

基于 TD-PSOLA 算法的汉语普通话韵律合成

张后旗¹, 俞振利², 张礼和²

(1. 浙江财经学院经济信息管理系, 浙江 杭州 310012; 2. 浙江大学信息与电子工程系, 浙江 杭州 310028)

摘 要: 结合汉语普通话的韵律特征, 采用 TD-PSOLA 算法实现了汉语普通话的韵律合成, 并对合成语音和原始语音的韵律参数作了比较分析. 实验结果表明, 这种方法能够有效地控制语音韵律参数, 实现较高质量的语音韵律合成.

关键词: 语音合成; 韵律参数; 时域基音周期叠加算法

中图分类号: TN912.3

文献标识码: A

文章编号: 1001-7119(2002)01-0006-04

0 引 言

在汉语普通话文语转换系统中, 韵律合成是保证合成语音具有良好自然度和可懂度的关键. 汉语普通话是一种具有声调和语调的语言, 因此在实际的韵律合成过程中, 既要进行时间尺度修改, 又要进行基频尺度修改. 显然, 这种修改尺度并不是均匀的. 通常可以先进行时间尺度修改, 然后再进行基频尺度修改. 由于对语音信号进行分帧处理, 如果分别进行时间、基频尺度修改, 必然导致运算开销太大, 很难满足韵律合成的实时要求. 为此, 我们结合汉语普通话的韵律特征^[1,2], 应用动态合成调型的方法, 采用时域基音同步叠加算法同时进行时间、基频尺度修改, 实现了较高质量的韵律合成.

1 时域基音同步叠加(TD-PSOLA)算法

基音同步叠加(PSOLA)算法的实质是将短时傅里叶变换方法中声源滤波器分解和声源谱修改这两个步骤合作一步来完成^[3,4]. 用 PSOLA 方法实现韵律修改时基本上按三个步骤进行. 第一步是分析阶段, 将语音信号波形分解成一系列短时分析信号 $x(t_a(s), n)$

$$x(t_a(s), n) = h_s(n)x(n - t_a(s)) \quad (1)$$

式中 $t_a(s)$ 为第 s 帧分析时刻或称作分析基音标记. 它的设置在浊音部分与基音周期同步, 在清音部分以恒定速度进行; $h_s(n)$ 为第 s 帧分析窗, 通常选 hanning 窗、hamming 窗或 Bartlett 窗, 窗长与该分析时刻所对应的合成时刻合成基音周期 $p(u)$ 成正比, 即 $T = \mu p(u)$. 对于时域

收稿日期: 2001-06-01

基金项目: 国家自然科学基金资助项目(69972064)

作者简介: 张后旗, 男, 1972 年生, 安徽无为, 硕士; 俞振利, 男, 1956 年生, 浙江余杭, 博士.

基音同步叠加(TD-PSOLA)算法,比例系数 μ 选为 2.

第二步是修改阶段,将短时分析信号序列转变为一组修改了的短时合成信号. 这些短时合成信号序列与一套新的合成信号基音标记同步. 这个步骤涉及到短时信号序列数量的更改、短时信号序列之间延迟的更改以及每一个短时信号波形可能要发生的更改. 在我们所采用的 TD-PSOLA 方法中,短时合成信号是由相应的短时分析信号直接拷贝而来. 若短时分析信号为 $x(t_a(s), n)$, 短时合成信号为 $x(t_s(u), n)$, 则有

$$x(t_s(u), n) = x(t_a(s), n) \quad (2)$$

式中 $t_a(s)$ 为分析基音标记, $t_s(u)$ 为合成基音标记.

第三步是合成阶段. 有多种叠加合成方程可以用来合成最终的合成信号. 例如,应用 LSEE-MSTFT 方案进行估计,便可以得到合成信号

$$y(n) = \frac{\sum_u y_w(u, n - t_s(u))}{\sum_u f_u(n - t_s(u))} \quad (3)$$

式中分母为时变的归一化参数,用来补偿由于相继两帧间不同程度的叠加所造成的能量修改;分子为短时合成信号.

我们所运行的算法流程如下:输入原始语音波形、分析基音标记、调域、调型序号、目标时长、停顿与否标记;根据输入的调域和调型给出基频尺度修改系数;根据输入的目标时长和原始音节的时长给出时间尺度修改系数;令第一个合成基音标记等于第一个分析基音标记;从第一个分析基音标记给出第一帧短时合成信号;根据基音轮廓和基音修改系数递推出下一个合成虚拟时刻最接近的分析基音标记,给出这一帧短时合成信号;如是最后一帧,则叠加合成所有的短时合成信号,否则继续递推算出下一个合成基音标记,直至给出最后一帧短时合成信号.

2 韵律合成实验系统的模块结构

基于 TD-PSOLA 算法的韵律合成实验系统的模块结构如图 1 所示. 首先将汉语普通话所有的音节录入音库,并将一些儿化音当作音节录入音库. 然后对这些音节进行人工基音标记,并对声母的清音部分按均匀间隔进行标记. 这样,音节的所有这些标记便构成分析时刻序列,并将其存入基音标记库.

语音学处理模块对输入的文本进行语义分析,并根据汉语语句发音的词调规则和语调规则给出每个音节的调域、调型曲线以及时长等参数.

韵律合成模块根据语音学处理模块给出的这些韵律参数对原始音库中的音节进行韵律合成(韵律修改),最后将修改后的所有音节进行拼接,构成合成语句.

在韵律合成模块中,我们应用 5 个模块来实现韵律合成. Dateinput 模块分别从音库及基音标记库中读入语音波形数据和音节的分析时刻序列, Pitchmodi 模块根据目的调型和原始分析时刻给出基频修改系数和分析基音轮廓, Timemodi 模块根据目的时长和音节原始时长给出时长修改系数, Synpmark 模块根据基频修改系数、分析基音轮廓、时间修改系数和原始分析时刻给出合成时刻和虚拟分析时刻, Synthesis 模块根据各合成时刻对应的虚拟分析时刻找出与它最接近的分析时刻及相应的短时合成信号,最后按式(3)叠加合成得到该音节的韵律

合成信号.

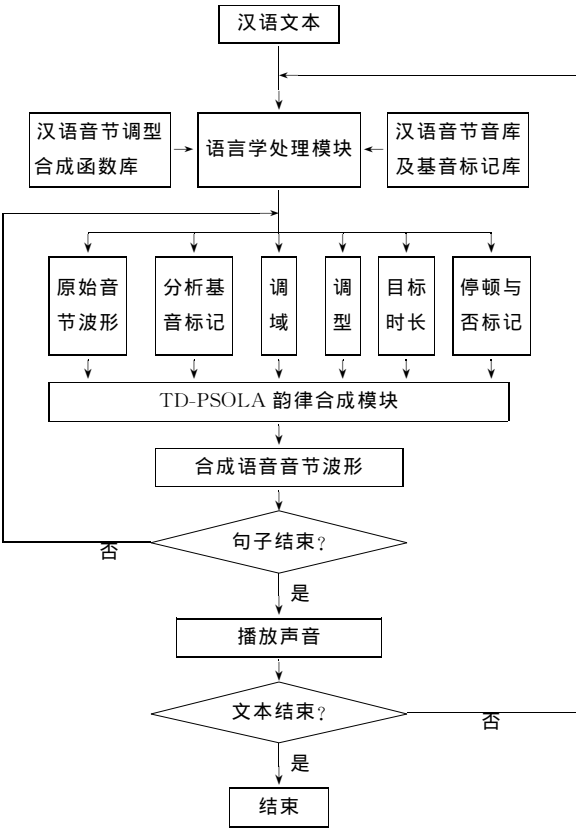


图 1 韵律合成实验系统模块结构

3 实验结果与分析

我们首先应用上述方法分析了多种汉语普通话的典型发音语句的主要韵律参数. 其中语音“迅猛发展”的基音轮廓如图 2 所示, 音节的调型曲线和调域以及时长如表 1 所示.

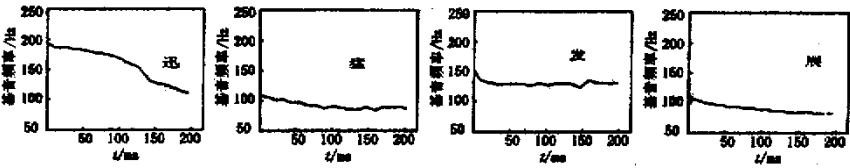


图 2 样句基音轮廓图

在合成语音之前,单独录下样句的单个音节,人工记录它们的基音标记,给出分析时刻,存于库中,建立基音标记库.在合成语音时运用图 1 所示模块结构,按表 1 所列韵律参数对音库中的单音节进行韵律修改,最后拼接成句.

我们对上述合成语音进行了韵律参数提取,其基音轮廓如图 3 所示. 比较图 2 和图 3 可见,合成语句的基音轮廓与样句的基音轮廓相符. 表 2 列出了输入的韵律参数与实际合成获得的韵律参数. 通过比较可以看出,文中给出的韵律合成方法能够相当精确地实现韵律参数修改.

表 1 样句韵律参数			
音节	调型	调域/Hz	时长/ms
迅	51	110~190	271
猛	211	85~185	224
发	55	118~118	212
展	211	85~205	190

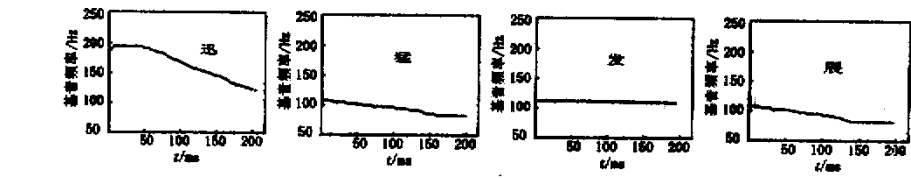


图 3 合成语音基音轮廓图

表 2 合成语音韵律参数				
音节	输入时长/ms	输入调域/Hz	输出合成信号 时长/ms	输出合成信号 调域/Hz
迅	271	110~190	271	110~190
猛	224	85~185	224	86~178
发	212	118~118	212	119~119
展	190	85~205	190	86~194

为了评估该方法的韵律合成质量,由 20 名听测人首先在未知内容的情况下对合成语句进行辨识实验,并对语调的自然度以及合成语句和样句的相似度进行评估. 听测结果表明,该方法具有较高的韵律合成质量,其结果如表 3 所示.

表 3 合成语句的质量评估				%					
听写正确率	合成语句自然度					语调相似度			
	自然	较自然	可忍受	难忍受	极不自然	相似	较相似	不太相似	极不相似
98	20	65	10	5	0	35	55	10	0

4 结 论

本文提出了一种基于 TD-PSOLA 算法的韵律合成方法. 该方法根据汉语普通话的音系特征,采用动态合成调型的方法,即根据输入的调型序号、调域和时长,动态地合成目标基音轮廓曲线,从而能够考虑汉语音节中清音的无调特性,使同一音节在清音段相邻合成基音标记间隔等于分析基音标记间隔,目标基音轮廓特征全部加于浊音段,并且实现时间和基频尺度的同时修改. 因此,这种方法不仅节约了运行开销,也减少了因计算合成基音标记不准确而产生的合成韵律参数的误差.

(下转第 13 页)

- [7] 金丽琴, 吕建新. 细脚拟青霉对小鼠血清及肝、脾、肾组织酸性磷酸酶活力的影响[J]. 温州医学院学报, 1995, 25(2): 83~85.
- [8] 吕建新, 金丽琴, 楼永良等. 细脚拟青霉对小鼠体液免疫功能的影响[J]. 中国实验临床免疫学杂志, 1996, 8(6): 8~10.

Effects of *Paecilomyces Tenuipes* Total Polysaccharides on Chief Biochemistry Index in Urine of Rats

JIN Li-qin¹, LU Jian-xin², YUAN Qian², LI An-le³

(1. Department of Biochemistry, Wenzhou Medical College, Wenzhou 325027, China;

2. School of Laboratory Medicine and Public Health, Wenzhou Medical College, Wenzhou 325027, China;

3. Centre of Xperimental Animal, Wenzhou Medical College, Wenzhou 325027, China)

Abstract: For studying the regulation of *paecilomyces tenuipes* total polysaccharides on substance metabolism of rats, the levels of trace-protein, urea nitrogen, creatinine, uric acid, calcium, magnesium, phosphorus, potassium, sodium, chloride, and γ -glutamyltransferase in the urine of rats were detected by ri-li 7170 automatic biochemistry analyzer. Compared with control group, the levels of UN, Cr, UA, P and GGT in the urine of rats showed remarkable differences. The levels of UN and Cr were increased and the levels of UA, P and GGT were decreased in the urine of rats induced by *paecilomyces tenuipes* total polysaccharides.

Key words: *paecilomyces tenuipes*; polysaccharides; urine; biochemistry index

(上接第 9 页)

参考文献:

- [1] 周同春. 汉语语音学[M]. 北京: 北京大学出版社, 1989.
- [2] 沈 炯. 北京话的声调和语调[M]. 北京: 北京大学出版社, 1985. 72~130.
- [3] Valbret H, Moulines E & Tubach J P. Voice transformation using PSOLA techniques[J]. Speech Communication, 1992, 11(2): 175~187.
- [4] Moulines E, Laroche J. None parametric techniques for pitch scale and time scale modification of speech[J]. Speech Communication. 1995, 16(2): 175~207.

Prosodic Synthesis of Chinese Speech Based on TD-PSOLA Algorithm

ZHANG Hou-qi¹, YU Zhen-li², ZHANG Li-he²

(1. Dept. of Informaiton Management, Zhejiang University of Finance & Economics, Hangzhou 310012, China;

2. Dept. of Information & Electronic Engineering, Zhejiang University, Hangzhou 310028, China)

Abstract: At the basis of Chinese speech prosodic feature, a high quality Chinese speech synthesis method using TD-PSOLA algorithm was put forward and realized. The synthetic speech prosodic parameters were analyzed and compared with objective prosodic parameters. Comparing the synthetic speech and the actual speech, it was found that the ability of this method to control prosodic parameters and its effect on synthetic speech quality was evaluated.

Key words: speech synthesis; prosodic parameters; pitch synchronous overlap add algorithm