基于混合幅度差函数的基音提取算法

刘 建,郑 方,邓 菁,吴文虎
(清华大学信息科学技术国家实验室语音技术中心,北京 100084)

摘 要: 为了减少基音周期提取中的倍频和半频错误,进行更准确的基音周期估计,提出一种基于混合幅度差函数的基音周期提取方法. 分析比较了不同幅度差和自相关函数估计基音周期错误率的分布,结合两类典型幅度差函数的优点定义了混合幅度差函数;进而基于混合幅度差函数,给出了使用历史信息进行校正的后处理方法. 分析表明,所提方法可提高基音周期估计准确率,接近实时地确定基音周期,减少了传统基音周期估计因平滑处理而带来的误差或者动态规划处理带来的延迟. 大量实验表明本文提出的基音周期提取方法比传统方法的错误率降低了 13.8 %.

关键词: 语音信息处理; 基音周期提取; 混合幅度差函数
中图分类号: TP391 **文献标识码**: A **文章编号**: 0372-2112 (2006) 10-1925-04

Combined Magnitude Difference Function Based Pitch Tracking Algorithm

LIU Jian, ZHENG Fang, DENG Jing, WU Wen-hu
(Center for Speech Technology, National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: To reduce the halving and the doubling errors in pitch tracking, an algorithm based on a combined magnitude difference function is proposed in this paper. By analyzing the error distributions in different algorithms based on magnitude difference functions and autocorrelation functions, a combined magnitude difference function (CMDF) is defined, which has the best performance compared with the others. Moreover, a post processing method is proposed which uses the history information to adjust the raw pitch estimated using the CMDF. The post processing method uses only the history information, thus the pitch tracking algorithm has low latency, which reduces the pitch estimation errors that often occur in traditional methods. Experiments show that the pitch tracking algorithm proposed in this paper can achieve a pitch error rate reduction of 13.8 % compared with the baseline traditional algorithm.

Key words: speech processing; pitch tracking; combined magnitude difference function

1 引言

基音周期是语音信号中的重要参数之一,它在语音识别、语音合成和语音编码中有广泛的应用. 语音信号是一种非平稳时变信号,其中浊音部分在一段相对短的时间内可以认为是准周期的,因此语音信号处理中通常采取短时处理技术. 常用的短时基音周期估计方法有 AMDF[1] (Average Magnitude Difference Function) 法和 ACF[2] (AutoCorrelation Function) 法. 当基音周期的估计值大于基音周期实际值时,认为发生偏长错误,即半频错误;反之,认为是偏短错误,即倍频错误. 利用传统的 AMDF 或 ACF 估计基音周期的结果一般都存在大量错误,相关文献提出了许多改进方法,其中:文[3]提出变长度 AMDF (LVAMDF),文[4]提出循环 AMDF(CAMDF),文[5]提出了改进的 ACF,文[6]提出了幅度差平方和函数,文[7]对幅度差平方和函数进行改进和推广. 不同文献中估计基音周期时利用的函数都有各自的优缺点,使用它们估计基音周期的错误率分布也不同,利用某一个固定的函数进行基音周期估计在特定情况下难免出现错误.

为此,本文首先在实验的基础上分析了各种函数在基音周期估计时错误率的分布特点,进而定义了一种混合幅度差函数,充分利用两类不同幅度差函数的优点,实验表明利用混合幅度差函数可有效降低基音周期估计的错误率. 在此函数的基础上提出一种低延迟的基音周期估计算法,进一步降低基音周期提取的错误率,避免中值平滑在语音过渡段产生的额外误差或是动态规划处理[8]造成的较高延迟.

2 实验设计

实验使用的评测数据库是文献[6]中提到的 Edinburgh 大学收集的一个基音周期提取评测数据库. 此数据库包括 100 个句子,其中男声和女声各 50 句,每个句子都有标准的基音周期标注. 这些句子中包括基音周期估计中容易出错的各种情况,如鼻音等.

收稿日期:2005-12-05;修回日期:2006-04-25

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

实验测试过程中,语音信号首先分成若干语音帧,帧长是 24ms 或者 48ms,帧移 12ms,即每隔 12ms 进行一次基音周期估计,基音周期限制在 2ms～20ms. 对于存在基音周期的浊音段,如果基音周期估计值偏离标准值超过 20 %,则认为出现估计错误.

实验中评估方法有 2 种:(1) 根据标准基音周期标注,只考虑存在基音周期的语音帧,计算出不同函数在这些帧上的基音周期估计错误率,用来衡量估计基音周期函数自身的性能. 总错误率 = 基音周期估计出错的帧数/标准标注中浊音的总帧数;偏长错误率 = 基音周期估计发生偏长错误的帧数/标准标注中浊音的总帧数;偏短错误率定义类似. (2) 考虑所有语音帧,计算出清浊音误判率,只对正确判定为浊音的语音帧统计基音周期估计错误率,用来衡量基音周期提取算法整体的性能. 清浊误判率 = 清音误判为浊音的帧数/标准标注中浊音的总帧数;浊清误判率 = 浊音误判为清音的帧数/标准标注中浊音的总帧数;总错误率 = 基音周期估计出错的帧数/正确判定为浊音的总帧数.

3 不同基音周期估计函数比较

下面比较 AMDF,LVAMDF,CAMD,改进的 ACF 和幅度差平方和函数. 下面是本文中实验使用的各种基音周期估计函数的定义. AMDF[1]定义为

$$d_1^t(\tau)=\frac{1}{N}\sum_{j=t}^{t+N-1}|s(j)-s(j+\tau)| \tag{1}$$

LVAMDF[3]定义为

$$d_2^t(\tau)=\sum_{j=t}^{t+\tau-1}|s(j)-s(j+\tau)|\Big/\left(\frac{1}{2}\sum_{j=t}^{t+2\tau-1}|s(j)|\right) \tag{2}$$

CAMDF[4]定义为

$$d_3^t(\tau)=\frac{1}{N}\sum_{j=0}^{N-1}|s_t(\mathrm{mod}(j+\tau,N))-s_t(j)| \tag{3}$$

改进的 ACF[5]定义为

$$d_4^t(\tau)=\frac{-1}{N-\tau}\sum_{j=0}^{N-\tau-1}s_t(j+\tau)s_t(j) \tag{4}$$

幅度差平方和函数[6]定义为

$$d_5^t(\tau)=\sum_{j=t}^{t+N-1}(s(j)-s(j+\tau))^2 \tag{5}$$

其中:$s(j)$ 是离散化的语音采样序列,N 是一帧语音中采样点的个数,$s_t(j)=\begin{cases}s(t+j), j=0,1,\cdots,N-1\\ 0,\text{其他}\end{cases}$. 式(4)相对原公式增加了一个负号,为了保证估计基音周期的方法与其他函数一致. AMDF 和幅度差平方和函数中,N 是 24 ms 帧长对应的采样点数. CAMDF 和改进 ACF 中,N 是 48ms 帧长对应的采样点数. 一帧语音基音周期的估计值可以由式(6)确定.

$$P=\arg\min_{\tau=P_{\min}}^{P_{\max}}(d(\tau)) \tag{6}$$

其中:arg min 表示函数达到最小值时自变量的取值,$P_{\max}$和 $P_{\min}$是语音基音周期最大和最小的可能取值,实验中取值分别为 20ms 和 2ms 对应的采样点个数,$d(\tau)$是估计基音周期使用的某种函数. 一般情况下 P 和基音周期是一致的,但是在实际应用中如果仅把 P 作为基音周期,难免会出现基音周期估计偏长或偏短错误. 表 1 是在实验评估方法(1)下得到的实验结果.

表 1 传统基音周期估计函数比较

方法	偏长错误率/ %	偏短错误率/ %	总错误率/ %
AMDF	4.96	**1.23**	6.19
LVAMDF	13.0	3.86	16.9
CAMDF	**1.16**	2.45	**3.61**
改进的 ACF	6.54	1.80	8.34
幅度差平方和函数	5.51	1.48	6.99

表 1 中每列的最小值都用黑体标识. 可以看出,基音周期估计偏短错误率最低的是 AMDF,偏长错误率最低的是 CAMDF,总错误率最低的是 CAMDF,其他函数的总错误率明显偏高. 对于基音周期偏短错误率,CAMDF 明显偏高,原因在于 CAMDF 取模循环之后计算的各项对偏长错误率有显著的抑制作用,但会引发额外的偏短错误率. AMDF 和 CAMDF 的错误率分布某种意义上是互补的.

4 混合幅度差函数

从以上实验结果可知利用 CAMDF 估计基音周期偏长错误率最低,利用 AMDF 偏短错误率最低. 针对 AMDF 和 CAMDF 估计基音周期错误率的分布特点,可以将两者结合,使基音周期偏长错误率和偏短错误率都达到最小值. 为了使 AMDF 和 CAMDF 有效的结合,先定义它们的归一化形式. 归一化的 AMDF 定义为

$$d_{\mathrm{amdf}}^t(\tau)=\frac{\sum_{j=t}^{t+N-1}|s(j)-s(j+\tau)|}{\sum_{j=t}^{t+N-1}|s(j)|+\sum_{j=t}^{t+N-1}|s(j+\tau)|} \tag{7}$$

归一化的 CAMDF 定义为

$$d_{\mathrm{camdf}}^t(\tau)=\frac{\sum_{j=0}^{N-1}|s_t(\mathrm{mod}(j+\tau,N))-s_t(j)|}{2\sum_{j=0}^{N-1}|s_t(j)|} \tag{8}$$

因为不等式 $\sum_{j=t}^{t+N-1}|s(j)|+\sum_{j=t}^{t+N-1}|s(j+\tau)|\geq\sum_{j=t}^{t+N-1}|s(j)-s(j+\tau)|\geq 0$ 成立,所以归一化 AMDF 和 CAMDF 函数取值区间都是[0,1]. 混合幅度差函数定义为

$$D^t=\alpha d_{\mathrm{amdf}}^t+(1-\alpha)d_{\mathrm{camdf}}^t \tag{9}$$

其中,α 是一个插值参数,混合幅度差函数可以认为是 AMDF 和 CAMDF 的线性插值. 表 2 是在实验评估方法(1)下得到的实验结果.

表 2 归一化 AMDF、CAMDF 和混合幅度差函数比较

方法	偏长错误率/ %	偏短错误率/ %	总错误率/ %
归一化 AMDF	4.32	1.38	5.70
归一化 CAMDF	1.16	2.45	3.61
混合幅度差函数($\alpha=0.35$)	1.49	1.79	3.28

由表 2 看出,混合幅度差函数具有最低的基音周期估计错误率,其错误率比 AMDF 降低 42.5 %,比 CAMDF 降低 9.14 %. 混合幅度差函数的性能是受插值参数 α 影响的,不同

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

α取值下混合幅度差函数的基音周期估计错误率曲线如图 1. 当 $\alpha=0$ 时,混合幅度差函数退化成归一化的 CAMDF;当 $\alpha=1$ 时,混合幅度差函数退化成归一化的 AMDF. 取适当的 α 可以得到基音周期估计错误率最低的混合幅度差函数,一般情形 α 可取 0.30～0.40,本文下面实验 α 均取值 0.35.

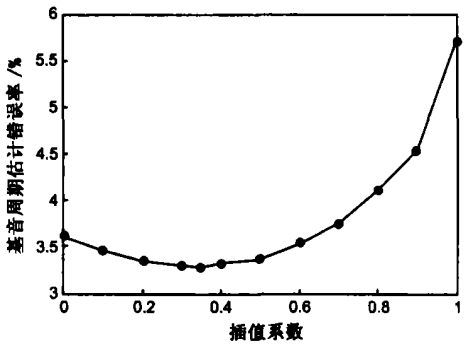

图 1　插值系数 α 不同取值时利用混合幅度差函数估计基音周期的错误率曲线

5　基音周期提取算法

5.1　清浊判定

为了便于清浊判定,增加一个辅助的函数,定义如下

$$V^t(\tau)=\begin{cases}1, \tau=0\\ D^t(\tau)\Big/\sum_{i=1}^{\tau}D^t(i), \text{其他}\end{cases} \tag{10}$$

清浊判定有两个判别阈值 s_D 和 s_V,当 $D^t(\tau)<s_D$ 且 $V^t(\tau)<s_V$,则认为是浊音,否则认为是清音或静音. 为了防止静音的干扰,还有一个能量判定阈值 e,如果当前语音帧平均能量 $\frac{1}{N}\sum_{j=0}^{N-1}s_t(j)<e$,则认为是静音,可防止能量过小造成计算 $D^t(\tau)$ 或 $V^t(\tau)$ 产生较大误差,一般 e 取值 100～300 即可. 如果当前语音帧的基音周期是 P^t,$D^t(P^t)$ 的值可以反映非周期性能量在总能量中所占的比重,$V^t(P^t)$ 可以反映函数 $D^t(\tau)$ 在基音周期位置 $\tau=P^t$ 谷值点的突出程度. 一般情况,s_D 可以取 0.4～0.6,s_V 可以取值 0.6～0.8.

5.2　中值校正

在处理实际语音信号时,以目前技术水平使用任何函数包括 AMDF、CAMDF 和混合幅度差函数等,都不能保证基音周期估计百分之百的正确. 为了进一步降低错误率,一般采用后处理对基音周期初步估计结果进行平滑. 后处理虽然能纠正部分基音周期估计错误,但往往会有以下 2 个问题:(1) 造成基音周期最终结果有较大误差,如:中值平滑,采用若干个点的中间值做为最终结果,由于语音信号只在短时稳定,特别是对一些过渡音使用这种方法会造成一定的偏差;(2) 造成基音周期估计不能实时完成,有较大的延迟,如:动态规划的平滑方法[7],这种方法用基音周期全局的信息,纠正基音周期的局部错误,每帧语音信号都需要保留若干个候选值. 用动态规划方法往往会得到相对满意的结果,但基音周期最优值在语音信号结束才能得到,造成较大延时.

针对以上问题,本文提出中值校正的方法. 如果当前处理第 t 帧语音,那么采用 $t-j$ 帧到 $t-1$ 帧之间距离第 t 帧语音最近的 k 个已经确认为浊音的基音周期的中值作为参考值,记参考值为 R^t. 注意 $j\geq k$,实际取值 $j=2*k$. 如果当前语音帧混合幅度差函数的最小值点是 P^t,那么可以利用式(11)得到 R^t 和 P^t 的近似比例关系 β.

$$\beta=\arg\min_m|R^t-mP^t|,\ P_{\min}\leq mP^t\leq P_{\max} \tag{11}$$

其中:$m=1,2,1/2,3,1/3,\cdots$;根据 β 可以利用式(12)得到当前语音帧基音周期的校正值 $\hat{P}^t$,如下

$$\hat{P}^t=\arg\min_{\tau}(D^t(\tau)),\ 0.75\beta P^t\leq\tau\leq 1.25\beta P^t \tag{12}$$

式(11)和(12)在实际应用中需要注意以下 5 点:(1) 清音的基音周期在算法中被设为 0,计算 R^t 时只用已经确认为浊音的语音帧,不到 k 帧不进行中值校正. (2) 如果当前帧的 $P^t\leq P_{\min}$ 或者 $P^t\geq P_{\max}$,则不进行中值校正. (3) $0.75\beta P^t$ 或 $1.5\beta P^t$ 如果超出 $P_{\min}$ 或 $P_{\max}$ 的范围,则使用 $P_{\min}$ 或 $P_{\max}$ 作为边界. (4) 如果在 $0.75\beta P^t$ 和 $1.5\beta P^t$ 之间不存在局部极小值点,则不进行中值校正,仍以 P^t 作为结果. (5) 进行中值校正后,不再进行清浊判定,以 P^t 判定的结果为准.

使用中值校正的方法与中值平滑的方法不同,中值平滑是使用若干个点的中值点作为基音周期的结果. 而中值校正方法中值点只是作为一个参考值,最终基音周期的估计值还是需要根据混合幅度差函数进行计算,避免了中值平滑用若干语音帧的中值近似作为前语音帧基音周期所带来的误差. 另外,中值校正可以采用更大的 k 值,从而充分利用历史信息.

中值校正的方法只利用了历史信息,而不需要使用当前语音帧以后的信息,所以其时间延迟很小,几乎是实时的. 中值校正主要是为了消除浊音段中的基音周期偏长或者偏短错误,但是不能改善清浊误判率.

6　实验结果

为了测试基音周期提取算法的可靠性,做了充分的比较实验,验证混合幅度差函数性能的实验前面已经叙述. 下面实验采用的基准系统是 Praat[5] 提供的分析基音周期的工具,估计基音周期使用的脚本是"To Pitch …0.012 50 500",即帧移是 0.012s,最小可能基音频率是 50Hz,最大可能基音频率是 500Hz. 基准系统是目前较实用的系统,其算法利用文[8]的动态规划方法进行了后处理. 表 3 是在实验评估方法(2)下得到的实验结果.

表 3　基于混合幅度差函数基音提取和基准系统比较(C 表示只使用混合幅度差函数,C+M 表示使用混合幅度差函数并利用 5 点中值平滑,C+A 表示使用混合幅度差函数并利用 5 点中值校正,$s_D=0.6$,$s_V=0.8$,$e=100$)

方法	清浊误判率/%	浊清误判率/%	总错误率/%	偏长错误率/%	偏短错误率/%
Praat	6.55	6.97	1.38	0.73	0.65
C	11.7	6.71	1.40	0.63	0.77
C+M	10.9	6.21	1.25	0.58	0.67
C+A	11.7	6.71	1.19	0.44	0.75

从表 2 可以看出,基于混合幅度差函数估计基音周期性

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

能是较高的. 基于混合幅度差函数并用 5 点中值平滑的基音周期估计总错误率比基准系统降低 9.42 %,使用 5 点中值校正的总错误率比基准系统降低 13.8 %,浊音误判为清音的错误率也比基准系统低,但是清音误判为浊音的错误率较基准系统高,有待改进.

7 结论

本文针对估计基音周期使用的不同函数的优缺点,在分析了不同函数错误率分布的基础上,定义了混合幅度差函数,结合了 AMDF 和 CAMDF 的优点. 对比实验表明,利用混合幅度差函数估计基音周期的错误率明显低于 AMDF 和 CAMDF. 在此基础上,提出了清浊判别的准则和中值校正的后处理方法,进一步提高了基音周期检测的精度,中值校正方法比中值平滑方法的实验结果要好. 本文算法仍然存在一些基音周期估计错误,主要是发生在清音浊音过渡段. 另外,清浊判别错误率还是比较高,特别是清音误判为浊音的错误率较高,需要进一步研究.

参考文献:

[1] Ross M, Shaffer H, Cohen A, et al. Average magnitude difference function pitch extractor[J]. IEEE Trans on Acoustics, Speech, and Signal Processing, 1974, 22(5): 353 - 362.

[2] Rabiner L R, Schafer R W. Digital Processing of Speech Signals [M]. Englewood Cliffs: Prentice Hall, 1978.

[3] 顾良,刘润生. 高性能汉语语音基音周期估计[J]. 电子学报, 1999, 27(1): 8-11.
GU Liang, LIU Runsheng. High-performance mandarin pitch estimation[J]. Acta Electronic Sinica, 1999, 27(1): 8 - 11. (in Chinese)

[4] 张文耀,许刚,王裕国. 循环 AMDF 及其语音基音周期估计算法[J]. 电子学报, 2003, 31(6): 896 - 890.
ZHANG Wenyao, XU Gang, WANG Yuguo. Circular AMDF and pitch estimation based on it[J]. Acta Electronic Sinica, 2003, 31(6): 896 - 890. (in Chinese)

[5] Paul B. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound [A]. Proc Institute of Phonetic Sciences 17[C]. Amsterdam: UVA, 1993. 97 - 110.

[6] Cheveign A D, Kawahara H Yin. A fundamental frequency estimator for speech and music [J]. J Acoust Soc Am, 2002, 111(4): 1917 - 1930.

[7] S Sood, A Krishnamurthy. A robust On-The-Fly Pitch (OTFP) estimation algorithm [A]. Proc of ACM Multimedia[C]. 2004. 280 - 283.

[8] Secrest B, Doddington G. An integrated pitch tracking algorithm for speech systems[A]. Proc ICASSP[C]. Boston, MA: IEEE, 1983. 1352 - 1355.

作者简介:

刘 建 男,1978 年 11 月生于北京,2001 年毕业于清华大学计算机科学与技术系,现为清华大学计算机系博士研究生,从事语音信号分析、语音识别技术等方面的研究. E-mail: liuj@cst.cs.tsinghua.edu.cn

郑 方 男,1967 年 3 月生于江苏,分别于 1990 年、1992 年和 1997 年获得清华大学计算机应用专业学士学位、硕士学位和博士学位,研究员,清华大学信息技术研究院副院长,清华信息科学技术国家实验室语音技术中心主任,IEEE 高级会员,主要从事语音识别、说话人识别与语言理解等方面研究,负责或作为骨干人员参与研发过 30 余项国家重点项目和国际合作项目,并获得教育部(委)、科技部(委)、北京市奖励和其他奖励 10 余次. 在国内外知名刊物和学术会议上发表了 130 多篇学术论文,并多次应邀出国访问和做学术报告.
E-mail: fzheng@tsinghua.edu.cn

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net