Spontaneous Mandarin Speech Recognition with Disfluencies Detected by Latent Prosodic Modeling (LPM)

Che-Kuang Lin¹, Shu-Chuan Tseng², and Lin-Shan Lee¹

National Taiwan University¹
Academia Sinica²

In this paper, a new approach for improved spontaneous Mandarin speech recognition using Latent Prosodic Modeling (LPM) for disfluency interruption point (IP) detection is presented. The basic idea is to detect the disfluency interruption points (IPs) prior to the recognition, and then to incorporate these information into the recognition process via the second pass rescoring. For accurate detection of disfluency interruption points (IPs), prosodic information from local to global, from observable to latent, were integrated using the proposed Latent Prosodic Modeling (LPM). A whole set of new features were first defined for each syllable boundary obtained in the first pass recognition by carefully considering the special characteristics of Mandarin Chinese, and the importance of each feature with respect to each disfluency type was analyzed. Then, a set of prosodic characters, prosodic terms, and prosodic documents were defined to be used in the Probabilistic Latent Semantic Analysis (PLSA), based on which the prosody can be modeled using a set of prosodic states representing various latent factors such as speakers, speaking rate, utterance modality, intonation behavior, etc. in terms of some probabilistic relationships with the observed prosodic features. Using all these different levels of information, the approach of incorporating the decision tree into the maximum entropy model training was developed to enhance the IP detection accuracy. Experimental results indicated that the proposed set of features and the IP detection approach based on Latent Prosodic Modeling (LPM) were very useful, and the obtained information about disfluency actually benefited the speech recognition performance.

Key words: spontaneous speech recognition, disfluency, prosody, latent modeling, Mandarin Chinese