Abstract

Speech assessment is one of the key tasks in Computer Assisted Pronunciation Training (CAPT). In this study, we develop a mispronunciation assessment system that checks the pronunciation of non-native English speakers and identifies which phonemes were pronounced incorrectly. Thus, this work mainly focuses on automatic estimation of the common mispronounced phonemes of Italian learners of English, both adults, and children, and presents an analysis and evaluation of the native and non-native pronunciation observed in phonetically annotated speech corpora. For this, we design an Automatic Speech Recognition (ASR) system that evaluates learners' speech at the phoneme level to detect pronunciation errors. For mispronunciation assessment, we select two corpora (Interactive Spoken Language Education (ISLE) and ChildEn), create a new list of phonemes extracted from manually annotated data from both corpora, then train an acoustic model with the new phoneme set. These evaluations rely on ASR, which could be performed using a database of non-native speakers of frequent mistakes.

In detail, two language models were considered; an n-gram model and an error language model. The output of the n-gram model helped us obtain insights into the phone errors in our audio files. Applying the error language model is the novel approach to the second language (L2) speech assessment, which combines both true transcriptions for a word with their adapted transcriptions for the same word. Quantitative evaluations are carried out for Italian learners of English. The output of the preliminary results reveals that Italian learners of English have the most problems in producing certain phonemes that do not appear in their language. Our results show that the selected error model can discriminate correct sounds from incorrect sounds in both native and nonnative speech, and therefore can be used to detect pronunciation errors in non-native speech. The Phone error rates show improvement in using the error language model. In summary, our ASR system shows high accuracy after applying the error language model on our selected corpora.