

# 一种高性能汉语语音基音周期估计方法

李祖鹏, 姚佩阳

(空军工程大学 电讯工程学院, 陕西 西安 710077)

**摘要:**在分析 AMDF 基音周期估计方法的基础上,提出一种新的汉语语音基音周期估计方法 MAMDF(Modified AMDF)。实验证明,MAMDF 方法可以获得更好的基音估计精度和时域分辨率。

**关键词:**语音基音周期估计;短时平均幅度差函数;时域分辨率

**中图分类号:**TP18 **文献标识码:**A **文章编号:**1009-3516(2000)04-0045-03

基音周期是语音信号的一个重要参数,在语音产生的数字模型中它还是激励源的重要参数。因此,在语音分析、语音合成、语音识别中,估计基音周期都是一个重要任务<sup>[1]</sup>。基音周期估计一般由两部分组成,即短时基音周期估计和基音周期轨迹平滑。常用的短时基音周期估计方法有线性预测分析法<sup>[2]</sup>、AMDF(短时平均幅度差函数)法<sup>[3]</sup>和 Seneff 法<sup>[4]</sup>等。其中,在 AMDF 方法中,短时平均幅度差函数定义为

$$D^{AMDF}(\tau) = \frac{1}{L} \sum_{j=1}^L |S_j - S_{j+\tau}|, \quad \tau = \tau_{mix}, \tau_{min+1}, \dots, \tau_{max} \quad (1)$$

其中, $S_j$  是语音采样序列 $(S_j) = (S_1, S_2, \dots, S_N)$ , $N$  为样本点总数; $\tau_{min}$  和  $\tau_{max}$  为基音周期范围的最小值和最大值, $L$  为计算范围,若语音采样率为 10kHz,则取  $\tau_{min} = 20, \tau_{max} = 200, L = 300$ ,即基音周期为 2~20ms,样本计算范围为 30ms。

基音周期

$$T_p^{AMDF} = \arg \min_{\tau} [D^{AMDF}(\tau)] \quad (2)$$

实验表明,当语音信号的波形比较平缓时,采用 AMDF 进行基音周期估计可以得到较高的估计精度和可靠性,所需的计算量相对自相关函数法而言也较小。但当语音信号的幅度和频率快速变化时,AMDF 方法的估计精度会明显下降,从而直接影响到语音识别的准确率。这主要是由 AMDF 方法数据长度选择上的不灵活性以及计算过程中存在的是误差累积问题所造成(这些问题我们将在后文具体讨论)。

鉴于 AMDF 法的以上不足,本文在其基础上提出了一种改进短时基音周期估计方法(MAMDF 法),并结合具体实验对 MAMDF 法和传统 AMDF 法进行了比较,最后得出结论。

## 1 MAMDF(Modified AMDF)方法的描述

在 MAMDF 方法中,为了提高基音周期估计的灵活性和准确性,重新定义短时平均幅度差函数为:

$$D^{MAMDF}(\tau) = \sum_{j=1}^{\tau} |S_j - S_{j+\tau}| / \sum_{j=1}^{\tau} S_j^2 \quad \tau = \tau_{min}, \tau_{min+1}, \dots, \tau_{max} \quad (3)$$

基音周期

$$T_p^{MAMDF} = \arg \min_{\tau} [D^{MAMDF}(\tau)] \quad (4)$$

由上述定义可见,在 MAMDF 方法中,样本幅度差的计算范围是随延时变化的,且延时越长,计算范围越宽。式(3)的分子项保留了 AMDF 法的特点,即保证了语音信号的波形比较平缓时,可以得到较高估计精

度;同时,为了排除窗宽的影响,式(3)引入了分母项  $\sum_{j=1}^{\tau} S_j^2$ , 其中  $S_j^2$  代表能量,该分母项的作用就相当于能量归一化,目的是保证  $D^{\text{MAMDF}}(\tau)$  在基音周期处的准确性和连续性。在实际应用中,根据不同发音人的发音周期不同,可以适当调整  $\tau$  的选择范围,如对男声可选取 4~20ms,对女声可选取 2~10ms。

与之不同,在传统的 AMDF 方法中,样本幅度差的计算范围是固定的(一般为 10~30ms),这是造成 AMDF 方法在语音基音周期估计中,特别是对元音过渡段的基音周期估计中的估计精度和时域分辨率不高的根源(这一点在后文将详细讨论)。

## 2 MAMDF 方法与 AMDF 方法的比较

针对 MAMDF 和 AMDF 两种算法,我们抽取了同一段语音信号,并分别用两种方法对它进行了基音周期估计,实验结果如图 1。下面结合实验结果,将两种方法的关系比较如下:

(1)当语音信号波形比较平缓,语音周期变化较慢时,MAMDF 方法和 AMDF 方法是等效的。若分析信号满足  $S_j = S_{j+T_p}$ , 其中  $T_p$  为基音周期。根据式(1)、(3),必有

$$D^{\text{AMDF}}(T_p) = \frac{1}{L} \sum_{j=1}^L |S_j - S_{j+T_p}| = 0$$

$$D^{\text{MAMDF}}(T_p) = \sum_{j=1}^{T_p} |S_j - S_{j+T_p}| / \sum_{j=1}^{T_p} S_j^2 = 0$$

又由式(2)、(4)得到

$$T_p^{\text{AMDF}} = \arg \min_{\tau} [D^{\text{AMDF}}(\tau)] = l \cdot T_p, \quad l \in N$$

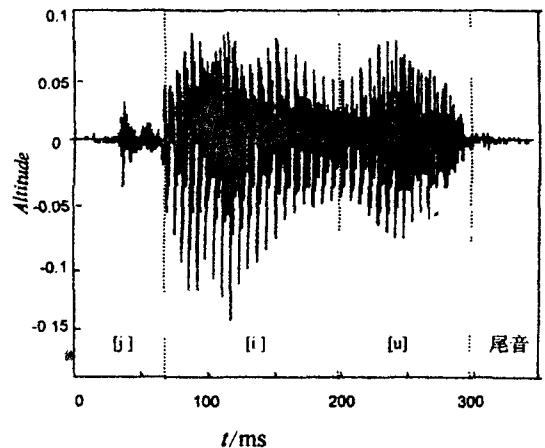
$$T_p^{\text{MAMDF}} = \arg \min_{\tau} [D^{\text{MAMDF}}(\tau)] = l \cdot T_p, \quad l \in N$$

当  $l=1$  时,  $T_p^{\text{AMDF}} = T_p^{\text{MAMDF}} = T_p$ , 显然,利用 AMDF 方法和 MAMDF 方法求得的基音周期都是正确的。

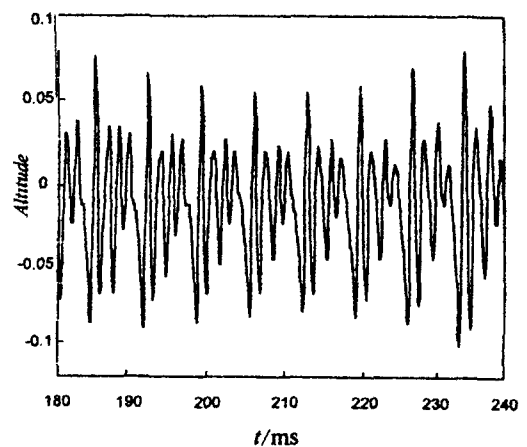
(2)针对语音中的元音过渡段,MAMDF 方法具有更好的时域特性。

AMDF 方法的数据长度是固定的,因此导致了 AMDF 方法在数据长度选择上的矛盾。一方面,数据长度越短,相应的 AMDF 时域特性越灵敏,估计出的基音周期也就越准确。这一点在汉语多音节语音识别中尤为重要,因为在语音由一个元音向另一个元音的过渡过程中,语音基音周期固定不变的时间是很短的,一般只能保持 20~30ms。另一方面,不同人的基音周期又是不同的,一般可以为 2~20ms,为保证能得到正确的基音周期,数据长度至少应大于一个基音周期的长度。因此,AMDF 方法在数据长度选择上很难兼顾。

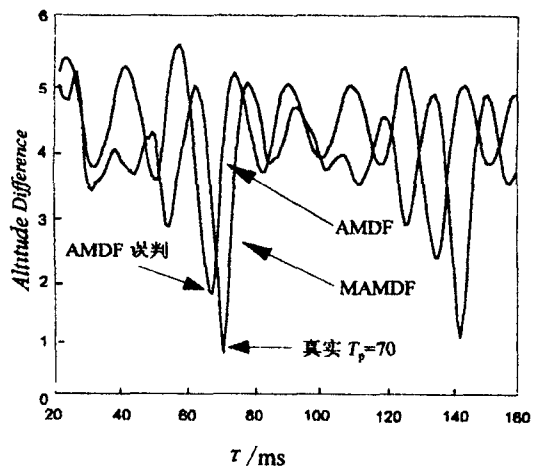
MAMDF 方法可以克服 AMDF 方法的上述问题。MAMDF 方法的有效数据长度一般集中在 1~2 个基音周期长度附近,并可以随不同人的基音频率变化而变化。因此 MAMDF 方法既具有良好的时域特性,又不会因数据过短而导致算法失败。从图 1c 的基



(a)汉语数码“9”的语音波形



(b)汉语数码“9”语音中的[i-u]



(c)AMDF 和 MAMDF 对应于图(b)语音段的短时平均幅度差

图 1 AMDF 与 MAMDF 实验结果比较

音周期估计结果可以看出, MAMDF 方法具有明显优于 AMDF 方法的谷值特性, 从而具有更好的精度和可靠性。

(3) 针对汉语的升调(二声)和降调(四声)语音段, MAMDF 方法优于 AMDF 方法。因为 MAMDF 方法克服了 AMDF 方法的误差累积效应。

在前面的分析中我们指出, 当基音周期和信号幅度不变时, AMDF 方法和 MAMDF 方法是等效的。但实际情况是, 语音信号的基音周期和信号幅度在绝大多数时间内是变化的。具体来说, 在汉语升调语音中, 基音周期一般是单调递减的; 在汉语降调语音中, 基音周期一般是单调递增的。与之类似, 在汉语语音起始段, 信号幅度一般是单调上升的; 在汉语语音结尾段, 信号幅度一般是单调下降的。当估计这些语音段的基音周期时, 在 AMDF 方法中, 基音周期对应的短时幅度差函数涉及十几个甚至几十个周期, 因此存在误差累积问题。这些累积误差会造成基音周期估计精度的下降甚至错误。在 MAMDF 方法中, 由式(3)可见分子项最大计算范围是  $2\tau_{\max}$ , 分母项最大计算范围是  $\tau_{\max}$ , 故基音周期对应的短时幅度差函数只涉及两个周期, 不存在误差累积问题, 因此基音周期估计的准确性是明显优于 AMDF 方法的。

在图 1 的实验结果比较中, MAMDF 法有效地纠正了 AMDF 法对元音过渡段 [i-u] 基音周期的误判, 并得到了更好的时域分辨率和更高的精度。综上所述, 采用 MAMDF 方法替代 AMDF 方法优点有: (1) 更好的时域特性; (2) 更适合于元音过渡段的基音分析; (3) 大大减少估计误差, 更适合进行汉语变调元音的基音分析。

### 3 结论

在数字语音处理中, 语音基音周期估计一直受到高度重视。语音基音分析的精确度、连续性会在很大程度上影响语音分析的效果。本文通过以 MAMDF 方法取代 AMDF 方法实现基音周期估计。实验表明, MAMDF 方法在时域分辨率和估计精度上明显优于 AMDF 方法。在实际应用中, 将其应用于汉语数码孤立语音声调判别中后, 有效地提高了声调识别率。

#### 参考文献:

- [1] 姚天任. 数字语音处理[M]. 武汉: 华中理工大学出版社, 1992.
- [2] 胡光锐. 语音处理与识别[M]. 上海: 上海科学技术文献出版社, 1994.
- [3] 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.
- [4] Seneff S. Real-time harmonic pitch detector[J]. IEEE trans. on Acoustics, Speech and Signal Processing, 1978, 26(4):358 - 365.

## A Method for Estimating High-Performance Mandarin Pitch

LI Zu-peng, YAO Pei-yang

(The Telecommunication Engineering Institute, AFEU., Xi'an 710077, China)

**Abstract:** A new Mandarin pitch contour estimation method MAMDF (Modified AMDF) is proposed on the basis of the AMDF method. According to the results of experiments, MAMDF method has gained a higher accuracy and better time resolution.

**Key words:** pitch estimation; average magnitude difference function; time resolution