GipHouse project:

Pipeline with forced alignment of speech and text

15 January 2020
Henk van den Heuvel , Helmer Strik

Outline

The core tool is Forced alignment (FA) of speech (audio) and text (the words that have been spoken). The result of FA is a segmentation at sound [phoneme] level.
In addition, there are several other tools that can be used [in a pipeline] before or after FA.
FA runs on Linux, the other tools run on MS Windows and/or Linux.
The goal is 1 environment for all the tools that can be accessed via MS Windows, with an accompanying lab-doc, to carry out experiments by students [education] and researchers.
The question is what the optimal environment is, e.g. virtual machine, docker, web portal, etc.

The project

Briefly, the procedure is as follows: The text [orthographic transcription, string of words, graphemes] is converted to a so called phonetic transcription [a string of sounds, phonemes], usually by using a lexicon. A tool should detect whether all words are present in the lexicon; if not, they should be added. FA aligns speech and sounds, resulting in an automatic segmentation at sound level. There are tools to manually check these segmentations, and if necessary, improve them. And then there are various tools that use these segmentations to calculate various measures.
The pipeline thus contains optional blocks, and input – output for the various tools should be organized well.
What the educational module should contain as pipeline building blocks are e.g.:

1. Arrange speech & orthographic transcriptions, incl. conversion of formats
2. Check the lexicon, and update of needed, e.g. using a G2P-convertor
3. Forced alignment (FA) for Dutch and English
4. Manual correction of the output of FA [the segmentations] using Praat-scripts
5. Analysis, using the segmentations with Praat or other [e.g. Python] scripts.
6. Meta-analysis for selected files, and export results [e.g. to Excel files].

It is important that input and output of the various toolkits are smoothly converted in the background so that the users do not need to bother, and the pipeline works as intuitively as possible.

Skills

Python programming skills, since many available scripts are in Python.
FA runs under Linux, other tools run on MS Windows and/or Linux. The whole architecture should be one easy to use environment that can be accessed via MS Windows, e.g. virtual machine, docker, web portal, etc. This is the most important aspect.
Dutch and English FA already exists, based on Kaldi (https://kaldi-asr.org/).
And there are many Praat scripts (https://www.fon.hum.uva.nl/praat/).
Expertise in the team [e.g. min. 1 person] on speech, Praat, Kaldi, would be good, but is not necessary.

Contact

Henk van den Heuvel (h.vandenheuvel@let.ru.nl) & Helmer Strik (w.strik@let.ru.nl)