

Measuring and Compensating for the Effects of Speech Rate in Large Vocabulary Continuous Speech Recognition

Matthew A. Siegler

This talk describes a series of experiments that measure speech rate and that attempt to improve speech recognition accuracy for rapidly-spoken speech. Descriptions of several measures of speech rate are presented, and their advantages and disadvantages are discussed. Speech recognition results obtained using several compensation methods are compared to identify methods by which compensation for the effects of fast speech may yield the greatest improvement in recognition accuracy.

Very simple measures of speech rate such as the word rate or phone rate are found to be inappropriate for the detection of both long-term and short-term speech rate because they are sensitive to the lexical content of speech. In contrast, the phone duration percentile, a comparison of measured versus expected phone duration, is shown to be robust with respect to lexical content and consistent with previous findings about the statistics of long-term and short-term speech rate. Using this metric, speakers with a speech rate in the highest 30% are found to have a word error rate of 11.7%, for an increase of 44% over the 8.1% baseline.

The compensation techniques explored consist of modifications to four components of the recognition system: the models of the acoustical characteristics of speech sounds, the models of the HMM state-transition probabilities, the pronunciations of words in the dictionary, and the weight with which acoustic and linguistic evidence are combined. Optimizing the language weight reduced the word error rate of fast speech to 10.3%. Adapting the state-transition probabilities to fast speech reduced the word error rate for fast speech to 11.3%. Using one of the modified pronunciation dictionaries also reduced the word error rate of fast speech to 11.4%. The other techniques yielded little or no reduction in the word error rate.