

MAIN COLLABORATORS

CSC, IT Center for Science, is a non-profit state enterprise with special tasks. As part of the national research system, it develops, integrates and provides high-quality information technology services and ensure that Finland remains at the forefront of development.

Aalto University Department of Signal Processing and Acoustics, Speech recognition

Our goal is to generally improve the speech recognition methodology with the help of the new algorithms developed in Aalto University. Speech recognition offers challenging benchmarking tasks for efficient algorithms that can process and learn to represent large quantities of data. In addition to improving the acoustic models of phonemes we aim at developing new learning statistical language models for difficult large vocabulary continuous speech recognition tasks.

University of Helsinki, Department of Digital Humanities

The department hosts FIN-CLARIN and has coordinated the collection and refinement of Donate Speech data. In the context of this project the department actively assists in making the speech processing tools available outside of CSC's supercomputing environment.

MAIN STAKEHOLDERS

All members of the FIN-CLARIAH consortium are requested to evaluate the iterations of the deliverables of this work package, especially:

- The Institute for the Languages of Finland (KOTUS)
- University of Oulu
- University of Eastern Finland
- The Faculty of Social Sciences, University of Helsinki
- The National Library of Finland

References

[1] <https://research.csc.fi/definition-of-sensitive-data>