

# Harish Krishnamurthy

TITLE: "STUDY OF ALGORITHMS TO COMBINE MULTIPLE AUTOMATIC SPEECH RECOGNITION (ASR) SYSTEM OUTPUTS".

Abstract and MS Thesis defense

Date: 13th April 2009

Automatic Speech Recognition systems (ASRs) recognize word sequences by employing algorithms such as Hidden Markov Models. Given the same speech to recognize, the different ASRs may output very similar results but with errors such as insertion, substitution or deletion of incorrect words. Since different ASRs may be based on different algorithms, it is likely that error segments across ASRs are uncorrelated. Therefore it may be possible to improve the speech recognition accuracy by exploiting multiple hypotheses testing using a combination of ASRs. System Combination is a technique that combines the outputs of two or more ASRs to estimate the most likely hypothesis among conflicting word pairs or differing hypotheses for the same part of utterance. In this thesis, a conventional linear system combination technique called Recognized Output Voting Error Reduction (ROVER) is studied. A weighted voting scheme based on Bayesian theory known as Bayesian Combination (BAYCOM) is implemented. BAYCOM is derived from first principles of Bayesian theory, a classical pattern recognition technique. ROVER and BAYCOM use probabilities at the system level, such as performance of the ASR, to identify the most likely hypothesis. These algorithms arrive at the most likely word sequences by considering only a few parameters at the system level. The motivation is to develop newer System Combination algorithms that model the most likely word sequence hypothesis based on parameters that are not only related to the corresponding ASR but the word sequences themselves. Parameters, such as probabilities with respect to the hypothesis and ASRs are termed word level probabilities and system level probabilities, respectively, in the thesis. Confusion Matrix Combination is a decision model based on parameters at word level. Confusion matrix that consists of probabilities w.r.t word sequences are estimated during training. The system combination algorithms are initially trained with known word sequences during which the corresponding parameters are estimated. Then a validation set is run on the trained system combination algorithms and the performances are compared. The word sequences are obtained by processing speech from Arabic news broadcasts in the thesis. CMC outperforms BAYCOM and ROVER over the training set but, requires computation of a larger set of parameters. ROVER still proves to be a simple and powerful system combination technique and provides best improvements over the validation set.

Thesis Committee:

[1] John Makhoul, Northeastern & BBN

[2] Jennifer Dy, Northeastern

[3] Spyros Matsoukas, BBN.

