ACOUSTICS2008/225
Automatic data enhancement for language identification using voice generation

Aaron Lawson[a], Matthew Linderman[b], Michael Carlin[b] and Allen Stauffer[a]

[a]RADC, Inc., 525 Brooks Road, Rome, NY 13441, USA

[b]Air Force Research Laboratory, 525 Brooks Road, Rome, NY 13441, USA

Approaches to LID require very large sets of training (often greater than 6 hours per language) for accurate results. This study looks at ways of automatically reducing the amount of training data required to train a LID model, while maintaining or increasing accuracy. Initial experiments found that speaker density, i.e. the number of speakers per time unit, had a very dramatic influence on the accuracy of models (absolute increase of 15%). In order to accomplish the goal of increasing the number of speakers available for LID training without having to collect additional audio, the STRAIGHT algorithm was used to synthesize "novel" speakers for use in training language models for a LID system. The mean pitch and vocal tract length of the speaker in each LID training file was scaled to generate four additional voices per original speaker to artifically augment the training data. The resulting models yielded an improvement of 10% over the baseline system (from 80% to 90%). This study shows that automatically generated speakers have a great impact on the accuracy of LID models, in this case an initial training set of 15 minutes per language was augmented and performed comparably to a set trained with six hours.