

汉语语音识别数据库研制¹⁾

王仁华 夏德瑜 倪晋富
(中国科技大学无线电系, 合肥 230026)

摘 要

本文探讨了汉语语音识别数据库的设计原则,扼要介绍了 CSDB 汉语语音识别数据库的内容和研制进展。

关键词: 语音识别, 数据库, 汉语语音库。

一、引 言

语音识别是发展人机语声通信和新一代智能计算机的重要组成部分。近年来,国际和国内的专家学者们都在积极地开展这一研究工作,并取得了很大进展,从特定人孤立字的语音识别到连呼数字、非特定人的连续语音识别,进而发展声控打字机、自动电话翻译系统等更高难度的前沿课题。随着语音识别理论和实践的不断深入,建立标准语音识别数据库的需要日益迫切。一方面,由于语音信号强烈依赖于发声者的个人特征,随机性又很大,建立高性能的语音识别系统,需要反复地用大量的语音样本来测试和修正识别算法并且训练识别系统;另一方面,标准的语音数据库可以对不同的语音识别算法和系统性能进行科学的评估,从而推动语音识别的发展,有利于国际上语音识别研究的交流。

二、关于汉语语音识别数据库

虽然语音识别的算法应该是与语言种类无关的,但事实上不可避免地有很多识别技巧是针对特定语种的。因此,世界各国都在动用大量的人力和物力建立适合自己语种的标准语音库,例如美国的 TIMIT 语音数据库^[1],日本的 ATR 语音数据库^[2],英国的 SCRIBE 语音数据库等等。汉语是世界上使用人口最多的语种之一,它除了有很多个性特点外,由于我国域广人多,即使讲普通话,不同地区的人也都带着浓郁的方言口音。为此,从语音识别的需要出发,建立一个包括不同地区、不同性别、不同年龄发音人,具有自己特色而实用的汉语普通话语音数据库显得尤为必要。作为标准的汉语语音识别数据库应具备如下条件:

本文于 1990 年 11 月 10 日收到。

1) 国家自然科学基金资助。

1) 数据库中的语音材料应包含全部汉语单音节,不同音节连读时可能出现的协同发音变化,包含一些能满足连续语音识别需要的句子和短文。

2) 应有一定的人数,大约 200 至 500 名发音人,其中小部分为专业播音员,考虑发音人年龄、性别和籍贯的适当分布。部分发音人对语音材料有多次发音。

3) 数据库不仅提供数字化语音样本,而且提供基本的声学特征参数,例如音素标记、基音参数、LPC 参数、例谱参数等等。

4) 有完善的数据库管理系统以及标准输入、输出接口,易于对各种语音识别系统进行训练、测试和评估。

完成这样的数据库,将可服务于各种语音识别的需要。但是,必须指出建立这样大型的数据库,不仅有巨大的工作量,而且需要兼顾科学性与实用性,需要一定的设备和语言学家的协助。因此这是一项应该全面规划、分阶段进行的工作,要有一定的人力、财力支持。

三、CRDB 语音数据库的研制

研制 CRDB 语音数据库是我国高技术发展计划中的一个项目。目前国内在汉语语音识别方面的研究工作,主要集中在认人的全音节识别和孤立词识别,离不认人的连续语音识别还有一段距离。因此 CRDB 主要面向国内及国际上进行特定人汉语孤立词语音识别研究,研制目标为一个包括 10 位发音人,每人对全部语音材料发音 5 遍以上的汉语普通话数据库。以下是 CRDB 的概况介绍:

1) 语音材料。包含四声的全部单音节约 1260 个,10 个数字及包含各种数字音节间过渡的 37 个数字串,易混淆字集和 420 个二字词组。二字词组考虑了汉语普通话中音节之间的所有音联现象,是以第一个音节的尾音素与第二个音节的首音素构成音联基素来选词,选词原则是:(1) 考虑各种可能的音素过渡;(2) 适当考虑对比和连用;(3) 考虑声调组合与词频。

2) 技术指标。语音材料在标准录音室用高信噪比的 DAT 数字记录仪录制, DAT 中数据以数字磁带方式存储。具体参数为:录音室的混响时间约 0.3 秒,录音室的隔音大于 40dB, DAT 录音的抽样率为 48kHz,量化精度 16bit,录音信噪比达 90dB 左右。

3) 成果形式。预期提供给用户的数据包括(1) 10 名发音人(5 男 5 女,年龄适当分布,其中 6 名为专业播音员)用标准普通话对以上语音材料发音 5 遍的盒式录音磁带。这些录音是分两次完成的,中间相隔约两个月;(2) 数字化后的语音材料。数字化采用 16kHz 抽样率, 14bit 量化,这些数据可以高密度软盘形式提供,读取格式满足 IBM 微机系列及其兼容机操作系统要求;(3) 提供数据库语音材料的详细分类清单及有关的使用说明。

四、结 束 语

CRDB 是汉语语音识别数据库系统工程的第一阶段,下阶段准备增加发音人的数目,

扩充语音材料,以满足不认人或连续语音识别的需要;同时对已数字化的语音材料进行语音特征标记,并建立语音识别评估控制系统等等。

笔者的最终目标是建成能满足各种不同语音识别需要的“标准汉语语音识别数据库”,希望它不仅能够得到有关专家学者的认可,而且能作为国家的一种标准在试用中改进,进而再推向世界。建立汉语语音识别数据库任重而道远,然而这一数据库的建立必将大大促进我国以及国际上汉语语音识别研究的发展,其意义将是十分深远的。

参 考 文 献

- [1] Galofolo, J. S., The Structure and Format of the DARPA TIMIT CD-ROM Prototype, Documentation of DARPA TIMIT, 1988.
- [2] Sagisaka, Y. et al., A Large-scale Japanese Speech Database, Proc. of ICSLP-90, (1990), Japan, 1089—1092.

THE DEVELOPMENT OF A CHINESE VOICE DATABASE FOR MACHINE RECOGNITION

WANG RENHUA XIA DEYU NI JINFU

(University of Science and Technology of China 230026)

ABSTRACT

This paper presents some of the design considerations of a Chinese voice database for machine recognition, and introduces briefly the contents and the development of a Chinese voice recognition database (CRDB).

Key words: Speech recognition; database; Chinese corpus.