语音合成语料库的设计与声学特征分析

蔡莲红，蔡锐，吴志勇，陶建华

（清华大学 计算机科学与技术系,北京 100084）

Design and analysis of acoustic feature for corpus of speech synthesis

CAI Lian-hong, CAI Rui, WU Zhi-yong, TAO Jian-hua

(Department. of Computer, Tsinghua University,Beijing 100084,China)

1 引言

近年来，基于大语料库的TTS技术迅速发展，推动了语音语料库的研究和建设。许多国家都建立了大量的语音语料库，如美国、日本、瑞典、芬兰等。我们面对语音合成的需求，设计并建立了相应的语料库。该语料库包括文本、语音数据、标注等，它较好的满足了TTS的需求。同时研制了语音分析工具，进行了语音特征分析、韵律建模等研究。本文基于语音的声学特性，提出用韵律位模型来界定韵律层级的尝试。

2 语料库的设计

在语料库的建设中，语音语料库的完备性和科学性是十分重要的。语音语料库的完备性体现在它能涵盖汉语音段、超音段等信息。科学性体现在它既能满足TTS系统研究和开发的需求，信息的冗余又小。我们的语料库力求覆盖汉语的音素、半音节、音节以及它们之间的音变现象。汉语是声调语言，词调、语调模式对合成的影响很大。连续语句能充分反映汉语语言的韵律结构，特别是语调信息。因此语句是本语音语料库的重要组成部分。尽量覆盖语言中的文本类型、句法结构以及其他可能对韵律有重要作用的因素。语料库包含了这三种类型文本和相应的数据：

- ✓ 成段的语句，基本为新闻体，覆盖多种题材，兼顾特殊的文字形式，如数字、地址等。
- ✓ 为覆盖语言中重要的句法和语调变化所选的语句。
- ✓ 为较好覆盖该语言中语音变化和标点所选的语句。

设计过程：从广播、电视、文艺作品、词典例句等选取了一批这三种句型的语句。对每一种句型，参考“现代汉语基本句型”，每种句型选出几个有代表性的语句，同时考虑到句子的长短不同，语气强弱，有无关键词及关键词的位置。语句长度是10～20个音节，包含连续的声调组合和基频稳定的语句。

语料库覆盖了'95汉语句型频度表中出现频度较高的句型中的95%以上，同时，对一些出现频度看来不高的句型，例如“主||"把"宾+"给"+动+其他成分”的句型， 按照频度表中的统计，出现概率低于0.01%，即不到万分之一，但考虑到它语调的特殊性，也特地包含在设计结果中。

3 语料库的标注和标注工具

建立语料库的工作还包括录音、标注等。根据语音合成研究的需求，我们对音库进行了音节、半音节的切分，同时增加了音段标音和韵律标注，包括文本的拼音、韵律边界，重音；语音数据的基频计算和标注等。韵律层级的划分为：韵律词（标记是 | ）、韵律短语（/）、语调短语（//）和

作者简介：蔡莲红，主要研究语音合成与处理，语音语料库，多媒体音频处理等。

语句边界（///）。音节的轻重划分为：轻声（标记是W）、轻读（L）、标准（N）、重读（H）、重重（V）。

例如：////缓和了|党/同|工(H)商界/知(H)识界/和(L)/民(H)主党派的|紧张关系/,//

huan3 he2 le5 dang3 tong2 gong1 shang1 jie4 zhi1 shi5 jie4 he2 min2 zhu2 dang3 pai4 de5 jin3 zhang1 guan1 xi5

音节切分和基频标注如图1所示。

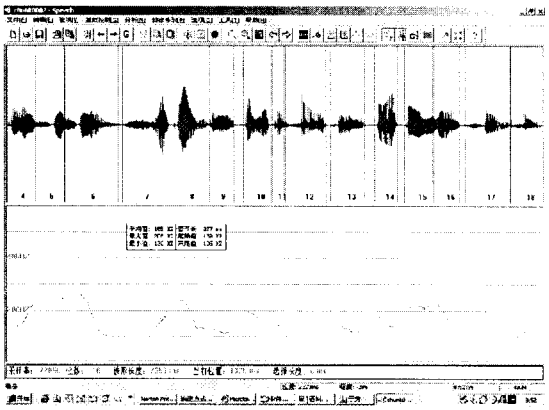

图1 语音分析工具 Speech

4 语音声学特征的韵律位(prosodeme)模型

语音携带了大量的韵律信息，我们利用统计的方法对语料库进行分析和研究。面对韵律层级划分中的困惑，提出用韵律位(prosodeme)模型综合描述语音的声学特征，并进而界定韵律层级。

我们统计了本语料库中韵律边界处的一些声学特性，如下表：

边界类型	音节	韵律词	韵律短语	语调短语
基频段时长（ms）	215.54	234.99	297.14	314.59
基频间断时长（ms）	86.93	114.53	151.85	369.18
边界前后音高重置（Hz）	-11.82	3.66	19.47	32.31

从表中可以看出，不同韵律边界的音节时长差别、音高差异都较小，而且实测表明它们的分布符合 Gauss 模型的正态分布。若单独以音节时长、音高来区分韵律边界就会出现混淆。

我们基于韵律边界处声学参数的统计特性，建立韵律位（prosodeme）模型，利用 Bayesian 决策的方法来进行韵律层级边界类型的划分和判定。具体做法是：首先人工标注大量语句，对已经标注的数据进行统计分析，获得各类韵律边界出现频率的先验概率。然后假设随机变量 x, y, z 分别代表韵律边界处音节的声学参数。在给定某音节边界的上述三个特征的取值的情况下，能够判定该边界的类型。韵律层级确定的问题转化为概率计算问题。定义 T 为边界类型，则韵律边界处音节的韵律位模型可表示为：

$$P(T = t \mid X = x, Y = y, Z = z) \qquad t \in \{A, B, C, D\}$$

其中，A, B, C, D 分别代表不同的韵律层级，上式表明韵律边界处音节的声学特征的联合属性决定了韵律的层级。最大概率所对应的 T 的类型则为该处边界最有可能的类型。

我们对人工标注的 1000 个句子采用上述的方法进行韵律层级边界等级的判定，对比人工标记的结果，计算机自动判断韵律层级的正确率为 80.0%。

5 结束语

在语料库的建设中，语料标注和分析的工作量极大。我们基于统计的方法研究了韵律层级的划分，得到了一些初步的结果。但如何提高分析的精度，减少标注的工作量还有许多问题有待研究。我们期望多学科的合作，让语料建设、语音分析更好地为语音处理服务。