Interactive Language Learning for Chinese Learners of English using Automatic Speech Recognition for Mispronunciation Detection and Diagnosis

Helen Meng,[1,2] Pauline Lee,[3] Lan Wang,[2] Alissa Harrison[1] and Wai-Kit Lo[1,2]

[1]*Human-Computer Communications Laboratory, The Chinese University of Hong Kong*
[2]*CUHK-CAS Shenzhen Institute of Advanced Technology*
[3]*Independent Learning Center, The Chinese University of Hong Kong*

We describe ongoing work that aims to develop automatic speech recognition technologies that cater for interactive language learning for Chinese learners of English. Our target user group consists of adult learners whose primary language (L1) is Chinese (including either the Cantonese or Putonghua dialects) and secondary language (L2) is American English. We began with a survey of existing commercial off-the-shelf tools that support English learning and analyzed their relative merits in terms of perceptual and productive training, as well as the importance of interactivity. In order to support interactive computer-aided pronunciation training (CAPT), we devised a linguistically-motivated approach for automatic speech recognition. This approach is grounded on the theory of language transfer and involves a systematic phonological comparison between L1 and L2 at the phonetic, phonotactic and prosodic levels. Major disparities across the language pair are identified as focal points where prominent language transfer effects may occur, which may in turn lead to pronunciation variants that constitute salient mispronunciations. Unlike conventional ASR systems that aim to accurately recognize words despite accentual variations, we attempt to develop a speech recognizer that recognizes the pronunciation variants (or mispronunciations) given a sequence of words. Our presentation will include a series of short talks that describe our approach, which includes a contrastive phonological analysis between L1 and L2 to identify prominent language transfer effects that may cause mispronunciations; automatic generation of pronunciation variants using context-sensitive phonological rules; data-driven pruning for elimination of implausible pronunciation variants; as well as the development of a speech recognizer that can detect and diagnose pronunciation variants from a learner's speech. We will also demonstrate our prototype system, Chelsea, which is intended to support Cantonese learners of English.