《语音研究报告》

2006
Report of Phonetic Research

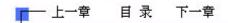# STANDARDIZATION OF SPEECH CORPUS (ABSTRACT)

Li Ai-jun, Yin Zhi-gang

Speech corpus is the basis for both analyzing the characteristics of speech signal and developing speech synthesis and recognition systems.
In mainland China, almost all of the speech research and development affiliations are developing their own speech corpora. We have so many different kinds and a large number of Chinese speech corpora that it is rather important to be able to conveniently share these speech corpora to avoid wasting of time and money and to make the research work more efficient.

The primary goal of this research is to find a standard scheme which can make the corpus be established more efficiently and be used or shared more easily.
A huge speech corpus on 10 regional accented Chinese, RASC863 (a Regional Accent Speech Corpus funded by National 863 Project，4 out of 10 regional data have been distributed until this time) will be exemplified to illuminate the standardization of speech corpus production.

全文：2006_11.pdf