文 章 编 号: 1000-1220(2000)03-0295-03

汉语语音合成语料库管理系统的建立

赵世霞　蔡莲红　常晓磊
(清华大学计算机系　北京 100084)

摘　要: 本文介绍的语料库管理系统主要用于语音合成的研究或开发工作. 语料的设计考虑了音段和韵律, 语料库中包括汉语的音节、词语、独白语句和情景对话语篇, 语音的录制是在卧限录音室完成. 管理系统对各种语音数据进行综合有效的管理, 它具有查询、浏览和更新等功能.

关 键 词: 语音合成; 语料库; 音段; 韵律结构
分 类 号: TP392; H11　　**文献标识码**: A

1　概　述

在发展计算机的人机交互技术方面语音交互倍受重视, 能听会说的计算机一直是人们梦寐以求的, 作为支撑技术的语音识别与合成近年已有突破并逐步走向实用化. 语音识别与合成处理的对象均为自然语音, 但是, 目前语音的识别率和合成声音的自然度还不尽满意. 其根本原因是对自然语音的研究不够深入, 不能准确归纳描述和模拟自然语音的规律, 因此语音的分析和合成工作也就越来越依赖语音语料库了. 国外从 80 年代就开始了语音语料库的建设, 如美国、日本、瑞典和芬兰都已经建立了本国语言的语音语料库以及语音特征标注标准 ToBI. 汉语是我们的母语, 建立一个面向汉语语音合成, 满足不同层次的研究和开发的需求, 具有综合性、典型性、实用性特点的语音合成语料库以及管理系统就迫在眉睫. 在 863 的支持下, 由清华大学、中国科学院声学所、中国社科院语言所共同完成并通过了鉴定汉语语音合成语料库 (Corpus of Speech Synthesis-1, 简称 CoSS-1).

对于语料库来说, 首先设计的语料应包含尽量多的语音现象; 需要高质量的录音和对语音数据的标注; 最后是建立一个综合有效的数据库管理系统. 对语料库的管理工作大致可归纳为两部分: 一是对原始语料文本资料的管理; 二是对大存储容量录音数据的管理. 建立管理系统目的在于将这两部分有机地结合起来, 以满足语音合成研究与开发的需要. 在语音合成中经过语言学处理后的文本还要做语音学处理, 如多音字、轻声词、儿化词以及变音、变调等处理, 语音合成系统中, 可选择的合成基元有音节、音素、词汇、句子等, 它们所需要的存储容量依次减少, 而对应的合成规则的复杂性却不断在增加, 有了语料库管理系统, 可对这些合成基元进行有效的管理, 做出规范的统计. 在语料库中我们将语音的文字描述(如汉字、拼音、声调, 标音文件)与语音数据、语音波形显示有机的结合起来, 让那些从事语言、语音和与之有关的研究人员, 可以根据各自的研究目的和具体要求, 通过这个系统完成对各类语音合成基元的分析、统计、检索等研究. 本系统也力图为语音合成的评测提供素材.

多媒体技术对数据库技术领域的发展具有十分诱人的应用前景. 我们把语料库的管理和多媒体应用技术有机地结合起来, 为 CoSS-1 建立的数据库管理系统, 提供给用户的是一个文字、声音、图形、综合管理的系统. 从而提高了语料库的使用效率.

2　系统的设计与实现

2.1　原始语料的设计

该系统的原始语料设计选自数部电子词典. 我们在吸收了国内外的语料库设计的最新技术并结合汉语普通话的特点后, 形成合成语料的自动挑选原则: 同时考虑韵律和音段两个方面, 尽量涵盖音段和超音段的音联现象. 在满足声调组合的情况下, 充分考虑音节间的音联现象, 即音节间的两音子(diphone)和大部分的三音子(triphone), 设计了声调组合音节组. 语料包括汉语单音节 1268 个、二音节组 640 个、三音节组 2048 个、四音节组 2048 个、轻声音节组 349 个和儿化音节组 233 个. 由于语调模式对语音合成的自然度影响很大, 本语料库还包括多种句型的语句, 其中独白体语句 265 句和情景对话语篇 52 个场景. 这些语句满足一定的语音、句型及韵律结构, 并尽量覆盖较多的音节间的三音子. 此外, 该系统还设计了普通话的韵律特征语料库, 并对其作了韵律标音, 这部分的语料全部以对话形式出现, 每一段对话就是一个语篇, 做到力争反映生活中的真实情景. 设计原则是: 重点考虑普通话语句层面的语调变化、轻重读、焦点分布和韵律短语问题等主要的韵律现象, 兼顾句子的句法结构和功能分类, 包括了陈述句、疑问句、祈使句和感叹句四大类. 对话语篇的设计主要考虑要满足韵律标音问题的研究需要, 值得注意的是韵律结构和从文字信息中得到的句法结构是不一致的, 韵律所反映的信息是文字无法传递的, 它对语义理解(包括歧义的消除)、对语气等语用意义、对提示语篇结构等都有贡献. 合成时必须

收稿日期: 1999-06-18　基金项目: 国家 863 高技术项目资助 (863-306-03-02-4)　赵世霞, 工程师, 主要从事微型计算机接口技术的教学和语音数据库方面的研究. 蔡莲红, 教授, 博士生导师, 主要从事语音合成方面的研究和多媒体技术的教学.

考虑语句的韵律结构, 才能合成出自然度较高的语音.

2.2 语料库的录音和数据采集

语料库的录音工作在中国科学院声学研究所卦限消声室进行. 在消声环境下, 完成高信噪比(SNR > 60dB)的数字录音. 用电容式传声器悬垂于发音人正前方接收语音声压波(speech wave)和发音人颈部的喉音器接收声门阻抗波(laryngograph), 这两路信号都经插入有源带通滤波器的测量放大器定量放大, 用双音轨数字录音机同步录制在数字磁带上. 在距发音人 1 米处监测发音声级和环境噪声, 以便为语音数据的研究提供定量的标度. 录音过程中采用计算机屏幕提示系统, 最大限度降低了发音人的工作强度并减少了人为噪声. 录音材料包括老、中、青三个年龄段的男和女 6 个发音人的八种语料.

经过录制的庞大的语音数据是存储在数字磁带上, 下一步还要将磁带上的庞大文件切割成一个个语音片断并转换成计算机的数据格式, 同时建立起语音数据与文字描述之间的联系. 这是一件繁琐细致的数据采集工作, 我们通过语音信号处理技术, 实现了对语音数据采集的半自动化, 提高了工作效率.

2.3 语料库的标音

语料库的标音实际是对语音信号中具有语言学功能的语音特征(音段特征和韵律特征)进行的定性描写, 属音系学的范畴. 参照 ToBI 的设计原则并结合汉语普通话既有声调又有语调的特点, 选择标注语句中有语言学功能的声调变化、语调模式、重音模式和韵律结构单位, 而不标注那些宜于定量描写的韵律现象, 例如音段时长和强度. 标音工作在中国社会科学院语言研究所语音室的 CSL Kay 4300B 上进行, 通过分析语音的基频曲线、宽带语谱图和窄带语谱图, 在语音波形上首先把音节切出来, 作上标记(tag), 然后对每个标记所对应的音节的韵律特征进行标注. 韵律标音系统分为五层: 拼音层、语句功能层、韵律结构层、声调/语调层、语句重音层. 每个音层标记不同的韵律现象. 拼音层是语句的汉语拼音, 语句功能层标记陈述句、疑问句、祈使句和感叹句的语气. 韵律结构层标记主要韵律短语、次要韵律短语和语调短语之间的界限. 如果短语后感知到的停顿较长, 它与后面的短语连接较松散, 则此短语为主要韵律短语, 反之为次要韵律短语. 声调/语调层标记每个音节的声调变化和全句的语调变化. 语句重音层标记每个语句的重音, 包括一般重音和强调重音.

2.4 语料库管理系统

该语料库管理系统在中文 Win98 环境下, 使用微软公司的可视数据库开发工具 Visual FoxPro 5.0(中文版)制作完成. 该系统功能结构简单, 清晰, 方便非计算机专业的人员使用和操作. 管理系统负责管理语料的设计文本、声音数据、标音文件等. 具有编辑、浏览、查询、放音、更新语音数据等功能, 还可以显示部分情景对话的波形、基频、韵律标音信息.

· 浏览功能

原始语料经过录音后, 次序杂乱, 为方便用户, 采用人们习惯的按字典音序方式浏览, 此外, 轻声音节组还可以按轻声音节浏览, 儿化音节组还可以按儿化音节浏览. 它们都能直接定位到某个音节组上. 如果用户对所浏览的语料库中某些语料感兴趣, 还可以存盘分析以及根据需要作一些相关数据的统计和研究工作.

· 查询功能

主要根据电子部发布的 SJ/T 11143-1997 "计算机用普通话语音库规范"中的标准设计查询功能. 语料库中的单音节、二音节组、三音节组、四音组的查询, 采用了按汉语拼音、汉字和声调组合方式的查询方法. 语料库中的轻声音节组, 可以按轻声的构词查询, 也可以按轻声音节的拼音查询; 儿化音节组可以按儿化韵母查询或按儿化词的拼音查询. 此外, 语料库中的独白体语句, 是按汉语的句型分类查询. 语料库中的情景对话语篇, 主要是为分析韵律特征而设计的, 可以按对话场景查询. 查询结果在图形界面下显示的韵律标音信息和基频标注波形如图 1 所示.

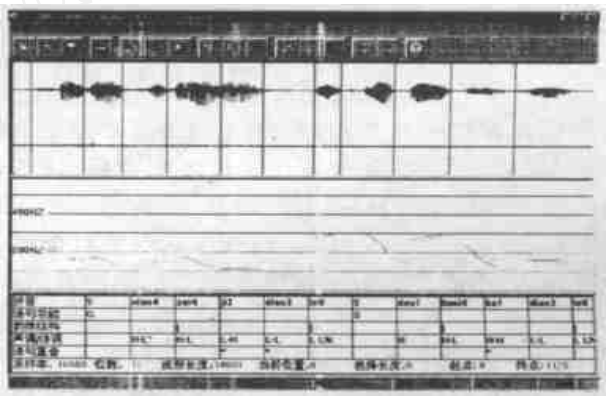

图 1 韵律标音显示界面

· 语音数据更新

该系统现有语音数据光盘 7 张, 用户可直接利用光盘或将语音数据转存到硬盘上使用这些资源. 随着研究工作的深入发展, 当原有语料库不能满足新的应用要求时, 随时可以更换或调整语音数据, 使语料库的数据处于最佳状态.

3 系统结构的组织与实现

语料库管理系统的特点是管理对象多(有原始文本数据、语音波形、韵律标注、基频标注波形等); 层次深(分录音环境、发音人属性、原始数据等); 要实现的功能也多. 要处理的实体对象主要是五种: 文本文件(.txt)、语音数据文件(.dbf)、语音波形文件(.wav)、韵律标注文件(.lab)、基频标注波形文件(.pit). 原始的语料文本, 我们称之为生语料, 首先要将它按照一定的规则组织成语料库的数据文件中的文本部份, 然后找到相应的录音数据, 并建立语音数据与语料库文本部份描述之间的联系, 最终组成了完整的语音数据文件(.dbf). 数据结构的确定要有利于系统功能的实现和方便于查询, 应找出各种处理对象的共同属性以某种有效的描述将它们有机地联系起来, 动态描述的办法是系统连接的关键. 语料库的数据结构包括汉字、声调组合、拼音、声母、韵母和语音数据文件名, 其中各音节的声母、韵母和声调按顺序排列. 例如三音节组的数据结构如下: word/ shdzh/ py/ s1/ y1/ d1/ s2/ y2/ d2/

s3/ y3/ d3/ filename. 在此基础上我们设计并实现了语料库管理系统, 并利用该系统对入库的语音数据进行检查校对. 最后是对入库的语音波形文件进行韵律标注和基频标注, 再根据标注所提供的信息编制韵律标音的显示程序, 同时实现了媒体的即时播放和显示. 该系统的原始语料有 8 种, 分别存放

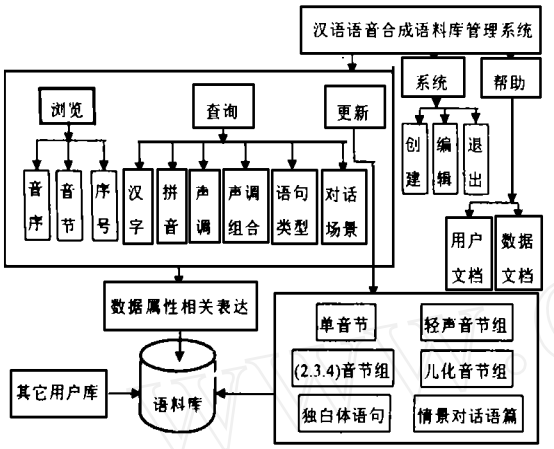

图 2　系统结构框图

在 word 目录和 sentence 目录. 语音数据按发音人存储, 每个人的语音数据 600 兆左右. 对于这样庞大的数据, 我们提供给用户的方式是可以装到硬盘上使用, 也可以直接在光盘上使用, 并可根据自己的需要作数据更新. 该系统中还有一个其它用户库, 主要是为用户对浏览和查询到的数据需要存盘时而设计. 原则上用户自己存盘的语料库, 可以利用本系统进行浏览和查询, 也可以任意编辑和修改. 而原始语料则不允许修改.

该系统提供的语音数据和管理系统可以独立使用, 也就是说用户只要将管理系统安装在自己的计算机上, 再将语音数据装载后就可以使用本系统; 另一方面也可以只使用提供的数据文件作分析和研究, 自己另外开发管理系统. 为了方便用户进行二次开发, 同时还提供了系统文档和帮助文档. 有关数据文件的索引以及原始语料库的数据结构都在帮助文档中给出, 系统文档提供给用户重组数据库和文本编辑的功能. 下面给出系统结构的框图和文件目录结构. 语料库管理系统的组织结构:

· 语料库的目录结构:

— data\	语料库的库文件
— txt\	语料库文本文件
— wav\	波形显示程序
— temp\	临时数据文件
— word\	音节组语音数据
— sentence\	独白体语句和情景对话语篇语音数据
— ylkm1. exe	语料库管理程序执行文件

word 子目录结构:

—— one\1xxxx. wav	单音节库
—— two\2xxxx. wav	二音节组库
—— tri\3xxxx. wav	三音节组库
—— for\4xxxx. wav	四音节组库
—— qingsheng\8xxxx. wav	轻声音节组库
—— erhua\9xxxx. wav	儿化音节组库

sentence 子目录结构:

—— discs\axxxx. wav	独白体语句库
—— dialg1\Xxxxx. wav	情景对话语篇库--女A, 男B
—— dialg2\Yxxxx. wav	情景对话语篇库--男A, 女B

4　结束语

本语料库是 863 信息领域智能计算机主题支持的项目, 结题成果已由主题办公室推广使用. 但语料的规模, 它的种类和数量还有待扩大, 特别是语音标注尚有差距. 语料库的建设需要语言和语音学、心理学、计算机科学等多学科的共同努力才能做好.

参 考 文 献

1　祖漪清、李爱军. 语音识别和语音合成语料库的设计. [C]第三届全国计算机智能接口与智能应用学术会议论文集. 1997, 174～179

2　李智强. 韵律研究和韵律标音. [J]语言文字应用. 1998. 1, 105～109

CONSTRUCTION OF MANDARIN CORPUS FOR CHINESE SPEECH SYNTHESIS

ZHAO　Shi-xia　CAI　Lian-hong　CHANG　Xiao-lei

(*Department of Computer Science and Technology, Tsinghua University　Beijing* 100084)

Abstract　This paper presents a corpus for Chinese speech synthesis. Segment and prosody are for designing database. Chinese syllables, words, sentences of soliloquy and paragraphs of dialogue are contained in the database. Speech is recorded in one-eight anechoic chamber. There is an effectual management system in the corpus. It has functions of searching, browse and renewing.

Key words　Speech synthesis; Mandarin corpus; Segment; Prosodic structure