Modeling Difficulties of Second Language Learners using Speech Technology

Tatsuya Kawahara
Kyoto University, Japan
kawahara@i.kyoto-u.ac.jp

ABSTRACT

Modeling difficulty for non-native speakers is a central issue in either listening or speaking training of foreign languages. This talk will address our recent findings in this issue for both listening and speaking training.

For listening training, we have proposed a novel scheme of Partial and Synchronized Captioning (PSC), which provides part of caption texts in sync with speech word by word (see Figure 1). This is intended for learners to focus on listening to audio rather than reading the caption texts by providing minimum help. The key issue is to select caption texts to be presented based on difficulty for non-native learners, which is variable according to their proficiency level. While there are a number of factors affecting difficulty such as speaking rate, word frequency and specificity, we exploit errors made by an automatic speech recognition (ASR) system. Here, we assume that difficult words for human are also difficult for ASR systems, but all errors are not necessarily useful. Therefore, we investigate useful error patterns and identified the following four categories: homophones, minimum pairs, negatives, and bleached error boundaries. By incorporating these error patterns, the quality of PSC has been significantly improved.

For speaking training, we have focused on automatic recognition of articulatory attributes such as the place of articulation and the manner of articulation, which provides effective feedback on articulation. The major challenge is modeling non-native learners' articulation without a large speech database of this kind, which is difficult to collect and annotate. We investigate multi-lingual learning of the deep neural network (DNN)-based articulatory attribute recognizer. Here, we assume that not a few of articulatory attributes are shared by the native language and the target language for learners, and also non-native learners' articulation is affected by their native language. For example, when we target English learning by Japanese learners, the DNN-based recognizer is trained by using a Japanese native speech database and an English native speech database, both of which are available in a large scale. This achieves a significant improvement in detection of articulation errors and also ASR performance for non-native speech uttered by language learners.

These methods will realize effective CALL systems.

Figure 1 Example of Partial and Synchronized Captionin (PSC)
©TED talk "The puzzle of motivation" by Dan Pink
PSC demo at http://sap.ist.i.kyoto-u.ac.jp/psc/

References

[1] M.Mirzaei, K.Meshgi, Y.Akita, and T.Kawahara. Partial and synchronized captioning: A new tool to assist learners in developing second language listening skill. ReCALL Journal, Vol.29, No.2, pp.178--199, 2017.

[2] M.Mirzaei, K.Meshgi, and T.Kawahara. Detecting listening difficulty for second language learners using automatic speech recognition errors. In Proc. Workshop Speech & Language Technology for Education (SLaTE), pp.164--168, 2017.

[3] R.Duan, T.Kawahara, M.Dantsuji, and J.Zhang. Effective articulatory modeling for pronunciation error detection of L2 learner without non-native training data. In Proc. IEEE-ICASSP, pp.5815--5819, 2017.

[4] R.Duan, T.Kawahara, M.Dantsuji, and H.Nanjo. Transfer learning based non-native acoustic modeling for pronunciation error detection. In Proc. Workshop Speech & Language Technology for Education (SLaTE), pp.50--54, 2017.