Computed-Aided Evaluation of Pronunciation and Prosody for Language Learning

Zhao Sixuan
PhD Student
Institute of Media Innovation
School of EEE, NTU

Supervisor: Koh Soo Ngee
Concept of CALL And Prosody

• Computer Aided Language Learning: to train language learners with assistance of computer software.

• Prosody: the super-segmental feature which combines of stress, rhythm and intonation of speech.

• Function of Prosody in CALL system: prosody can be evaluated to reflect stress, emotion and nativeness of learners’ utterance.
Basic Steps of Prosody Evaluation

1. Feature Extraction: to extract important prosody features from continuous speech for processing.

2. Forced Alignment: to align the input utterance to transcription. This step can enable detailed processing for each part of speech.

3. Evaluation of prosody: to evaluate the prosody with the help of mathematical models, e.g. HMM or DTW.

4. Machine Scoring: to score the speech according to the result in evaluation step.
Flowchart of Evaluation

Learner’s Speech

Feature Extraction
- Rhythm Score Evaluation
- Intonation Score Evaluation

Forced Alignment
- Database of Teachers’ Speech
- Word Boundary Information

Rhythm Score
Intonation Score
Scoring

• Human-Human Correlation: native speakers are invited to score the utterance of language learners and the correlation between scores given by different evaluators are calculated as references. Normally, a 5 score standard is adopted with 5 level of the pronunciation of the speakers, from 1 (very poor) to 5 (excellent).

• Human-Machine Correlation: automatic evaluation system scores the utterance of each sample, and then the scores given by machine are compared to that given by human to calculate the correlation to indicate the effectiveness of the system. Higher correlation indicates better evaluation system.