# Educational pipeline module speech - text

# For MCS / Radboud AI

26 Aug. 2019, v1.1

Helmer Strik, Henk van den Heuvel (notes)

The idea is to make an educational / research environment where people can learn and experiment with the alignment of spoken word material with corresponding transcripts. The environment should consist of a pipeline set- up where this alignment is addressed and experienced in a series of task ranging from simple to more complex (offered in an accompanying labdoc).

Material

There are recordings for different types of speech, ready for use in the pipeline:

1. COPD
2. Children reading aloud
3. CHASING dysarthric speech
4. ISLA idiomatic expressions (Germans speaking Dutch)
5. MPC (Dutch children speaking English)

Types 1-4 have already gone through the pipeline, and there are manually corrected segmentations.

All these speech types can be considered atypical speech.

What has to be done is to select ‘normal’, ‘typical’ speech, that can be used as reference [benchmark], e.g. adult read speech from CGN. The reference speech should also go through the whole pipeline.

In terms of software there is:

* Lexica and G2P convertors for Dutch and English
* Forced alignment software for Dutch and English, yielding automatic segmentations
* Praat scripts for manually checking and correcting the segmentations
* Praat scripts to yield statistics per aligned file in terms of durations, pitch, intensity, formants, etc.
* Other [open source] software to analyze these segmented audio files, e.g. OpenSmile.
* Python scripts to export Praat results to an Excel file

.

Contents

What the educational module should contain as pipeline building blocks:

1. Arrange speech & orthographic transcriptions
   * Incl. conversion between audio and transcription formats
2. Forced alignment (FA) for Dutch and English
   * For aligning new material students must update the lexicon with the words in the transcription, making use of a G2P-convertor
3. Optional building block for manual correction of output FA
   * Using Praat-scripts for quickly processing the files
4. Analysis of statistics per aligned file in terms of durations, pitch, intensity, formants, etc. (Praat scripts)
5. Meta-analysis of the aforementioned statistics for selected files

Needs

* Tool to convert audio and transcription formats (secondary priority)
* Reference material [see above]
* More user friendly implementation of forced alignment tool
* Output of one part of the pipeline should be immediately accessible as input for the next: predictable location on disk, filenames and formats
* An environment that is suitable for students [e.g. Docker/VM]
* Manuals containing the subsequent tasks for the student must be written (Helmer and assistant)

Implementation

Series of pipelines chaining into each other in terms of output of one building block used as ready-made input for the next.

The first version will be realized on ponyland in such a way that it can be wrapped into a docker/VM

ICT personnel

Maarten de Vos, a student assistant working on his MSc thesis, is familiar with the material. Until November he is available for one day per week. After that he is interested in a junior position to complete the job. We have 3 PM available for the task.

Schedule

|  |  |
| --- | --- |
| Aug/Sep 2019 | Deciding on data to use |
|  |  |
| Sep/Nov. 2019 | Setting up the educational plan (things to learn) |
|  | Collecting & installing software packages needed |
|  | Set up specs for the building blocks with stud.ass |
|  |  |
| Dec 2019 / Feb 2020 | Implement module |
|  |  |
| March / April 2020 | Test and refine module |
|  |  |