

MSc Thesis Topics

Laboratory of Phonetics and Speech Technology
Institute of Cybernetics

Laboratory of Phonetics and Speech Technology

- Research in various areas of speech processing
- Development of Estonian speech recognition open source tools
 - Android app *Kõnele*
 - Online speech recognition server <https://github.com/alumae/kaldi-gstreamer-server> (234 stars, 74 forks)
 - Interface to automatically transcribed Estonian radio news and talkshows <http://bark.phon.ioc.ee/tsab/p/index>
- Most research is language agnostic (not specific to Estonian), no knowledge of Estonian is needed

Topics

- Topic segmentation of speech
- Speech summarization
- Modelling (foreign) name pronunciations
- Speech activity detection
- Spoken language identification

Most topics involve deep neural networks (DNNs)

Topic segmentation of speech

Given a transcription of speech (produced using speech recognition):

- Automatically find places where topic changes (e.g. in a news broadcast or a radio talkshow)
- Possibly also provide section heading, summary or keywords
- Approach: train a recurrent neural network model on a corpus of manually segmented speech

Speech summarization

Given a transcription of speech (produced using speech recognition):

- Produce a coherent summary of the speech content
- Approach: use a recurrent neural network, train in on manually produced summaries

Modelling (foreign) name pronunciations

Person names, and especially foreign names are very difficult for current speech recognition systems

- New names enter into mainstream use every day e.g. politicians, sportsmen, company names, products
- Current ASR system replace the previously unseen name with something similar
- Foreign names especially complex: e.g. Kurt Cobain
 - Is it a foreign name?
 - If foreign, which language rules to use to pronounce it?
 - Maybe it's foreign, but pronounced according to Estonian rules?

Speech activity detection

Given a long recording, find segments that contain speech

- More complex than it sounds
- Noise sometimes “looks like” speech
- Speech sometimes “looks like” silence
- Music, background noise, make things complicated etc

Approach: deep neural networks

Usage: given a long audio recording containing speech, silence, music, apply speech recognition only to the segments that contain speech

Spoken language identification

Classify speech in a short or long segment according to the spoken language

- Usage: use appropriate speech recognition system to transcribe a speech segment
- Use a training dataset with speech in different languages
- Train a (recurrent) neural network