

基于线性预测残差倒谱的基音周期检测

金学成^{1,2} 汪增福¹

¹(中国科学技术大学 自动化系 合肥 230027)

²(江西省电力公司 调度中心 南昌 330077)

摘 要 提出一种基于线性预测残差倒谱的基音周期检测算法. 该算法对语音信号的线性预测残差信号做倒谱变换, 将其作为基音检测特征. 并综合残差倒谱峰、短时能量和短时过零率三种特征, 构造一个清浊音判决函数, 简化清浊音判决过程, 提高判决精度. 在基音周期检测过程中, 根据基音连续原则, 提出峰值重定位方法, 有效降低基音倍频和半频的错误率. 对比实验表明, 本文算法的性能不仅较之传统的倒谱方法有明显改善, 同时也优于目前效果较好的 YIN 算法和多尺度小波算法.

关键词 基音周期检测, 清浊音判决, 线性预测 (LP), 倒谱
中图法分类号 TN912.3

A Pitch Detection Algorithm Based on Linear Prediction Residual Cepstrum

JIN Xue-Cheng^{1,2}, WANG Zeng-Fu¹

¹(*Department of Automation, University of Science and Technology of China, Hefei 230027*)

²(*Dispatching Center, Jiangxi Electric Power Corporation, Nanchang 330077*)

ABSTRACT

An algorithm based on the linear prediction (LP) residual cepstrum for pitch detection is presented. The cepstrum of linear prediction residual of the speech signal is used to be the information for pitch determination. Voicing decisions are made based on a decision function consisting of prediction residual cepstral peak, energy and zero-crossing rate of short-time segments of the speech signal. By this decision function the procedure of voicing decision is greatly simplified and the accuracy of voiced/unvoiced classification is improved significantly. Based on the consecution of pitch, a peak relocation method is introduced in the process of pitch determination to resolve the problems of pitch doubling and pitch halving. The results of the contrast experiment show that the proposed algorithm not only obtains a considerable improvement compared with the conventional cepstrum method, but also performs better than YIN estimator and multiscale wavelet method, which are effective admittedly.

Key Words Pitch Detection, Voiced/Unvoiced Classification, Linear Prediction (LP), Cepstrum

1 引言

基音周期作为一个重要参数在语音压缩编码、分析合成、语音识别和说话人识别等研究中具有重要的意义. 现有的基音周期检测方法主要可分为基于事件的和非基于事件的两大类. 非基于事件的方法主要包括:自相关法、倒谱法、简单逆滤波法、平均幅度差法等^[1-2]. 近年来众多学者在此基础上做了大量研究,提出许多改进方法,如基于改进自相关函数的方法^[3-4]、基于修正倒谱的方法^[5-6]、基于循环AMDF的方法^[7]等. 其中 A. de Cheveigné 和 H. Kawahara 于 2002 年提出的 YIN 算法^[8]是这一类方法中的佼佼者. 基于事件的基音检测方法是通过定位声门闭合时刻来估计基音周期. 目前这类方法并不多见,主要以小波方法^[9]为代表,其中,李香春和杜立民于 2003 年提出的基于多尺度空间的小波系数加权的方法^[10]是目前效果较好的方法之一. 虽然上述各种方法都已经在不同的系统中得到一定程度的应用,但是语音信号固有的特性使得准确、鲁棒的清浊音判决和基音周期检测仍然是语音信号处理研究中最困难的课题之一.

本文提出了一种基于线性预测 (Linear Prediction, LP) 残差倒谱的基音周期检测算法. 该算法对语音信号的线性预测残差信号做倒谱变换,然后根据倒谱的性质对其进行简单修正,将其作为基音检测特征,通过这种方法进一步去除声道共振峰的影响. 同时,在基音周期估计过程中,提出基于基音连续原则的峰值重定位方法,有效降低发生基音倍频和半频错误的概率.

在现有的基音检测方法中,大多采用语音信号的某一特征作为参量,根据事先确定的阈值做出清浊音判决. 这些方法对于强浊音语音段效果较好,但在清浊音过渡段和弱浊音段,其判决效果则不够理想. 虽然有学者针对这一问题提出多特征综合的清浊音判决方法^[6],但其判决规则较为复杂,过程较为繁琐. 本文突破这种规则式的判决方法,综合 LP 残差倒谱峰、短时能量和短时过零率三种特征,构造一个判决函数,并通过对其统计特性的分析,选取合理的阈值,使得仅依据该判决函数即可做出准确的清浊音判决,从而避免繁琐的判决过程,提高判决精度.

实验结果表明,本文算法的性能不仅较之传统的倒谱方法有非常明显的改善,而且也优于目前效果较好的 YIN 算法^[8]和李香春等人提出的多尺度

加权的小波算法^[10]. 同时该算法对加性噪声具有较好的鲁棒性.

2 线性预测残差倒谱

一个离散时间信号 $x(n)$ 的倒谱定义为其功率谱对数的离散逆傅立叶变换:

$$c_1(n) = IDFT\{\log | DFT\{x(n)\} |\}.$$

语音信号 $s(n)$ 可以看成是由声门激励信号 $e(n)$ 和声道冲激响应 $v(n)$ 卷积而成:

$$s(n) = e(n) \otimes v(n).$$

在计算倒谱的过程中,通过同态解卷积使声门激励信号与声道响应的卷积效应变成相应的加性效应. 从而使声道响应在倒谱中主要体现在相对较低的倒频部分,而激励信号则体现在相对较高的倒频部分.

浊音语音的激励信号可以近似看作周期冲激序列,若假设其周期为 N_p ,则可以证明其倒谱具有如下性质^[2]:

$$c_e(n) = \begin{cases} \text{非零值}, & n = N_p, 2N_p, 3N_p, \dots \\ 0, & \text{otherwise} \end{cases}$$

另一方面,由于倒谱会随着 n 值的增大而衰减,所以浊音语音的倒谱会在对应于基音周期的时间点上出现尖峰. 而清音语音因其激励信号为随机噪声,不具有周期性,其倒谱中则不会出现这样的尖峰. 通过在倒谱中检测这个尖峰,就可以做出清浊音判决并求出基音周期^[11].

然而,在实际语音信号的倒谱域中激励信息与声道信息并不是完全分离的. 在过渡音和弱浊音情况下,由于周期激励信号能量较低,声道响应共振峰的影响将突显出来,使得能够反映基音信息的倒谱峰变得不清晰甚至完全消失. 这种现象在噪声情况下将更为严重. 图 1(b) 所示为一段过渡性弱浊音语音信号 (图 1(a)) 的倒谱,在计算倒谱的过程中采用了 1 024 点的快速傅立叶变换,为了观察方便,将 2ms 以下的倒谱幅度置为零. 可以看出,在图 1(b) 中不存在明显的代表基音周期的倒谱峰. 此时,传统的基于倒谱的基音检测方法将失效. 如果能够设法去除语音信号中的声道响应信息,并对噪声加以抑制,则基于倒谱的基音检测结果将会得到改善,特别是在过渡音和弱浊音段.

语音信号 LP 分析的基本思想是语音抽样信号 $s(n)$ 和声门激励信号 $e(n)$ 之间的关系,可用以下的差分方程来表示:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Ge(n),$$

等式右边的第1项为 p 阶线性预测器的输出,若将其记为

$$\hat{s} = \sum_{k=1}^p a_k s(n-k),$$

则 LP 残差信号可表为

$$\begin{aligned} \varepsilon(n) &= s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \\ &= Ge(n). \end{aligned}$$

上式的过程又称为逆滤波,从其结果可以看出,一个 p 阶线性预测器与产生实际语音信号的系統越接近,则残差信号 $\varepsilon(n)$ 的特性也就越接近于激励信号 $e(n)$.理想情况下, $\varepsilon(n)$ 中将不包含声道响应信息,但包含完整的声门激励信息,对它进行倒谱分析,可期望获得更为清晰和精确的基音信息.

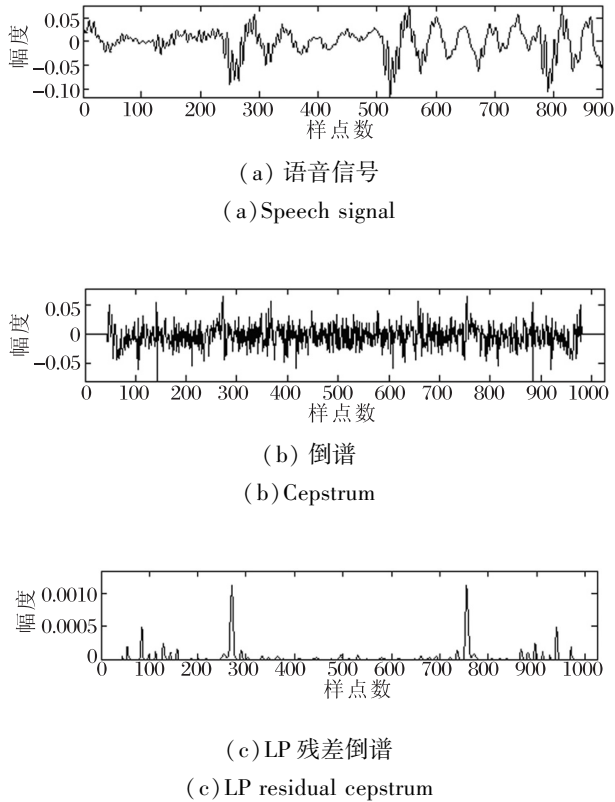


图1 一段过渡性弱浊音的倒谱和 LP 残差倒谱
Fig.1 Cepstrum and LP residual cepstrum of a transitional weak voiced speech segment

图2给出计算语音信号 LP 残差倒谱的流程.语音信号经过分帧加窗之后,首先对语音帧进行 LP 分析,得到线性预测系数,并由此构成逆滤波器对语音帧进行逆滤波,得到残差信号 $\varepsilon(n)$,然后对 $\varepsilon(n)$ 做倒谱变换.在倒谱变换的过程中,将对数谱的高频分量置零,这样做,1)能够对清音的类噪激励加以抑制;2)有利于抑制随机噪声^[5];3)能够使倒谱变得

平滑.实验发现,置零频率门限 F_b 在1 400 ~ 1 600Hz 范围内取值,本文算法可取得较好效果.

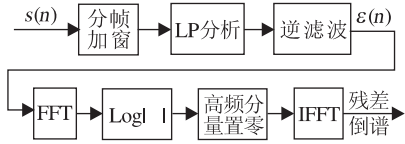


图2 计算 LP 残差倒谱的流程
Fig.2 Flowchart for calculation of LP residual cepstrum

为了使代表基音周期的倒谱峰值更为突出,根据倒谱的性质对其进行如下简单修正:

$$c_{\varepsilon}(n) = \begin{cases} c_{\varepsilon}(n)^2, & \text{if } c_{\varepsilon}(n) > 0 \\ 0, & \text{if } c_{\varepsilon}(n) \leq 0 \end{cases}$$

即只取大于零的倒谱值,并将其平方作为修正后的倒谱值,小于零的倒谱值则置为零.图1(c)所示为 $FB = 1500\text{Hz}$ 时图1(a)中语音段的 LP 残差倒谱,其中代表基音周期的倒谱峰清晰可见.

由于倒谱在高倒频部分(随着 n 的增大)会发生衰减,所以在1 ~ 15ms(对应于基频70 ~ 1 000Hz)范围内对倒谱进行线性加权提升是必要的.通过对变周期的冲激串信号进行实验,发现在这个范围内权重取为1 ~ 5效果较好.

3 清浊音判决

用于清浊音判决的语音特征可以分为两类:1)因清浊音而呈现不同值的特征,例如短时能量、短时过零率等;2)能反映出浊音周期性的特征,例如短时自相关函数、倒谱峰等.传统的基于单一特征的清浊音判决方法,由于所用特征的动态取值范围在清音段和浊音段通常会有一部分重叠,从而使得相关算法的效果在清浊音过渡段不够理想.为此,本文在进行清浊音判决时,综合考虑上述两类特征.

对于浊音而言,除了倒谱中会出现对应于基音周期的尖峰之外,还应具有较高的能量和较低的过零率,而清音则相反.于是可以构造如下的判决函数:

$$DF_n \equiv C_n + \lambda \cdot \log \left\{ \frac{E_n}{[ZCR_n + \frac{\text{mean}(E_n)}{1 \leq n \leq N}]} \right\}, \quad (1)$$

其中, C_n 、 E_n 和 ZCR_n 分别表示第 n 帧语音信号的 LP 残差倒谱峰值、短时能量与短时过零率, N 为帧数, λ 为一常数.如此构造的 DF_n 函数在清音段会具有较低的值,而在浊音段则具有较高的值.式中的

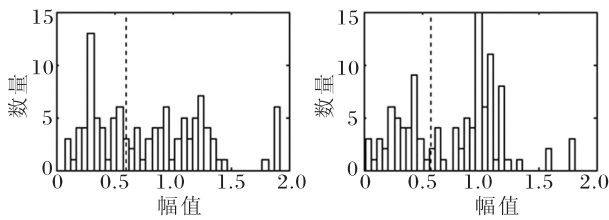
$\text{mean}(E_n)$ 算子表示短时能量的平均值,它的作用是能够避免将一些过零率很低,但不具有明显周期性的语音段判为浊音段,这在进行基音周期检测时是合理的.在实际应用中,可以把式(1)中的 C_n 和对数算子都归一化到 0 ~ 1 范围内,经过这样的处理之后,系数 l 就可以取为 1.

如果 DF_n 大于一个给定的阈值,则可以判决该语音段为浊音,否则判决其为清音.这个判决阈值的选取是否合理,直接影响着判决结果的可靠性.理想情况下,该阈值的最大后验概率估计应该是使如下误判概率最小的 θ 值:

$$P(\theta) = P(V) \int_{-\infty}^{\theta} p(x|V) dx + P(UV) \int_{\theta}^{+\infty} p(x|UV) dx,$$

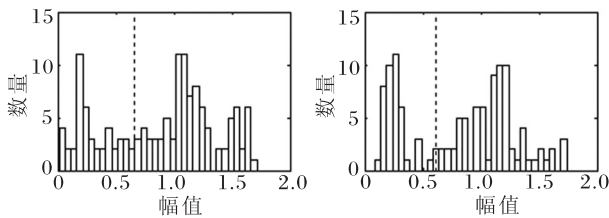
其中, $P(V)$ 和 $P(UV)$ 分别表示语音段为浊音和清音的先验概率, $p(x|V)$ 和 $p(x|UV)$ 则分别为浊音和清音情况下 DF_n 的条件概率密度函数.若这些先验概率和概率密度已知,则可以通过求极值的办法得到 θ 值.但是一般情况下这些先验概率和概率密度都是不可知的,所以无法得到一个理论上的最优阈值.可行的办法是通过实验找到一个能够使误判概率最低的阈值.通过大量的统计分析实验发现, DF_n 的统计均值的 0.7 ~ 0.8 倍是相对最优的选择.

图 3 所示为分别来自 2 男 2 女 4 位话者的 4 句话的 DF_n 值分布直方图,虚线标出了 $0.75 \text{mean}(DF_n)$ 的位置.



(a) 2 个不同的男性话者语料的 DF_n 分布

(a) Distributions for 2 different male speakers



(b) 2 个不同的女性话者语料的 DF_n 分布

(b) Distributions for 2 different female speakers

图 3 DF_n 的分布直方图

Fig. 3 Histograms of DF_n

4 基音周期估计

如果语音段被判定为浊音段,则 LP 残差倒谱峰的位置一般是基音周期的合理候选.但是倍频和半频错误是基音周期检测中较难克服的一类问题,常用的中值滤波方法虽然能起到一定的平滑作用,却不能很好地解决此类问题.对此,一些学者提出有针对性的解决方案.例如,文献[11]采取在以当前倒谱峰位置的一半为中心的 $\pm 0.5\text{ms}$ 范围内搜索超过实时阈值的峰值的方法避免半频现象.文献[12]中提出元音中心定位和基音邻近搜索算法.前者无法避免倍频现象,后者则需要进行元音中心定位,而且没有考虑自然的基音周期加倍或减半.为此,本文提出一种基于基音连续原则的峰值重定位方法,在考虑自然的基音周期加倍或减半的前提下,能够较好地克服基音倍频和半频错误.

语音信号是连续的,其基音变化也应该是连续的,即相邻帧的基音周期变化应该不会很大.若已知第 n 帧语音信号的基音周期为 T_n ,则其相邻帧的基音周期应满足

$$\begin{cases} \left| \frac{T_{n+1} - T_n}{T_n} \right| \leq D_T \\ \left| \frac{T_{n-1} - T_n}{T_n} \right| \leq D_T \end{cases}$$

由上式可得

$$\max \left\{ \frac{1}{1 + D_T}, 1 - D_T \right\} \leq \frac{T_n}{T_{n-1}} \leq \min \left\{ \frac{1}{1 - D_T}, 1 + D_T \right\}, \quad (2)$$

在实际应用中,一般取 $D_T = 0.4$ 就能取得满意的效果.

根据式(2),可以采用如下的搜索算法避免错误的半频和倍频现象.

step1 如果 $T_n > 1.4T_{n-1}$, 视为发生半频现象,在 $T_{n-1} \pm 1\text{ms}$ 范围内重新搜索 LP 残差倒谱峰值;否则转到 step4.

step2 如果新峰值大于 0.5 倍原峰值,则取新峰值位置为 T_n ; 否则视为自然的基音周期加倍.

step3 转到 step6.

step4 如果 $T_n < 0.7T_{n-1}$, 视为发生倍频现象,在 $T_{n-1} \pm 1\text{ms}$ 范围内重新搜索 LP 残差倒谱峰值;否则转到 step6.

step5 如果新峰值大于 0.8 倍原峰值,则取新峰值位置为 T_n ; 否则视为自然的基音周期减半.

step6 算法结束.

由于 LP 残差倒谱滤除了大部分声道共振峰的影响,大大降低了发生倍频错误的概率,所以在倍频检测中的系数因子取得较大,而半频检测中的系数因子则取得相对较小.

可以看出,上述搜索算法的准确性受其起始点的制约,如果起始点的基音周期估计错误,则有可能导致其后面的搜索随之错误. 为了避免这种现象的发生,在每段浊音区的开始部分,考察连续的 3 帧,做如下处理.

1) 如果
$$\frac{|T_{n-1} - T_{n-2}|}{\min(T_{n-2}, T_{n-1})} > 0.4,$$
并且
$$\frac{|T_{n-1} - T_n|}{\min(T_{n-1}, T_n)} > 0.4,$$
则
$$T_{n-1} = (T_{n-1} + T_n)/2.$$

2) 如果
$$\frac{|T_{n-1} - T_{n-2}|}{\min(T_{n-2}, T_{n-1})} > 0.4,$$
并且
$$\frac{|T_{n-1} - T_n|}{\min(T_{n-1}, T_n)} \leq 0.4,$$
则
$$T_{n-2} = (T_{n-1} + T_n)/2.$$

3) 如果
$$\frac{|T_{n-1} - T_{n-2}|}{\min(T_{n-2}, T_{n-1})} \leq 0.4,$$
并且
$$\frac{|T_{n-1} - T_n|}{\min(T_{n-1}, T_n)} > 0.4,$$
则
$$T_n = (T_{n-2} + T_{n-1})/2.$$

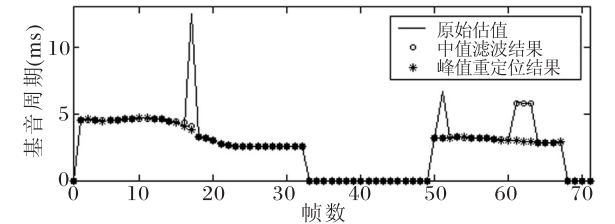


图4 峰值重定位方法与中值滤波平滑的比较

Fig.4 Comparison between peak relocation method and median smoother

图4 给出采用本文提出的峰值重定位方法与 5 点中值滤波方法对基音周期轨迹进行平滑的实验比

较结果. 从图中可以看出,基音周期估值在“帧数”为 17、51、61、62、63 处有 5 个估计错误,其中后面 4 个属于典型的半频错误. 中值滤波纠正了其中“帧数”为 17 和 51 两处孤立的错误点,而对另外 3 处连续的错误却无能为力. 而采用上述峰值重定位方法后,所有的错误都被纠正了.

5 实验结果及分析

为了确保实验语料的有效性,我们请不同年龄的 2 男 2 女以不同情感不同内容录制了 40 句话(每人 10 句),共 50 236ms,4 909 帧,其中男性语音 25 891ms(2 531 帧),女性语音 24 345ms(2 378 帧). 这些语音样本以 22.05kHz 采样率采样,16bit 量化. 窗函数选为 40ms 的 Hamming 窗,帧移为 10ms. 在求残差倒谱时,LP 分析的阶数取为 16,对数谱高频分量置零的频率门限取为 1 500Hz.

表 1 给出采用峰值重定位方法前后,在 4 909 帧中发生倍频和半频错误的帧数对比结果. 从表中数据可以看出,本文方法大大降低了发生基音倍频和半频错误的概率.

表 1 采用峰值重定位方法前后的对比结果

Table 1 Comparison between results with and without peak relocation method

方法	倍频错误 (帧数)	半频错误 (帧数)
采用峰值重定位方法之前	53	99
采用峰值重定位方法之后	2	29

为了精确评估本文算法的性能,将其与传统倒谱方法^[11]、YIN 算法^[8]以及李香春等人提出的多尺度加权的小波算法^[10](以下简称“多尺度小波算法”)进行了比较实验. 在传统倒谱方法中,清浊音判决门限取为整段语音信号倒谱峰值统计均值的 0.8 倍. 在多尺度小波算法中,采用二次样条函数作为小波函数,选取 2⁵、2⁶ 和 2⁷ 这 3 个尺度的小波系数进行加权,清浊音判决门限取为整段语音信号小波系数加权和最大值的 0.1 倍.

图 5 给出分别利用本文算法、传统倒谱方法、YIN 算法以及多尺度小波算法对一段典型的男性语音信号进行基音周期检测的结果,其中实线是采用 CoolEdit 软件手工标注的基音周期曲线. 从图中可以看出,本文算法在清浊音过渡段和弱浊音段的效果好于另外 3 种方法.

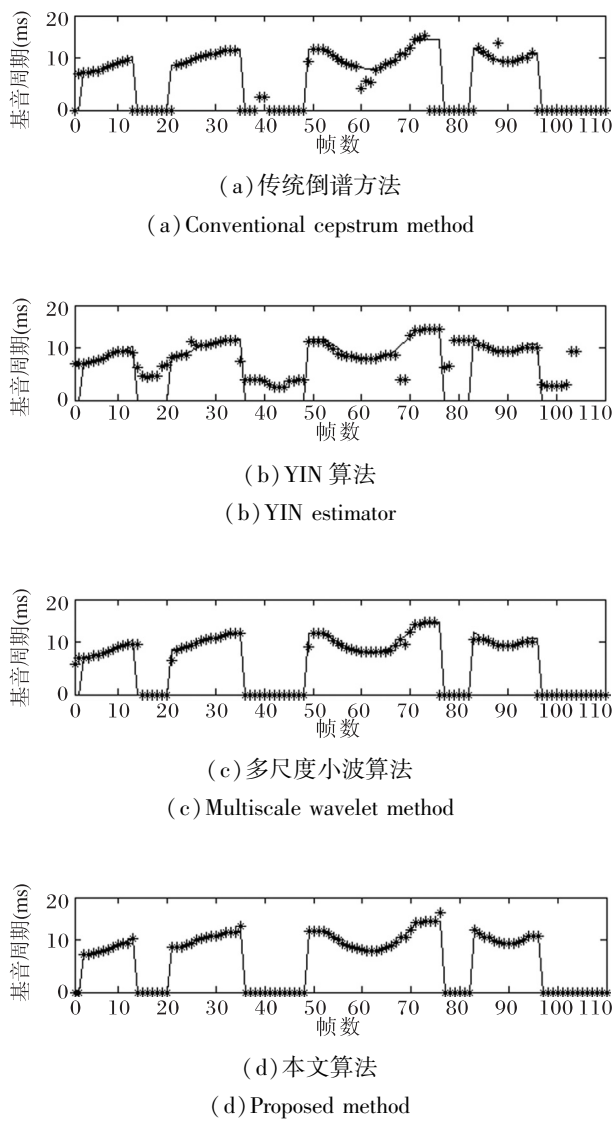


图5 4种方法的基音周期轨迹

Fig.5 Pitch tracks of 4 different methods

当基音周期估计值与标注值之间的误差超过10%时,称为发生基音毛误差(Gross Pitch Errors, GPE).表2给出了分别利用本文算法、传统倒谱方法、YIN 算法以及多尺度小波算法对4 909 帧语音信号进行基音周期估计的误差统计数据,其中 V-UV 和 UV-V 分别表示把浊音误判为清音和把清音误判为浊音的误判率.由于 YIN 算法中不进行清浊音判决,所以其相应的 V-UV 和 UV-V 项中没有对比数据.由表2 中数据可见,本文算法无论是清浊音误判率还是基音毛误差率,都远远优于传统的倒谱方法.基音毛误差率小于 YIN 算法,略小于多尺度小波算法. V-UV 略高于多尺度小波算法,但 UV-V 远远低于该算法.总体来说,本文算法略优于多尺度小波算法,较之另外两种方法具有较明显的优势.

表3 给出本文算法在不同信噪比(Signal to Noise Ratio, SNR)情况下的清浊音误判和基音毛误差的统计结果,其中噪声为加性白噪声.实验中发现,在噪声情况下,本文算法的稳定性受清浊音判决阈值的影响较大,这种影响不仅仅体现在清浊音判决的精度上,而且体现在基音周期估计准确率上,这是由于错误的初始点会对本文的基音搜索算法产生一定的影响.在低信噪比的情况下,很难找到一个普适的阈值,使得该算法对所有语料都能取得较好的效果.但是同时也发现,针对特定的语料,总能找到一个合适的阈值,使得清浊音判决和基音周期估计都能达到很高的精度.通过大量实验发现,噪声情况下 DF_n 均值的0.3~0.5 倍是相对较优的选择,因此,在实验中选取其均值的0.4 倍作为判决阈值.此时误差的分布不是很规则,但是从表3 中数据可以看出,从统计意义上讲,在信噪比高于5dB 的情况下,本文方法仍能取得较令人满意的结果.

表2 本文算法与其它3种方法的比较结果

Table 2 Results comparison among the proposed algorithm and other methods

	%		
方法	GPE	V-UV	UV-V
传统倒谱方法	3.01	6.38	2.99
YIN 算法	2.30	—	—
多尺度小波算法	1.94	1.45	2.85
本文算法	1.20	1.69	0.43

表3 本文算法在不同信噪比下的误差率

Table 3 Errors rates of the proposed algorithm at different SNR

SNR(dB)	GPE(%)	V-UV(%)	UV-V(%)
20	2.06	3.44	1.04
10	3.24	5.03	3.42
5	4.79	6.17	4.87

6 结束语

本文针对基音周期检测中存在的一些问题,提出有针对性的解决方案.将语音信号的 LP 残差倒谱用于基音周期检测,进一步去除声道共振峰的影响.并综合 LP 残差倒谱峰、短时能量和短时过零率构造一个判决函数,用于清浊音判决,取得较好效果.在基音周期估计过程中,根据基音变化连续的原则,采用峰值重定位方法有效降低基音倍频和半频的误差率.实验结果表明,本文算法在清浊音过渡段和弱浊音段能够取得较佳效果,并且在信噪比高于

5dB 的加性白噪声情况下也具有较好的表现. 该方法适用于对清浊音判决和基音周期检测精度要求较高的场合.

本文算法中清浊音判决阈值是通过大量实验分析选取的一个经验值, 这种选取方法在噪声情况下可靠性不是很高, 如何自适应地确定清浊音判决阈值将是进一步研究的重点. 另外, 本文未对更复杂的环境噪声加以讨论, 但是只要借鉴文献[5]的结论, 对本文算法加以改进, 可以期望其在其他噪声环境中也将具有良好的表现.

参 考 文 献

[1] Hess W. Pitch Determination of Speech Signals. Berlin, Germany: Springer-Verlag, 1983

[2] Yang Xingjun, Chi Huisheng. Digital Processing of Speech Signals. Beijing, China: Publishing House of Electronics Industry, 1995 (in Chinese)
(杨行峻, 迟惠生. 语音信号数字处理. 北京: 电子工业出版社, 1995)

[3] Paul B. Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound. Proc of the Institute of Phonetic Sciences of the University of Amsterdam, 1993, 17: 97-110

[4] Bao Changchun, Fan Changxin. Pitch Detection Algorithm Based on Normalized Cross-Correlation Function. Journal of China Institute of Communications, 1998, 19(10): 27-29,31 (in Chinese)
(鲍长春, 樊昌信. 基于归一化互相关函数的基音检测算法. 通

信学报, 1998, 19(10): 27-29,31)

[5] Kobayashi H, Shimamura T. A Modified Cepstrum Method for Pitch Extraction // Proc of the IEEE Asia-Pacific Conference on Circuits and Systems. Chiangmai, Thailand, 1998: 299-302

[6] Ahmadi S, Spanias A S. Cepstrum-Based Pitch Detection Using a New Statistical V/UV Classification Algorithm. IEEE Trans on Speech and Audio Processing, 1999, 7(3): 333-338

[7] Zhang Wenyao, Xu Gang, Wang Yuguo. Circular AMDF and Pitch Estimation Based on It. Acta Electronica Sinica, 2003, 31(6): 886-890 (in Chinese)
(张文耀, 许刚, 王裕国. 循环 AMDF 及其语音基音周期估计算法. 电子学报, 2003, 31(6): 866-890)

[8] de Cheveigné A, Kawahara H. YIN: A Fundamental Frequency Estimator for Speech and Music. Journal of the Acoustical Society of America, 2002, 111(4): 1917-1930

[9] Kadambe S, Boudreaux-Bartels G F. Application of the Wavelet Transform for Pitch Detection of Speech Signals. IEEE Trans on Information Theory, 1992, 38(2): 917-924

[10] Li Xiangchun, Du Limin. A Pitch Detection Algorithm Using Multiscale Edges Feature Extraction. Acta Electronica Sinica, 2003, 31(10): 1500-1502 (in Chinese)
(李香春, 杜利民. 一种基于多尺度边缘特征提取的基音检测算法. 电子学报, 2003, 31(10): 1500-1502)

[11] Noll A M. Cepstrum Pitch Determination. Journal of the Acoustical Society of America, 1967, 41(2): 293-309

[12] Gu Liang, Liu Runsheng. High-Performance Mandarin Pitch Estimation. Acta Electronica Sinica, 1999, 27(1): 8-11 (in Chinese)
(顾良, 刘润生. 高性能汉语语音基音周期估计. 电子学报, 1999, 27(1): 8-11)