第26卷 第1期 计 算 机 工 程 2000年1月
Vol.26 № 1 Computer Engineering January 2000

·多媒体技术及应用· 文章编号：1000－3428(2000)01－0084－03 中图分类号：TP391.43

PSOLA技术在汉语文-语转换系统中的应用

陈 愉，张宗红，李 炜，李宗葛，宋 彬
（复旦大学计算机系，上海 200433）

摘 要：首先简述了语音合成技术和文-语转换的基本原理，然后介绍了一种可用来进行韵律修正的基音同步叠加（PSOLA）技术。在此基础上，提出用基于PSOLA的波形编辑语音合成技术应用于汉语TTS系统中，以提高输出语音的自然度和听觉效果。

关键词：文-语转换系统；基音同步叠加；韵律特征

Application of PSOLA Technique to Chinese Text-to-Speech Conversion System

Chen Yu, Zhang Zonghong, Li Wei, Li Zongge, Song Bin
(Department of Computer Science, Fudan University, Shanghai 200433)

【Abstract】This paper first narrates the principles of the speech synthesis and the text-to-speech technology. Then a new technology named PSOLA(Pitch-Synchronous Overlap Add) which can be used to prosodic modifications is introduced. A PSOLA based waveform coding synthesis technology in the Chinese TTS system is proposed and implemented to improve the speech's naturality and the auditory effect.

【Key words】Text-to-speech conversion system; PSOLA; Prosodic feature

计算机语音输出是智能化计算机的重要特征之一，文-语转换系统(TTS)是一种比较高级的、有广泛应用价值的计算机语音输出方式。目前，关于文-语转换系统的研究主要集中于如何提高其输出语音的自然度。

语音的韵律特征和语音自然度有着密切关系。基于PSOLA的波形编辑语音合成技术是在保证基音周期完整性的前提下，将波形数据库中的波形级联起来，输出连续语流的技术。它既能保持原始发音的主要音段特征，又能在拼接时灵活调整其基频、时长和强度等超音段特征(韵律特征)。

1 文-语转换系统TTS(Text-to-speech)

一般来说，实现计算机语音输出有两种方法：一是录音/回放。先把模拟语音信号转换成数字序列，编码后，存放在储存设备中(录音)；需要时，再经过解码，重建语音信号(回放)。这种方法所产生的音质能保留个人的音色，但存储量随发音时间线性增长。所以仅适用于语音输出时间短且不经常变化的场合。另一种方法就是文语转换(TTS)。文语转换是一种高级的语音输出，它能把文本转换成连续自然的语流。采用这种方法，先建立语音数据库、发音规则库。需要输出语音时，只要输入待发音的字符，系统便能按语音规则输出语声流。文-语转换系统的语音库不随发音时间的增长而加大，但规则库会随语音质量的要求而增大。

完备的文-语转换系统一般由语言学处理、语音学处理和语音合成这三大部分组成。整个系统包括以下几个组成部分：文本预处理、分词处理和分词词典、句法分析、多音字处理和多音字典、音变处理及韵律规则、语音合成器以及语音数据库，如图1所示。输入的文本材料经语言学处理、语音学处理，得到语流控制参数，然后读取语音数据库，经语音信号处理，输出连续语声。

目前，包括汉语在内的各个语种的TTS系统的研究，都面临着如何提高输出语音自然度的问题。对于这个问题，目前的研究主要围绕两个方面：一是通过自然语言理解，从输入语本中提取语音韵律特征；二是根据韵律规则，利用韵律修改算法，对欲输出的语流进行修改，从而得到良好的语音输出。PSOLA算法主要对后者提出改进方案。

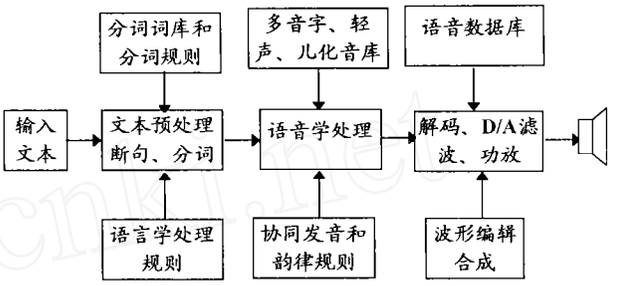

图1 文-语转换系统的组成

韵律特征(Prosodic Feature)就是语流中由音高、音长和强度等方面的变化所表现出来的特征，也叫作超音段特征(supra-segmental feature)。

语音合成从技术上讲主要分参数合成和波形合成。参数合成的系统结构较复杂，合成的音质也较差，在这就不作讨论。语音的波形编码合成方法是以语句、短句、词或音节为合成单元，这些单元被分别录音后直接进行数字编码，组成一个合成语音库；重放时，根据待输出的信息，在语音库中取出相应单元的波形数据，串接或编辑在一起，经解码还原出语音。合成单元越大，合成的自然度越好。在波形编辑语音合成技术中,要对语音信号的超音段特征的韵律修改及语音学中一些协同发音及音变现象进行分析和处理。韵律在时域波形上的主要体现是时长、音高、音强及语音波形的形状之别。它们反映了语音在基频、共振峰、能量及谱分布特性上的差异。波形编辑语音能对韵律进行灵活方便的修改，使语气、语调、重音达到我们所要求的效果。

PSOLA是一种韵律修改算法。它以基音周期(而不是传

作者简介：陈 愉（1973～），男，研究生，主研语音识别

收稿日期：1999－05－14

统的定长的帧)为单位进行波形的修改。算法直接作用于语音波形的数据，实现语音的拼接、韵律的修改。

2 PSOLA(基音同步叠加)算法

PSOLA(Pitch-Synchronous Overlap Add)算法是在90年代初提出的。基于PSOLA算法的波形编辑技术和早期的波形编辑有原则性的差别。PSOLA算法在编辑和拼接语音波形前能根据上下文的要求，对拼接单元的韵律作出调整，而且音库中的采样波形中保留了一部分原发音人的语音特征，因而合成语音的自然度和清晰度都得到了显著提高。

PLOSA算法对合成基元的超音段特征的调节分为3步：

1) 对原始波形进行分析，产生非参数的中间表示。

2) 对这些短时信号作必要的修正，形成一系列短时合成信号。首先根据原始语音波形的基音曲线和超音段特征与目标基音曲线和超音段特征修正的要求，建立合成波形与原始波形之间基音周期的映射关系，再由此映射关系确定合成所需的短时合成信号系列。

3) 将合成短时信号系列与目标基音周期同步排列并重叠相加得到合成波形。此时，合成语音波形就具有所期望的超音段特征的修改。

2.1 基音同步分析

数字化的语音波形的中间表示形式是由基音同步分析窗$h_m(n)$对原始数据加权得到的短时信号$x_m(n)$：$x_m=h_m(t_m-n)x(n)$

其中，t_m为基音标注点，$h_m(n)$一般取用Hanning窗，其窗函数为：$h_m(n)=0.5-0.5\cos[(2n\pi)/(N-1)]$

窗长大于原始信号的一个基音周期，因此窗间有重叠。窗长一般取为原始信号的基音周期的2～4倍，则有：$h_m(n)=h(n/lp)$

h(n)为归一化窗长，p为基音周期，l为表明窗覆盖基音周期数的比例因子。p既可选分析基音周期p_m，也可选合成基音周期$\overline{p_q}$。一般情况下，选l =2可使合成方法简化；当提高基频时选$p=p_m$，降低基频时选$p=\overline{p_q}$，也可合成简化。

2.2 基音同步修改

短时分析信号$x_m(n)$将修改为合成信号$\overline{x_q(n)}$，同时原始信号的基因标注t_m也相应地改为合成基音标注$\overline{t_q}$，这转换有3个基本操作：1) 对短时信号的数量进行修改；2) 对短时信号之间的延时进行修改；3) 对每个独立的短时信号波形的修改。

基音标注$\overline{t_q}$的数目依赖于音高量和时间量上的修改因子β和γ。任意两个正确的基音标注的间隔就是合成信号的基频。在TD(Time Domain)-PSOLA中，从$x_m(n)$到$\overline{x_q(n)}$的映射只要选择一段$x_m(n)$信号，按延时序列$d_q=t_q-t_m$转换成$\overline{x_q(n)}$：$\overline{x_q(n)}=x_m(n-d_q)=(n+t_m-t_q)$

(1)时间量的修改　时间量的修改可以与音高量同时进行，也可以独立变换。我们用后者。最简单的情况是，时间量修改因子γ为常数，此时，从$\overline{t_q}\to t_m$基音标注的映射简化为寻找最接近γt_q的t_m。当需要减慢语速时，基音标注的映射为几个短时分析信号的重复；相反情况时，为使语速加快，需删去短时信号中的某些波形段。

(2)音高量的修改　音高量的修改总是与时间量的修改互相交叉的，相对来说要复杂一些。最简单的情况是时间量和音高量的修改因子相同：$\beta=\gamma$，则合成基音标注和分析标注成一一对应的关系：$\overline{t_q}\to t_m$。但是一般情况下，时间量和音高量是不相关的，这就要对短时分析信号进行复制或删除，这可看作是两个转换过程的结合：其一，用相同的因子修改音高量和时间量；其二，用因子γ/β对时间量进行补偿。这两步映射可被结合为一个映射，时间量和音高量的修改在一步之内同时完成。由于时间量因子γ/β的倒数比较大，当同时增高浊摩擦音的音高并减慢其语速时，也会产生很小的噪声。

2.3 基音同步叠加法合成

采用原始信号谱与合成信号谱差异最小的最小平方叠加合成法(least-squares overlap-add scheme)：

$$\overline{x_q(n)}=\sum_q \partial_q \overline{x_q(n)}\overline{h_q(\overline{t_q}-n)}/\sum_q \overline{h_q^2(\overline{t_q}-n)}$$

其中，分母是时变单位化因子，是窗之间时变叠加的能量补偿，$h_q(n)$为合成窗序列，α_q为相加归一化因子，是为了补偿音高修改时能量的损失而设的，上式可简化为：

$$\overline{x_q(n)}=\sum_q \partial_q \overline{x_q(n)}/\sum_q \overline{h_q(\overline{t_q}-n)}$$

式中的分母是一个时变的单位化因子：补偿相邻窗口叠加部分的能量损失。该因子在窄带条件下接近于常数，在宽带条件下，当合成窗长为合成基音周期的两倍时该因子亦为常数。此时，若设$\alpha_q=1$，则有：

$$\overline{x_q(n)}=\sum_q \overline{x_q(n)}$$

一般地，除加窗操作之外其它运算都是线形的，因此当PSOLA与LPC滤波或低通滤波等线性滤波结合使用时，不能颠倒计算顺序。

3 采用PSOLA算法的汉语文-语转换系统实验

3.1 语音库的建立

本实验的语音库利用声霸卡CREATIVE-SB16的功能建立，系统选用16kHz的采样频率，16为量化单位。音库除了包括1281个汉语有调音节外，还有英语字母、数字、标点和各种符号的发音，共1340个声音文件。

3.2 语言学处理

对本系统的输入先进行预处理、断句和分词处理、建立停顿规则、文本替换规则及多音字的处理。为了使计算机合成的语音也能抑扬顿挫、可表达一定语气的发音，要建立完善的韵律规则库。本系统主要建立了重音的韵律规则。

3.3 利用PSOLA对重音的韵律进行修正

本TTS系统的输入文本经过各项语言学、语音学处理后得到的拼音码还不能直接发音，下一步要进行波形编辑，波形编辑时使用了基于PSOLA的语音合成技术。

韵律修正处理中除了要用到有汉语带调音节采样数据构成的语音库外，还要用到基音同步标记库。这个库中存放音节波形的基音同步标注文件，每个基音标注文件与音库中相应的音节波形编码文件一一对应，其中内容是音节波形的基音周期的个数和每个基音周期的起始点的位置序列。

韵律修正中，首先根据拼音码从语音库中提取相应的语音基元的波形数据，再根据重音韵律控制符得到PSOLA算法中所需的音高、音长修正因子，而后就可利用TD-PSOLA算法对各音节波形进行基音同步调节，使其满足韵律规则库中所确定的该音节的音高、音长和音强。经过实时韵律修改的波形数据在滑滤波后即可经发声设备输出。

经实验测试，对于语音合成的简单句，经过重音韵律规则和PSOLA算法处理后的语音输出，在音字、音调、字与字、词与词之间的平滑上比未经过处理的有很大进步。

4 小结

90年代后，随着PSOLA算法的提出，语音合成技术获得了很大提高。并因修改所针对的侧面不同，提出了TD(Time Domain)-PSOLA，FD(Frequence Domain)-PSOLA和LP(Linear Prediction)-PSOLA等几种不同的算法。语音合成技术正与语音识别技术一起一步步走向成熟及实用。

参考文献

1 Moulines E，Charpentier F. Pitch-synchronous Waveform Processing Techniques for Text-to-speech Synthesis Using Diphones. Speech Communication，1990，9：453-456

2 Hamon C，Moulines E，Charpentier F A. Diphone Synthesis System Based on Time-domain Prosodic Modifications of Speech. Proc. Int. Conf. ASSP，1989：238-241

3 Lee LS.The Synthesis Rules in a Chinese Text-to-speechSystem. IEEE Trans. ASSP，ASSP-37，1989(9):1309-1320

4 林焘，王理嘉. 北京语音实验录. 北京：北京大学出版社，1985

※※

(上接第18页)

入多个条件，还是对同一种攻击程序调用多次(不同脚本名，不同程序参数，或不同后续操作步骤等)，只要其搜索的关键字串相同，它们的结果也应该相同。唯一不同的就是前者对攻击的可执行代码只运行一次，虽节约时间，但灵活性小；而后者是将可执行代码运行两次，虽耗时，但灵活性大，有时甚至非要这样不可。再加上NSATS的并行测试机制，使得后者所占用的测试时间也可大大缩短。

例如，用户可能扩充NSATS以搜索本文前面讲的FTP服务器bounce(反弹)问题。ftpbounce.nsats的目标在于查出远程ftpd服务器是否允许一个客户指定任意远程客户IP地址及TCP端口以便进行文件传输。如果ftpd允许一个不同于初始源的带有IP地址的PORT命令，以及一个被保留的TCP端口，那么该ftpd存在bounce问题。

构造ftpbounce.nsats的最快途径是利用本系统所提供的交互式辅助工具来完成此项任务。

这里所需要做的仅仅是查找试图发出PORT命令的200应答，如表1所示填写页面上的参数：

表 1 新增方法的输入参数

Arguments	-nv
Executable file	ftp
Operation Steps	Open $target
	Quote user anonymous Quote pass ftp@ftp Quote port 1,2,3,4,0,25 Quit
Key String 1	200 PORT command successful
$SecurityLevel	3
$Status	A
$ServiceOutput	BOUNCE
$InternalText	Provide ftp server bounce

然后，按照页面上的提示继续操作，将此项检测手段列入重型攻击列表中。最终，ftpbounce.NSATS脚本就会被列在NSATS.conf的再度扫描列表中。生成的"ftpbounce.NSATS"文件如下：

```
open (FTPBOUNCE,"ftp -nv <<EOF
open $target
quote user anonymous
quote pass ftp@ftp
quote port 1,2,3,4,0,25
quit
EOF ") || die "Cannot run bin/ftpbounce";
While (<FTPBOUNCE>){
        If (defined ($opt_v)) {
                Print;
        }
        if (/^200 PORT command successful/) {
                $status = "a";
                $SecurityLevel = "3";
                $ServiceOutput = "BOUNCE";
                $InternalText = "Provide ftp
server bounce";
        }
};
```

如果有可能的话，应该将所增加新方法的说明用HTML方式存在pages\vuls\ 目录下。文件名就取与攻击方法名相同的文件名加上".HTML"。这样，在报表生成和漏洞检索时会非常有用的。

3.4 报表接口

报表生成的总体框架主要利用HTML的超链接特性，将各种主机信息、各种安全漏洞和主机间的信任关系联系起来。用户可以从任意一种信息出发，而遍历所有的检测结果。同时，由于客户端是基于浏览器的，所以可以方便地提供打印和将报表用E-mail形式发送的功能。

报表的生成是由一系列的Perl程序完成的。每一种结果的排序方式对应一个Perl程序。当客户端发出请求时，服务器端就调用相应程序。

4 结束语

NSATS提供了方便的挂靠接口技术,随着安全攻击手段的不断更新，在具体应用中，应收集尽可能多的攻击方法，以提高系统安全检测水平。另外，NSATS还需不断地完善与改进，如人工智能测试技术和差别测试技术的应用等。