Sentence design for speech recognition database

Zu Yiqing

The material of database for speech recognition should include phonetic phenomena as much as possible. At the same time, such material should be phonetically compact with low redundancy[1,2]. The phonetic phenomena in continuous speech is the key problem in speech recognition. This paper describes the processing of a set of sentences collected from the database of 1993 and 1994 "People's Daily"(Chinese newspaper) which consist of news, politics, economics, arts, sports etc.. In those sentences, both phonetic phenomena and sentence patterns are included.

In continuous speech, phonemes always appear in the form of allophones which result in the co-articulary effects. The task of designing a speech database should be concerned with both intra-syllabic and inter-syllabic allophone structures. In our experiments, there are 404 syllables, 415 inter-syllabic diphones, 3050 merged inter-syllabic triphones and 2161 merged final-initial structures in read speech. Statistics on the database from "People's Daily" gives and evaluation to all of the possible phonetic structures.

In this sentence set, we first consider the phonetic balances among syllables, inter-syllabic diphones, inter-syllabic triphones and semi-syllables with their junctures. The syllabic balances ensure the intra-syllabic phenomena such as phonemes, initial/final and consonant/vowel. the rest describes the inter-syllabic jucture. The 1560 sentences consist of 96% syllables without tones(the absent syllables are only used in spoken language), 100% inter-syllabic diphones, 67% inter-syllabic triphones(87% of which appears in Peoples' Daily). There are roughly 17 kinds of sentence patterns which appear in our sentence set. By taking the transitions between syllables into account, the Chinese speech recognition systems have gotten significantly high recognition rates[3,4].

The following figure shows the process of collecting sentences.

[People's Daily Database] -> [segmentation of sentences] -> [segmentation of word group] -> [translate the text in to Pin Yin] -> [statistic phonetic phenomena & select useful paragraph] -> [modify the selected sentences by hand] -> [phonetic compact sentence set]