文语转换系统韵律置标方法的研究*

蔡莲红　罗　恒　汪　泳　谭　晖　　涂相华

（清华大学计算机系　北京　100084）　　（信阳师范学院计算机系　信阳　464000）

摘要　韵律理解是言语合成的基础.本文分析了文语转换系统TTS(text to speech)的研究现状,提出了韵律置标的方法,设计了韵律符号,并将其用于汉语TTS系统中,实现了重音和语调的模拟,改善了输出语音的自然度.

关键词　韵律,置标,文语转换.

韵律理解(理解口语和文字间的关系)是自然语音处理、高质量语音合成、话语理解系统的基础性课题.韵律不仅描述了语音信号的变化,还能表达语音的深层信息,而这些信息超出从孤立字符所抽取的信息.

文—语转换系统的功能是将文字转换成声音,其研究的内容已超出"见字发音".如英语TTS(text to speech)中见到e,系统要决定是发e音还是发ə音.汉语TTS中,系统要处理单字在连续语流中的音变,如:多音字、儿化轻声、语流中插入的停顿、音的加重或轻化.有时,同样一段文字说出来时却采用不同的语调等.统称之为韵律特性.

计算机生成自然语音,首先要对语音的韵律特性进行抽取、描述,归纳其规律,有效地理解和研究韵律的计算模型,然后在TTS中进行模拟.在韵律描述方面,已制定了相关的标准(ToBI)[1],这是一个韵律标注系统.

当输入仅仅是文字时,如何使TTS输出的语音带有自然语音的韵律呢?理想的解决方法是利用自然语言理解的结果,指导声学参数的调整.显然这是困难的.当前,可以人为的借用一些符号来表述所期望的韵律特性,并将其插入字里行间,来实现韵律置标.在输出语音时,再将这些符号转换成声学参数,对合成语音信号进行修饰.这方面的研究刚刚起步.我们进行了韵律置标方法的研究和系统的设计,以此方式控制TTS系统的语音输出,实现了重音和语调的模拟,改善了语音的自然度.

1　韵律描述和韵律控制符

1.1　韵律描述

* 作者蔡莲红,1945年生,教授,主要研究领域为语音合成与识别,多媒体.罗恒,1974年生,硕士,主要研究领域为多媒体.汪泳,1972年生,硕士,主要研究领域为语音合成.谭晖,1973年生,硕士,主要研究领域为语言信号处理.涂相华,1964年生,讲师,主要研究领域为语音合成,计算机应用.

本文通讯联系人:蔡莲红,北京100084,清华大学计算机系

本文1996－01－30收到修改稿

ToBI 韵律描述系统[2]的功能是对自然语音加以韵律标注.韵律标注可分层次进行,每一层都设有表示韵律事件、事件相伴时刻的符号.如:

·音调(tonal)层:用一个线性的基音序列描述音调.音调描述贯穿于整个语音数据中.根据基音的变化规律,基音描述如下:

H*　简单的高音　　L*　简单的低音　　L+H*　从低音升到高音

L*+H 后升音　　H+!H* 最后落到重音

·断引(Break Index)层:用 7 个等级(0…6)来标识两相邻音的关系.ToBI 标准引入了 3 个最高的 Break Index,表示音调段.并将级 0 和 2 的定义进行了修改使其更明确,增加了标注的稳定性.其中 0…4 的含意如下:

0　在相邻 2 个词间,有很类似的声学特征

1　2 词属于不同的韵律

3　音的边界

4　全音调的片断

对于一个韵律描述标准来说,ToBI 具有以下特点:能够描述自然语音的最重要的韵律特征;与语法分析的输出、语义学的正确表达一致;易学易用,适用于不同地点收集的数据;不同描述者使用时,其标注结果的共同之处不少于 80%.

目前,ToBI 仅对语音作韵律描述,韵律规律的抽取还有待进行.该标准已用于英语语音韵律分析和标注,但还没有用于汉语.

1.2　韵律控制符

在文语转换的研究中,已采用了一些韵律控制符,如:

(1) TTS 控制标记(Control Tags)

Microsoft 公司推荐了一些控制语音韵律的符号,用于控制变音、重音、时长等,可以插入在原文正文间或由程序自动插入.简述控制标记如下:

Chr:置声音的风格.如:Angry(生气)　Happy(高兴).

Com:在文本中嵌入注释.

Ctx:设置随后的文本的类型,以决定符号如何被读出.

Emp:强调待读的下一个字.

Eng:嵌入一个 TTS 引擎命令.

Mrk:指出文本中的书签.

Pau:暂停语音数毫秒.

Pit:设置语音的基频.

Prn:指定文本如何发音.

Pro:指定韵律规则是否有效.

Prt:指定下一个字的词类.

Rst:复位所有的韵律标记.

Spd:设定讲话的速度.

Vce:通知 TTS 引擎改变语音的特性.

Vcl:设置语音的音量.

这套标记符是以英语为背景的.

(2) 汉语 TTS 中的韵律控制符

我们针对汉语韵律的特点设计了 10 个韵律控制符,并使用这些控制符实现了一些增强 TTS 的自然度和表现力的规则.[3]

对某一音节的韵律控制,例如,"衣\i3\服",表示音节"衣"的基频提高 3 级. 表 1 为韵律控制符及控制级数的说明.[4]

表 1 韵律控制符号和控制级数说明

符号	控制说明	级数 n 的意义(有效值为 0~9)
h	修改音域上限	$n>4$ 上限升高,n 越大,升高越多; $n<5$ 上限降低,n 越小,降低越多.
b	修改音域下限	$n>4$ 下限降低,n 越大,降低越多; $n<5$ 下限升高,n 越小,升高越多.
p	修改基频模式	n 为基频模式号.
r	修改音域宽度	$n>4$ 音域加宽,n 越大,加宽越多; $n<5$ 音域变窄,n 越小,变窄越多.
i	调值提高	n 越大,提高越多.
e	调值降低	n 越大,降低越多.
d	修改音节时长	$n>4$ 时长增加,n 越大,增加越多; $n<4$ 时长减短,n 越小,减短越多; $n=4$ 时长不变.
c	截掉某些周期	$4<n<9$ 截去尾部周期,n 越大,截去越多; $0<n<5$ 截去头部周期,n 越小,截去越多; $n=9$ 截去尾部未加标注的波形段; $n=0$ 截去头部未加标注波形段的前 1/4.
a	修改音节幅度	$n>4$ 幅度提高,n 越大,提高越多; $n<5$ 幅度降低,n 越小,降低越多.
u	音节前后加停顿	$n>4$ 在尾部加停顿,n 越大,停顿越长; $n<5$ 在头部加停顿,n 越小,停顿越长.

2 韵律置标的设计

为了使韵律描述更具一般性,对 TTS 输出语音具有更强的控制能力,我们参考 Microsoft 的控制标记和 ToBI 标准,采用韵律置标(Prosody Markup)的方法,改进了我们设计的韵律控制符.

2.1 韵律置标的设计原则

利用置标符号控制合成汉语语音的韵律特性. 考虑到 TTS 系统采用不同合成算法,韵律置标重在标记语音的整体特性. 各系统可将这些置标符号转换成底层参数,修饰合成的语音. 韵律置标系统应具有较强的通用性. 置标符号系统应具有可扩充性.

2.2 韵律置标的语法

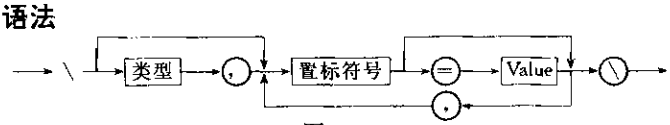

图1

图 1 中标注以反斜杠(\)引导,2 个反斜杠之间是一个有效的置标符号. 若在正文中引

用反斜杠,请用连续的 2 个来表示.标注与大小写无关,系统不识别并忽略错误的标注.置标符号的使用方式举例如下:

天\read=tian1\,其中置标符号 read 指定前一个字的读音.tian1 是该读音的拼音.

2.3 置标符号

置标符号用于控制相应文字的读音、语速、语调、基频、标点读出、数字读出、停顿、音量等,还可用"复位置标符号"恢复系统的置标参数为默认值.根据标注的作用域不同有全局标注和局部标注.

- 类型置标符号 Local:指明在同一标注内的其后的标注为局部标注,即只对当前标注前的一个字有效.
- 置标符号 Language:指明语音输出的语种,默认为汉语.本置标符号默认为全局标注.
- 置标符号 Read:指明标注前那个字的读法,默认为局部标注.
- 置标符号 Pitch:设置语音基频,默认为全局标注.
- 置标符号 Speed:设置语音语速,默认为全局标注.
- 置标符号 Volume:降低语音音量,默认为全局标注.
- 置标符号 Pause:设置语音每个字后停顿时间,默认为全局标注.
- 置标符号 Sig:设置标点读出与否,默认为全局标注.默认为 Off.
- 置标符号 Digital:设置数码读出方式,默认为全局标注.默认为 Telegram.
- 置标符号 Style:设置语音风格,如:男/女,老/少,高兴/悲哀.默认为全局标注.
- 置标符号 Silent:指出其后的字符是注释.
- 置标符号 Stress:指明标注前那个字的重音的级别.默认为局部标注.
- 置标符号 Intonation:指明标注前那个语句的语调.默认为局部标注.
- 置标符号置标符号 Reset:复位所有标注为默认值.
- 置标符号 Type:放置随后的文本类型.
- 置标符号 Part:指定下一个词的词类.

2.4 置标符号处理

定义置标符号的文法如下图所示,为简便起见,仅以 read 为例,同时拼音码不一一列举;其它置标符号的文法可能有一些差别,如对置标符号 local,〈tag equation〉→rst,但都可以类似描述.

〈tag〉→\〈tag list〉\

〈tag list〉→〈tag list〉,〈tag equation〉|〈tag equation〉

〈tag equation〉→〈tag〉=+〈value〉|〈tag〉=-〈value〉|〈tag〉=〈value〉|ε

〈tag〉→read

〈value〉→〈pingyin〉+〈yindiao〉

〈pinyin〉→a|ai|…

〈yindiao〉→0|1|2|3|4|5|6|7|8|9

其中带尖括号的为非终结符,否则为终结符.

置标符号文法包含 1 个左递归,消去后是 1 个 LL(1)文法,可以用递归子程序法处理.

程序实现时还要注意置标符号的错误处理.由于置标符号处理只是文语转换过程的文

本分析部分，错误的置标符号跳过即可，本系统是通过返回消息表明置标符号是否包含了语法错. 错误的置标符号也可作为正文处理.

置标符号处理之后从正文中滤去，代之为数据索引、声学参数等.

3 结束语

计算机模拟自然语音还有很长的路要走. 我们提出为 TTS 系统设计韵律置标，是希望 TTS 按置标要求实现各种韵律特性. 进一步的工作是借助自然语言理解的成果，自动置标，自动模拟自然语音.

参考文献

1 Silverman Kim, Beckman Mary, Pitrelli John *et al*. TOBI: a standard for labeling English prosody, ICSLP, 1992. 867～870.

2 Pitrelli John F, Beckman Mary E, Hirschberg Julia. Evaluation of prosodic transcription labeling reliability in the TOBI framework, ICSLP, 1994.

3 周俏峰，蔡莲红. 提高合成语音表现力的研究. 第 2 届全国计算机智能接口与智能应用学术会议论文集，威海，1995. 149～154.

4 周俏峰，蔡莲红. 汉语的重音及在 TTS 系统中的模拟. 全国第 4 届多媒体技术学术会议论文集，广州，1995. 36～41.

RESEARCH OF PROSODY MARKUP METHOD IN CHINESE TTS

Cai Lianhong Luo Heng Wang Yong Tan Hui

(*Department of Computer Science Tsinghua University Beijing* 100084)

Tu Xianghua

(*Department of Computer Science Xinyang Normal College Xinyang* 464000)

Abstract Prosody understand is the foundation of speak synthesis. After investigating the state of research in TTS. A method of prosody markup is proposed. Grammar and symbols of prosody markup are designed. The proposed method is used in a Chinese TTS system. As a result of it, simulation of stress and intonation is achieved. The naturalness of synthesized speech is improved.

Key words Prosody, markup, text to speech.