

Osman Büyük
Thesis Supervisor: Prof. Levent M. Arslan

Telephone-based Text-dependent Speaker Verification

In this thesis, we investigate model selection and channel variability issues on telephone-based text-dependent speaker verification applications. Due to the lack of an appropriate database for the task, we collected two multi-channel speaker recognition databases which are referred to as text-dependent variable text (TDVT-D) and text-dependent single utterance (TDSU-D). TDVT-D consists of digit strings and short utterances in Turkish and TDSU-D contains a single Turkish phrase.

In the TDVT-D, Gaussian mixture model (GMM) and hidden Markov model (HMM) based methods are compared using several authentication utterances, enrollment scenarios and enrollment-authentication channel conditions. In the experiments, we employ a rank-based decision making procedure. In the second set of experiments, we investigate three channel compensation techniques together with cepstral mean subtraction (CMS): i) LTAS filtering ii) MLLR transformation iii) handset-dependent rank-based decision making (H-rank). In all three methods, a prior knowledge of the employed channel type is required. We recognize the channels with channel GMMs trained for each condition. In this section, we also analyze the influence of channel detection errors on the verification performance.

In the TDSU-D, phonetic HMM, sentence HMM and GMM based methods are compared for the single utterance task. In order to compensate for channel mismatch conditions, we implement test normalization (T-norm), zero normalization (Z-norm) and combined (i.e., TZ-norm and ZT-norm) score normalization techniques. We also propose a novel combination procedure referred to as C-norm. Additionally, we benefit from the prior knowledge of handset-channel type in order to improve the verification performance. A cohort-based channel detection method is introduced in addition to the classical GMM-based method. After the score normalization section, feature domain spectral mean division (SMD) method is presented as an alternative to the well-known CMS. In the last set of experiments, prosodic (energy, pitch, duration) and spectral features are combined together in the sentence HMM framework.

PUBLICATIONS

Journals

- 1) **Osman Büyük**, Levent M. Arslan, “Model selection and score normalization for text-dependent single utterance speaker verification”, *Turkish Journal of Electrical Engineering and Computer Sciences 2011*. (accepted for publication)

Conferences

- 2) **Osman Büyük**, Levent M. Arslan, “HMM-based text-dependent speaker recognition with handset-channel recognition”, *Proc. of the IEEE Eighteenth Signal Processing and Communication Applications Conference 2010 (SIU 2010)*, Diyarbakir, Turkey.
- 3) Ali Haznedaroglu, Levent M. Arslan, **Osman Büyük**, Mustafa Erden, “Turkish LVCSR system for call center conversations”, *Proc. of the IEEE Eighteenth Signal Processing and Communication Applications Conference 2010 (SIU 2010)*, Diyarbakir, Turkey.
- 4) Oytun Turk, **Osman Büyük**, Ali Haznedaroglu, Levent M. Arslan, “Application of voice conversion for cross-language rap singing transformation”, *Proc. of the IEEE*

International Conference on Acoustics, Speech and Signal Processing 2009 (ICASSP 2009), Taipei, Taiwan.

- 5) **Osman Buyuk**, Ali Haznedaroglu, Levent M. Arslan, “Turkish speech recognition software with adaptable language model”, *Proc. of the IEEE Fifteenth Signal Processing and Communication Applications Conference 2007 (SIU 2007)*, Eskisehir, Turkey.
- 6) Ali Haznedaroglu, **Osman Buyuk**, Levent M. Arslan, “Keyword spotting using keyword adapted language model”, *Proc. of the IEEE Fifteenth Signal Processing and Communication Applications Conference 2007 (SIU 2007)*, Eskisehir, Turkey.
- 7) Hakan Erdogan, **Osman Buyuk**, Kemal Oflazer, “Incorporating language constraints in sub-word based speech recognition”, *Proc. of the IEEE Automatic Speech Recognition and Understanding Workshop 2005 (ASRU 2005)*, Cancun, Mexico.
- 8) **Osman Buyuk**, Hakan Erdogan, Kemal Oflazer, “Using hybrid lexicon units and incorporating language constraints in speech recognition”, *Proc. of the IEEE Thirteenth Signal Processing and Communication Applications Conference 2005 (SIU 2005)*, Kayseri, Turkey.

Defense Jury Members

Prof. Levent M. Arslan

Assoc. Prof. Burak Acar

Assist. Prof. Hakan Erdođan

Assoc. Prof. Engin Erzin

Assoc. Prof. Murat Saraçlar

Bogazici University

Bogazici University

Sabanci University

Koc University

Bogazici University

Defense Date: 09.06.2011