Expose Bachelorarbeit - Augmenting accuray of automatic lecture transcriptions by using context information from slide contents

Jonathan Werner

Abstract

Automatic speech recognition (ASR) of university lectures is an important topic because of two main reasons: 1) providing accessibility for deaf people in the form of subtitles and 2) making the spoken text searchable, thus making it feasible to skim the content efficiently.

The main problem concerning ASR on academic lectures is the error rate for technical terms. This paper explores the possibility of improving recognition accuracy by adapting the vocabulary model with words that are extracted from the speakers slides.

Motivation

Scannability is crucial for academic research – but audio/audio-visual sources make this really hard right now. A concrete use case exists at the university of Hamburg, were we have the lecture2go system which makes video recordings of many lectures available. It would be a great benefit for students that want to rewatch the lectures in preperation for an exam to have searchable transcriptions of the lectures. If the results of the augmented transcriptions are good enough they could be rather easily integrated into the lecture2go website and provide an actual benefit to students.

During a project at the University Hamburg in the winter semester 14/15 we implemented a system which automatically generated subtitles for a given audio/video and a transcription. The system that I plan to implement for this thesis will integrate with this project: it would supply the missing first step in the toolchain: automatically generating the transcriptions (For the project we only used human generated transcriptions).

Work schedule

I would start by testing if and how the open source tool 'CMU Spinx' is able to adapt the recognition process in favor of specific set of words

Gliederung, konkrete Kapitelnamen

Absatz, der konkret benennt, wie die Kapitel aufeinander aufbauen

Literatur

- Cerva et al. (2012)
- Maergner, Waibel, and Lane (2012)
- Miranda, Neto, and Black (2013)
- Kawahara, Nemoto, and Akita (2008)
- Cho et al. (2013)

Bibliographie

Cerva, Petr, Jan Silovsky, Jindrich Zdansky, Ondrej Smola, Karel Blavka, Karel Palecek, Jan Nouza, and Jiri Malek. 2012. "Browsing, Indexing and Automatic Transcription of Lectures for Distance Learning." In *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*, 198–202. IEEE.

Cho, Eunah, Christian Fügen, Teresa Herrmann, Kevin Kilgour, Mohammed Mediani, Christian Mohr, Jan Niehues, et al. 2013. "A Real-World System for Simultaneous Translation of German Lectures." In *INTERSPEECH*, 3473–77.

Kawahara, Tatsuya, Yusuke Nemoto, and Yuya Akita. 2008. "Automatic Lecture Transcription by Exploiting Presentation Slide Information for Language Model Adaptation." In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 4929–32. IEEE.

Maergner, Paul, Alex Waibel, and Ian Lane. 2012. "Unsupervised Vocabulary Selection for Real-Time Speech Recognition of Lectures." In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 4417–20. IEEE.

Miranda, Joao, Joao Paulo Neto, and Alan W Black. 2013. "Improving ASR by Integrating Lecture Audio and Slides." In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 8131–35. IEEE.