循环 AMDF 及其语音基音周期估计算法

张文耀,许 刚,王裕国
(中国科学院软件研究所,北京 100080)

摘 要: 基音周期是语音压缩、合成以及识别中的一个重要参数. 传统的基于 AMDF 的基音估计算法容易导致估计的基音周期加倍. 本文针对该现象,分析了 AMDF 函数的特性及其用于基音检测时存在的不足,提出了新的 CAMDF 函数. CAMDF 有效地克服了 AMDF 函数的不足,简化了基音检测过程. 在此基础上,本文给出了新的基于 CAMDF 的基音检测算法. 该算法不仅简化了基音检测过程,而且降低了误判率,提高了估计精度. 大量实验表明其性能优于其它 AMDF 或 LVAMDF 的方法.

关键词: 平均幅度差函数; 循环平均幅度差函数; 基音周期估计
中图分类号: TN912 **文献标识码**: A **文章编号**: 0372-2112 (2003) 06-0886-05

Circular AMDF and Pitch Estimation Based on It

ZHANG Wen-yao, XU Gang, WANG Yu-guo
(Institute of Software, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: Pitch period is a key parameter in speech compression, synthesis and recognition. The well-known AMDF is often used to determine this parameter. But it is easy to make the estimated pitch doubled. With the problem, this paper analyses the features of the AMDF and its disavantages that occur in pitch detection. And a new function, Circular AMDF (CAMDF), is proposed. The CAMDF conquers the defect of the AMDF effectively, and simplifies the process of pitch detection. With the properties of CAMDF, a new pitch detection algorithm based on CAMDF is described. The algorithm not only simplifies the pitch detection, but also effciently decreases the estimation errors and improves the precision of estimated values. Lots of experiment show that the performance of the algorithm is better than other methods that are based on the AMDF or the LVAMDF.

Key words: average magnitude difference function (AMDF); circular AMDF; pitch period estimation

1 引言

基音周期是语音压缩、合成以及识别中的重要参数. 基音周期估计也因此成为大多数语音信号处理系统的重要组成部分. 基音估计 (Pitch Estimation) 也常称为基音检测 (Pitch Detection). 由于语音信号是非平稳的时变信号,只有其中的浊音部分能够看作是准周期的,所以语音信号处理中通常采取短时处理技术. 最常用的短时基音周期估计方法有:自相关法和 AMDF 基音检测算法[1~3].

与自相关法相比,AMDF 方法具有运算量小、精度高等优点,在军用语音编码中得到广泛应用. 但是在 AMDF 方法中,经常发生基音估计结果为实际基音周期的 2 倍或 1/2 的情况[1],这不仅与语音信号波形复杂有关,还与短时 AMDF 函数的特性有关. 短时 AMDF 函数随着滞后时间的增加,峰值幅度逐渐下降[3]. 这使得谷值点检测以及谷值点的清晰度检查比较困难;加倍误判、减半误判的概率比较高. 针对该问题,相关文献提出了不少改进方法[1~3,6]. 这些方法中有些需要增加算法的复杂度,有些效果不太理想.

为此,本文首先分析了传统的 AMDF 函数的特性及其造成基音估计结果加倍与减半的原因. 在此基础上建立了一个新的函数 ——循环的 AMDF 函数 (Circular AMDF, CAMDF),同时对 CAMDF 的性质进行了分析和理论证明,提出了一种新的基于 CAMDF 的基音检测算法,CAMDF 有效地克服了 AMDF 的不足,使得基音检测算法大为简化. 对比实验表明,本文基于 CAMDF 的基音检测算法明显地减少了加倍误差、清浊误判等现象,极大地提高了基音估计的精度与性能,其实验结果与传统的 AMDF 方法和 LVAMDF[6]方法相比有很大的改进.

2 传统的 AMDF 函数及其不足

传统的平均幅度差函数 (Average Magnitude Difference Function, AMDF) 是 Ross 等人于 1974 年提出的[4],其定义为:

$$D_1(k) = \frac{1}{N}\sum_n |s(n) - s(n-k)| \qquad (1)$$

其中 $s(n)$ 为离散化的语音采样序列. 当采用短时处理技术加

收稿日期:2001-04-25;修回日期:2002-04-26

方窗时[1,3],式(1)变为

$$D_2(k)=\sum_{n=0}^{N-k-1}|s_w(n+k)-s_w(n)| \tag{2}$$

其中 $s_w(n)=s(n)w(n)$, $w(n)=\begin{cases}1, & n=0\sim N-1\\ 0, & \text{其他}\end{cases}$. 式(1)中的均值系数 $1/N$ 不影响函数特性,因此在式(2)中被省略了. 本文后面的定义也做了同样的处理,但仍保留平均幅度差函数的叫法.

一般基于短时 AMDF 函数的基音周期初步估计为:

$$TP=\arg\min_{k=TP_{\min}}^{TP_{\max}}(D_2(k)) \tag{3}$$

其中 $TP_{\max}$和 $TP_{\min}$分别为预先设定的最大、最小基音周期.

从式(2)可以看出计算 $D_2(k)$ 的差值项是不同的,随着 k 的增加,求和的差值项数将逐渐减少,结果导致 $D_2(k)$ 峰值幅度随着滞后时间 k 的增加而逐渐下降,如图 1 所示.

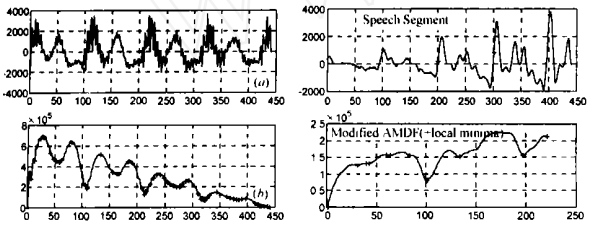

图 1　传统的 AMDF 函数示例. (a) 一段浊音波形; (b) 对应于波形(a)的 $D_2(k)$ 函数曲线

图 2　修正的幅度差函数 $D_3(k)$ 示例

函数 $D_2(k)$ 逐渐下降的趋势将使式(3)的基音估计失效. 为此,文献[1,2]中基于 AMDF 的基音周期估计算法不得不设置多个阈值,以此筛选侯选基音值并进行有效性检验. 然而,阈值与误判率之间很难达到理想的均衡,估计结果中加倍误判、减半误判、清浊误判的情况非常多. 虽然基音平滑过程能够纠正某些错误判别与估计,但是增加了处理时间,增加了算法的复杂度,还可能引入新的误差. 况且基音平滑对于一连串的误判也往往无能为力[5].

为克服平均幅度差函数的峰值幅度随着滞后时间的增加而逐渐下降的现象,文献[3]中提出将方窗 $w(n)$ 的长度加长到 $N+k_{\max}$,此时式(2)被修正为:

$$D_3(k)=\sum_{n=0}^{N-1}|s_w(n+k)-s_w(n)| \tag{4}$$

这实际上,相当于使用两个不同长度的语音帧来计算平均幅度差函数. 虽然在大多数情况下能够改变下降的趋势,但却使整个函数趋势变成向上走(如图 2 所示),而且当下一帧为清音或静音时,由于信号幅度的跃变,仍会出现下降的现象. 这些都将影响到基音周期估计的精度与算法的复杂度.

此外,文献[6]采用变长度的 AMDF(Lengthr-Varied AMDF, LVAMDF)方法实现短时基音周期估计. LVAMDF 的定义为:

$$D_4(k)=\sum_{n=0}^{k-1}|s_w(n+k)-s_w(n)|/\left(\frac{1}{2}\sum_{n=0}^{2k-2}|s_w(n)|\right) \tag{5}$$

LVAMDF 函数采用长度可变的定义且进行了归一化处理,具有诸多优点,但是也存在两个不足的地方:(1) LVAMDF 受语音帧首部的影响比较大. 如图 3(a)所示,基音周期的估计值对于语音帧的前部分比较吻合,而后面的偏差就越来越大. 这容易造成以部分帧(前一部分)的估计结果代表整个语音帧的情况,使估计精度下降. 对于以静音开始的语音帧,还会出现函数分母为零的奇异情况. (2) LVAMDF 加倍误判的情况比较严重. 如图 3(b)所示,依据式(3)估计的结果为实际值的两倍. 对清浊混合帧(或浊音的起始帧)的估计结果通常还是实际值的好几倍. 这种浊音开始部分的估值偏离,由于缺少先前的参考信息,更难以纠正. 即使是对于周期特征很强的元音, LVAMDF 也会出现多个基音周期被判为一个基音周期的现象,见图 3(b).

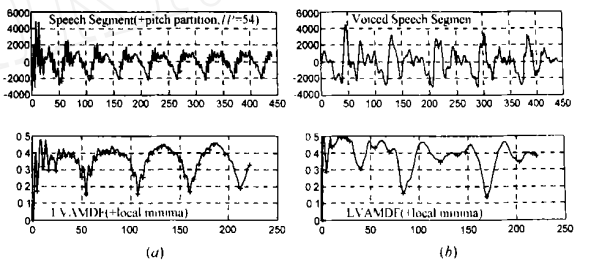

图 3　LVAMDF 函数示例

3　循环 AMDF 函数及其性能

传统的 AMDF、修正的 AMDF 以及 LVAMDF 用于基音估计时存在不足的主要原因之一是函数本身的定义. 为此,本文提出循环的 AMDF(Circular AMDF, CAMDF)函数,采取类似于循环卷积[7]的方式将式(2)重新定义为:

$$D(k)=\sum_{n=0}^{N-1}|s_w(\mathrm{mod}(n+k,N))-s_w(n)|,\quad k=0,1,\cdots,N-1 \tag{6}$$

其中 $\mathrm{mod}(n+k,N)$ 表示对 $n+k$ 进行模为 N 的求模取余操作.

从该定义出发,可以推导 CAMDF 函数的如下性质:

性质 1　$D(0)=0$.

性质 2　在定义域内, $D(k)$ 关于 $k=N/2$ 对称,即 $D(k)=D(N-k)$.

性质 1 和性质 2 从式(6)的定义可以直接推导出来,而且性质 2 使得我们在实际应用中只需要计算 $k\in[0,N/2]$ 内的函数值.

另外,对于最小周期为 T 的严格周期信号, CAMDF 函数还具备以下性质(推导与证明过程略):

性质 3　$D(aT)<D(aT+b)$,其中 $0\le aT+b\le\lfloor\frac{N}{2}\rfloor$, $0<b<T$, $a=0,1,2,\cdots$.

性质 4　$k=aT$ 是 $D(k)$ 的局部最小点,其中 $0\le aT\le\lfloor\frac{N}{2}\rfloor$, $a=0,1,2,\cdots$.

性质 5　$D(aT)\le D(aT+T)$,其中 $0\le aT<aT+T\le\lfloor\frac{N}{2}\rfloor$, $a=0,1,2,\cdots$.

而对于象浊音那样的短时平稳准周期信号,统计实验表明,在绝大多数情况下,以上性质仍然成立. 这些性质在图形

(如图 4(a)) 上表现为:CAMDF 函数在整数倍基音周期的位置上呈现显著的谷值点特征,而且这些谷值点的函数值依次递增.这说明 CAMDF 函数可以象 AMDF 那样用于基音周期检测.

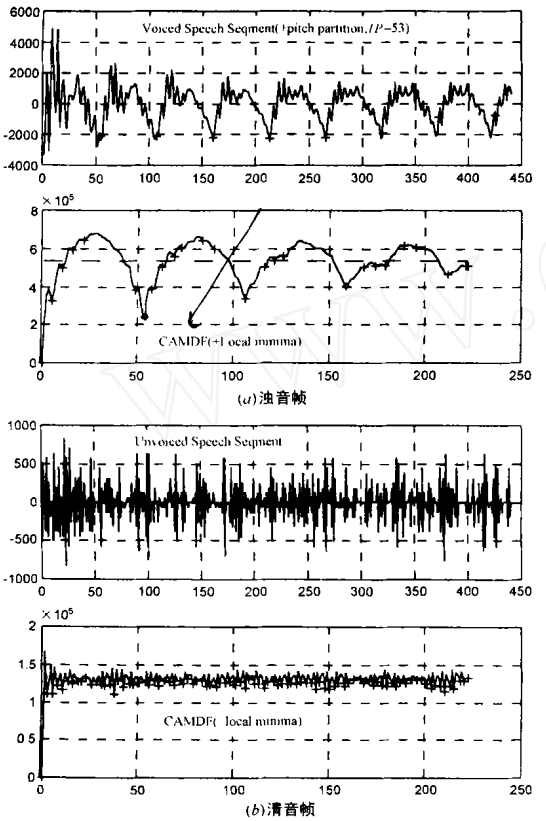

图 4 典型的 CAMDF 函数示例

然而,与 AMDF 函数不同的是,在计算式(6)所定义的幅度差函数 $D(k)$ 时,当前加窗语音帧内的每个样本点都被使用且仅被使用一次,求和的差值项数也相同.其直接结果是克服了原来的 AMDF 函数不同的 k 值之间因求和项数不同而造成的函数峰值幅度逐渐下降的趋势.如图 4 所示,新的 CAMDF 函数围绕均值线(图 4 中的短横线)水平波动,而且峰值幅度基本保持不变.

CAMDF 函数的这种特性给基音检测带来了极大的便利.(1)使谷值点检测更容易.因为水平的波动趋势,更容易确立一致的谷值点显著性检验标准;(2)简化了基音周期的检测过程.因为基音周期位置上谷值点的函数值依次递增,利用式(3)的方法可以一次定位到估计的基音周期位置,而不需要复杂的基音检测逻辑[1,2,4],与此同时,还可避免因不适当的基音判决逻辑而造成基音周期误判的情况;(3)增加了基音估计的精度.因为每次计算 $D(k)$ 时使用的样本点都一致,使得幅度差函数更能反应不同 k 值之间的差别.对比图 3(a)和图 4(a),可以看到利用 CAMDF 的估计结果进行基音划分与实际情况的吻合程度更高.

此外,CAMDF 函数对清音帧与浊音帧有着明显的区别特征(如图 4).对于类似白噪声的清音,CAMDF 也呈现类噪声特性.信号的噪声特性越强,该函数的波动就越剧烈,局部最小点的分布就越密集,函数值的变化范围也越来越小,这为我们提供了新的类似于过零率的清浊判决依据.

CAMDF 函数还具有较好的抗噪性能(见图 5),对谐波和共振峰的影响也不十分敏感.在清浊混合的情况下(见图 6),CAMDF 函数也能够估计出浊音部分的基音周期.

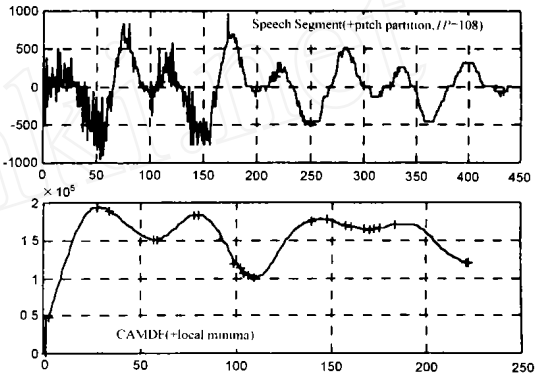

图 5 CAMDF 函数抗噪示例

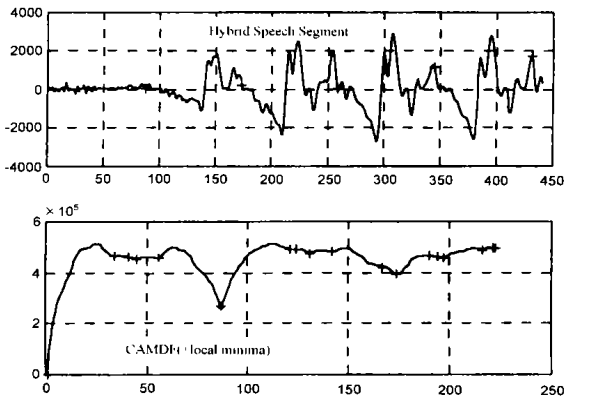

图 6 清浊混合帧的 CAMDF 函数示例

4 基于 CAMDF 的基音检测算法

上一节的分析表明 CAMDF 函数有利于简化基音检测算法,改善基音估计的性能,降低误判率.本节在此基础上给出如下新的基于 CAMDF 的基音检测算法.由于通常的基音检测算法都将基音检测和清浊判别结合在一起,因此本文也不例外.

令加窗的语音信号帧为 $s_w(n)$, $n=0,1,\cdots,N-1$, N 为帧长;最小基音周期为 $TP_{\min}$,最大基音周期为 $TP_{\max}$;估计的基音周期为 TP,若 $TP=0$ 表示该帧为清音或静音.具体的算法步骤如下:

(1)依据式(6)计算循环的平均幅度差函数 $D(k)$.

(2)查找并统计 $D(k)$ 的局部最小点(即谷值点).令总的局部最小点个数为 M(不包括端点在内).

(3)如果 $M\leq 1$,则原始语音帧为静音,将该帧判定为清音,令 $TP=0$,结束.

(4)如果 $M=2$,则令函数值较小的局部最小点的位置为 P_1,另一个最小点的位置为 P_2.

如果 P_1、P_2 都小于 $TP_{\min}$,则令 $TP=0$,结束;否则

如果 $P_1 > TP_{\min}$,则令 $TP = P_1$,否则 $TP = P_2$,结束.

(5) 利用式(3)的方法估计基音周期,即令 $TP = \arg_k MIN(D(k))$, $k = TP_{\min}, \cdots, TP_{\max}$.

(6) 进行清浊判别. 如果 $ValleyJudge(TP) = 0$,则将该帧判定为清音,令 $TP = 0$;否则将该帧判定为浊音,TP 即为估计的基音周期.

(7) 算法结束.

该算法除了使用最小基音周期 $TP_{\min}$ 作为估计结果的下限外,还在估计的基音周期位置进行了谷值特征检验. 如果谷值特征不明显,则判定为清音,将估计结果修正为 0.

谷值特征检验函数 $ValleyJudge(TP)$ 判断 TP 是否为一显著谷值点. 判别方法是计算 $D(k)$ 的全局平均值 D_{avg},令 $VP_d = D(TP)/D_{avg}$;再从点 TP 开始计算左右两边函数值低于 D_{avg} 的样本数,令其结果为 VP_w. 以 VP_w 和 VP_d 作为谷值点特征的衡量指标. VP_d 越小,VP_w 越大,TP 点的谷值特征就越明显. 为此,设置三个谷值判别阈值 TH_{d1}、TH_{d2} 和 TH_w. 当谷值点足够深(即 $VP_d < TH_{d1}$)或者是有相当的深度和宽度(即 $VP_d < TH_{d2}$ 且 $VP_w > TH_w$)时,判定 TP 为一显著有效的谷值点,$ValleyJudge(TP)$ 返回 1;否则认为当前帧为清音帧,返回函数值 0.

5 实验结果

为了检验 CAMDF 及其基音估计算法的性能,进行了大量的对比实验. 本文在此给出两组代表性的对比实验结果. 实验数据的采样率为 11.025kHz,最小基音周期设定为 2.5ms. 实验一为"祝你好运"的男声发音,结果如图 7 所示. 实验二为"Nice to meet you"的女声发音,结果如图 8 所示. 从这两组实验结果可以看到,CAMDF 估计的基音轮廓要好于 AMDF 和 LVAMDF. 在 AMDF 和 LVAMDF 中都出现加倍误判的情况,其中又以 LVAMDF 更为突出. 此外,还可以看到 CAMDF 的清浊误判率也低于其它两者.

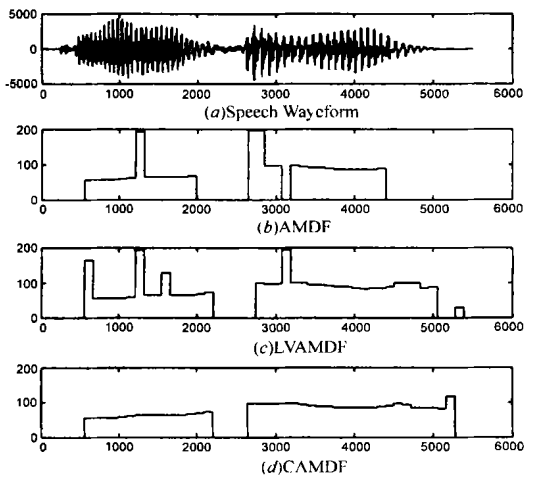

图 7　对比实验结果一

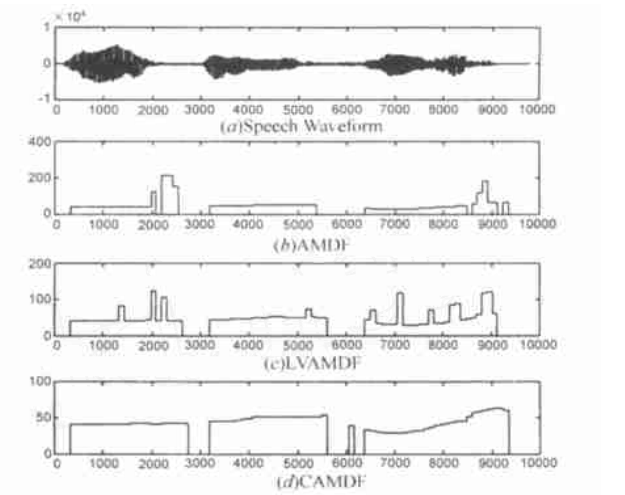

图 8　对比实验结果二

6 分析与讨论

6.1 帧长的选择

由于 $D(k)$ 的对称性(性质 2),为了估计到可能的最大基音周期 $TP_{\max}$,要求语音帧长 $N > 2\,TP_{\max}$,即语音帧长至少包含两个以上的基音周期. LVAMDF 与此相同. AMDF 的帧长则大于一个基音周期就可以,但通常为了取得较好的估计结果也使其包含两个以上的基音周期. 因此 CAMDF 的帧长限制并不特别. 在本文的实验中,采用了固定帧长的方案,最大基音周期设定为 20ms,帧长 40ms,帧移 10ms.

6.2 清浊判决与阈值的选择

基音估计和清浊判决紧密地偶合在一起,两者都是难以处理的问题. 本文的检测算法中以 CAMDF 函数的全局均值作为参考,选择了两条清浊判别标准(三个阈值):谷值点深度 VP_d(TH_{d1}和 TH_{d2})和谷值点宽度 VP_w(TH_w). 这是因为 CAMDF 函数围绕着水平均值线上下波动(见图 4). 一般浊音的 CAMDF 波动比较缓慢,周期性变化的特征比较明显;而清音的波动比较剧烈,类似于随机噪声. 当 CAMDF 函数波动缓慢时,谷值点宽度 VP_w 将比较大,反之则比较小. 谷值点宽度可以看作类似过零率的度量标准. 谷值点深度为一相对量,一般浊音的谷值点深度要小于清音. 然而,由于清音的随机特性,清音的谷值点深度值有时也会比较小,但是此时谷值点宽度也将很小. 本文运用这两条标准取得了非常好的清浊判别效果. 但是由于语音信号的复杂性,至今也没有那种方法能够将清音与浊音准确无误的区别开来. 本文的方法也不能完全避免清浊误判. 在实际应用中,增加判别逻辑或者是借助语音信号的其它特征,如能量、过零率等,可以作进一步的校正.

在语音信号中除了准周期的浊音和类噪声的清音外,还存在大量的持续时间很短的暂态信号和过渡信号,以及尾音部分的一些残缺信号. 当采用二元判决策略时,将这些信号归入浊音还是清音依赖于阈值的选择,具体的情况视应用场合而定. 实验分析表明,TH_{d1} 取 0.5～0.7,TH_{d2} 取 0.7～0.9 比较合适. TH_w 与采样率有关. 这是因为对于不同的采样率,信号包含的频率成分不一样,CAMDF 函数的波动也有变化. 当采

样率为 11.025KHz 时，TH_w 取为 5～10 比较可行.

6.3 误判率

由于 CAMDF 克服了原有 AMDF 的不足，本文算法不易造成估计结果的加倍或减半，因而显著地降低了基音估计中的误判率. 但是在少量浊音的尾音部分，特别是/ r/ 音尾部依然可能发生加倍或减半估计的情况. 原因在于尾音部分能量快速衰减，出现信号缺损现象，准周期特性被严重破坏. 此外，语音信号中有时会出现人眼也难以分辨其基音的情况，此时也可能产生误判. 这些误判现象都是由语音信号复杂多变的特性决定的. 本文基于 CAMDF 的算法只是极大地减少了误判，改善了性能，但不能完全避免. 再加上少量不可避免的清浊误判，本文算法的误判率可以控制在比较低的水平.

6.4 算法的复杂度

与文献[1,2,4]中基于 AMDF 的基音检测算法相比，本文算法的基音检测逻辑更为简单明了. 因为基于 CAMDF 的检测过程，不需要利用多重判别条件进行侯选基音位置的筛选. 这也正是 CAMDF 的优越之处. CAMDF 的运算量约为 AMDF 的两倍. 为了减少数据访问次数，基音估计算法中的幅度差函数及其均值的计算、局部最小点的检测与统计、基音位置的查找都可以在一个数据遍历中完成. 此外，采取文献[2]中介绍的一些技术，如计算部分幅度差函数，或者是借助其它参数进行清浊判别从而只计算浊音帧的幅度差函数等，都可以进一步提高算法的效率，但此时算法的复杂度也将有所提高.

7 结论

传统的基于 CAMDF 函数的基音检测算法中容易出现基音估计加倍或减半的现象. 本文针对该现象，分析了 AMDF 函数的特性及其用于基音检测时存在的不足，提出了 CAMDF 函数. CAMDF 函数具有许多优越性能，有效地克服了传统 AMDF 函数的不足，简化了基音检测过程. 在此基础上，本文提出了新的基于 CAMDF 的基音检测算法，并在该算法中采用了新的清浊判别准则. 对比实验表明，基于 CAMDF 的基音检测算法明显地降低了误判率，提高了基音估计精度，而且算法的复杂度低，实验结果比 AMDF 和 LVAMDF 都要好.

本文算法虽然优于传统的 AMDF 和 LVAMDF，但是仍然存在一些误判现象. 如果能够借助其它方法或标准区分语音信号中的清音、浊音、暂态信号、过渡信号等，那么基于 CAMDF 的基音检测性能将进一步提高. 此外，本文只考察了 CAMDF 函数针对原始语音信号的情况，对于残差信号以及中心削波等非线性处理后的情况，有待进一步研究.

参考文献：

[1] 杨行逵，迟惠生，等. 语音信号数字处理 [M]. 北京：电子工业出版社，1995.

[2] Wolfgang Hess. Pitch Determination of Speech Signals [M]. New York: Springer-Verlag, 1983.

[3] 姚天任. 数字语音处理 [M]. 武汉：华东理工大学出版社，1992.

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

[5] Thomas W Parsons. Voice and Speech Processing [M]. New York: McGraw-Hill, 1986.

[6] 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.

[7] A V 奥本海姆，R W 谢弗. 离散时间信号处理 [M]. 北京：科学出版社，1998.

作者简介：

张文耀 男，1974 年 11 月出生于江西萍乡，博士研究生，主要研究领域为语音信号处理，模式识别.

许 刚 男，1963 年 10 月出于安徽合肥，博士后，高工，主要研究领域为图像分析，多媒体通讯.