Pairwise audio comparison for visualisation of mispronunciation

Amy V. Beeston, Mauro Nicolao and Thomas Hain
Department of Computer Science, University of Sheffield, UK
{a.beeston, m.nicolaos, t.hain}@dcs.shef.ac.uk

Overview
- Funded by ITSLanguage in the Netherlands, we develop training tools for Dutch students of English.
- Assessment of pronunciation quality in adolescent learners of foreign languages is a challenging task.
- Detecting common Dutch mistakes (frequently substituted speech sounds) is particularly difficult.

Audio comparison
- Parallel audio analysis quantify student/teacher acoustic dissimilarity on a phone-by-phone basis.
- The output is provided in a range of 0 (poor performance) to 100 (good).

Datasets
- Classroom Recordings: lesson of 193 items recorded by c. 500 students (age 12-18) in 8 schools across The Netherlands.
- Syllabic Testset: Articulation Index Corpus (AIC, LDC2005S22) is used for preliminary tests.

System output scores
- System score distributions for Syllabic Testset. Correctly spoken phones (blue), mispronounced (red).

System duration analysis for Classroom Recording corpus.

Common mistakes
- Common mistakes were defined in an inventory of phones frequently substituted by native Dutch learners of English.

Objective assessment
- Cross-correlation (CC) [Witt & Young, 2000, Speech Comm. 30] quantifies correlation between ground truth reference, \( x_{ref} \), and the automatic system scores, \( x_{test} \).

Results
- Objective evaluation for Syllabic Testset corpus.

Conclusions
- The framework for analysing student recordings was defined, and test sets were devised to examine mispronunciation in the difficult problem of common Dutch mistakes.
- Good-pronunciation comparisons had relatively low acoustic difference values; distribution shifted to the right (larger difference) for bad-pronunciation comparisons.
- System performance was broadly consistent with the reference pronunciation using objective assessment metrics for cross-correlation, agreement, and normalised cross entropy.
- The Syllabic Testset provided a useful test bed for investigating thresholds in the signal analysis methods presented, and revealed appropriate parameter regions for reliable operation of the visualisation software.
- This work provides a baseline from which to develop tools for automatic assessment of student pronunciation in the Dutch classroom recordings.