論文内容の要旨

論文題目

Chinese Dialect-Based Speaker Classification and Pronunciation Assessment Using Structural Representation of Speech　音声構造表象を用いた中国語方言に基づく話者分類と発音評価

氏名　馬学彬

Current situation of Chinese dialects is very complicated. There are several general dialect regions in China and they are further classified into many different sub-dialects and sub-sub-dialects. Although these dialects are developed from the same root and have inherited a lot of common features, they are still different to each other in varying degrees grammatically, lexically, phonologically and phonetically, because of my different social, historical and geographical reasons. Therefore, people from different general dialect regions cannot communicate orally, and even for the people from adjacent cities, their dialects are different and they have difficulty in oral communication sometimes. So since 1954, standard Mandarin has been popularized all over the country as the official language. Every dialect speaker began to learn Mandarin just like learning a second language, but their Mandarin pronunciations always have some regional accents affected by their native dialects. Meanwhile, these dialects are still developing. For example, affected by the popularization of Mandarin and people are moving across different dialect regions, many of these dialects are still developing and losing some of their special features. Strictly speaking, every speaker has his/her individual dialect, because speakers of the same dialect are often speakers of different sub-dialects and the dialect of this speaker may already change affected by other dialects or Mandarin.

In modern speech processing technologies, segmental features of speech are usually represented acoustically by spectrum, which contains not only linguistic information but also extra-linguistic information corresponding to age, gender, speaker, microphone and so on. But in the case of dialect-based speaker classification, only the acoustic features which are relevant to dialectal information should be focused on and the extra-linguistic features should be canceled. Just like this problem, in conventional speech processing frameworks, speaker-independent acoustic model is often built by collecting the data of many different speakers trying to cover the different speaker features and extract the linguistic features. Then for the processing of different dialects, speaker-independent but dialect-dependent models are always built using the data of many speakers for every dialect. However, this method doesn't work in Chinese dialect-based speaker classification. It is not only because collecting the data of so many dialects and sub-dialects is a very challenging work, but also because creating a dialect model from utterances of different speakers of the same dialect is conflict with the target of finding the intra-dialect

relations among speakers. For this problem, the linguistic features should be extracted for every individual speaker.

In my study, a novel structural representation of speech is proposed to represent Chinese dialect pronunciation. Using the dialect utterances of every speaker, the dialect pronunciation structure can be built by the distances between any pair of these utterances. As the extra-linguistic features are already removed, the dialect-based speaker classification can be achieved by classifying these structures based on the distances between these dialect structures. After that, two experiments of dialect and sub-dialect based speaker classification were carried out separately and linguistically-reasonable results were obtained. Then in order to prove that the structural method can still work well in the situation of minimum or maximum speaker differences, some more experiments are carried out. Corresponding to the data of some speakers from different dialects and sub-dialects, a new data set with constant speaker identity was created by an expert of Chinese dialect. She transcribed all the data and read the linguistic content of each original utterance in her voice through looking at the transcript and listening to the original utterance. Using these two data sets, classification experiments using our structural method were carried out and very similar results were obtained, which means our method can extract the speaker-invariant dialectal features. At last, corresponding to these two data sets, new data sets like pronounced by very tall and short speakers were simulated. Using the original data and simulated versions of these data sets, classification experiments based on structural comparison and conventional spectral comparison were carried out separately. Then about the results obtained using structural method, unlike the result of spectral comparison was affected greatly by the speaker features, speaker-invariant dialect-based are obtained.

Besides dialect-based speaker classification, we also applied the structural method to calculating the utterance similarity between two dialect speakers. By comparing the pronunciation of speakers from the same dialect regions with standard Mandarin, very similar similarity orders of the utterances were found and they were robust to the genders of the speakers. Then this method is further applied to pronunciation assessment of accented Mandarin. For every utterance of accented Mandarin, a structural score was given by comparing the pronunciation structure of accented Mandarin with the structure of standard Mandarin. Meanwhile, two kinds of scores were given by pronunciation evaluation manually and speech recognition with a recognizer. At last, these scores were compared by calculating the correlation coefficients and the results were discussed.