Abstract

Current human evaluation of public speaking performance is cost-inefficient, may be unreliable, and yields variable provision of constructive feedback. The present thesis aimed to investigate whether machine-generated prosodic features could predict human holistic scores in descriptive-word public speaking assessment, towards research on the development of a system for automatic assessment of public speaking performances. Experiments have been performed on which prosodic features can be utilized as speaker representative features, and results suggest that prosodic features related to the pitch range and pitch profile of a speaker can be speaker representative, and can be used to discriminate between similar and different speakers. In addition, clustering of similar feature vectors has performed, using the mean Euclidean distance of same speaker comparison as a non-similarity threshold, suggesting that speaker representative feature vectors can be classified as similar using a distance measure as threshold. Furthermore, the classification and regression models, utilizing Support Vector Machines (SVM), show promising results when focusing on target data related to prosodic features, suggesting the feasibility of automatic prediction of human holistic scores on public speaking performances.