

改进 LVAMDF 及综合多因素基音检测算法

薛帅强, 陈波, 陈菲

(西南科技大学 计算机科学与技术学院, 四川 绵阳 621010)

摘要: 在对语音信号静音、清音、浊音划分的基础上, 针对语音信号周期特征明显段分布随机性问题, 提出改进的变长度平均幅度差函数 LVAMDF 及综合多因素基音检测算法, 该算法对语音信号进行周期特征明显段和周期特征不明显段的聚类划分, 同时, 获取周期特征明显语音段的所有基音周期的起止端点, 针对少数基音周期划分倍频或半频问题, 提出识别、修正方法, 其识别、修正率极高; 在对大量真实语音处理中, 能够精确地检测出语音特征明显段的基音周期端点, 基本没有倍频和半频划分, 并且和 AMDF、ACF 算法作了对比。

关键词: 改进的变长度平均幅度差函数; 综合多因素; 精确基音周期; 修正倍频或半频

Improved Algorithm of Length Varied Average Magnitude Difference Function (LVAMDF) and Comprehensive Multi-factor for Pitch Frequency Detection

Xue Shuaiqiang, Chen Bo, Chen Fei

(School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China)

Abstract: On the basis of the methodology that categorizes speech signal into three types, silence, voiceless sound, voiced sound, in view of the random distribution of obvious periodic property speech, the improved algorithm of length varied average magnitude difference function (LVAMDF) and comprehensive multi-factor for pitch frequency detection is put forward to categorize voiced sound into two types, one is obvious periodic property speech, the other one is unobvious periodic property speech. At the same time, the starting and ending points of all accurate pitch period in the obvious periodic property speech is achieved. For a few pitch periods divided into frequency doubling or half frequency, the recognition and correction method is proposed which has a high recognition and correction rate. Finally, in a large amount of the real speech processing experiments, the method can detect the pitch period in the obvious periodic property speech accurately, and there is hardly frequency doubling and half frequency. The results show that the proposed algorithm performs much better on pitch detection compared with AMDF and ACF.

Keywords: improved algorithm of length varied average magnitude difference function (LVAMDF); comprehensive multi-factor; accurate pitch period; identify and correct frequency doubling and half frequency

0 引言

语音基音周期是人类在发语音过程中, 声带振动的最小周期。语音基音周期检测的准确性, 对语音识别、语音压缩、语音编解码、语音合成、说话人识别与语种辨识等等有至关重要的作用^[1]。语音基音周期检测方法有很多种, 平均幅度差函数法 (average magnitude difference function, AMDF)^[2-4]、变长度短时平均幅度差函数法 (length varied average magnitude difference function, LVAMDF)^[5-7]、自相关函数法 (autocorrelation function, ACF)^[8]、希尔伯特—黄变换方法 (hilbert-huang transform, HHT)^[9]、小波变换法^[10] 等等, 其中, LVAMDF 是运算最简单的方法, 在实时语音处理过程中, 底层运算应尽可能简单、稳定, 由于传统的 LVAMDF 没有对处

理对象做选择, 所以在处理不适应此方法的语音段时, 会划分错误, 而且在处理适应此方法的语音段时, 有较多的倍频和半频划分。

本文提出了提取适用 LVAMDF 方法划分基音周期语音段落的方法, 改进了 LVAMDF 方法, 并且综合周期长度、端点振幅和最优相似度选取因素精确估计基音周期端点, 最后通过倍频和半频识别、修复和排除方法, 大大减小了倍频和半频周期划分。

1 基音周期检测原理

基音周期检测原理如图 1 所示。

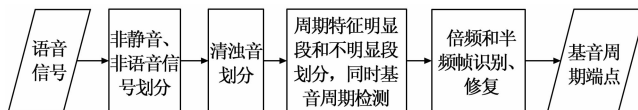


图 1 基音周期检测原理图

1.1 预处理

在接收到的语音信号中包含静音、清音、噪音等复杂成分, 这些成分在基音周期检测中, 是冗余部分, 会使基音周期检测错误和检测速度下降, 所以在基音周期检测之前要将这些成分划分开。

收稿日期: 2015-11-02; 修回日期: 2015-11-26。

基金项目: 四川省教育厅重点项目 (ZD1109); 绵阳市网络融合工程实验室开放基金项目 (12ZXWK11)。

作者简介: 薛帅强 (1989-), 男, 河南省原阳县人, 硕士研究生, 主要从事语音信号处理方向的研究。

陈波 (1963-), 男, 博士, 教授, 主要从事图像处理、嵌入式技术方向的研究。

排除静音段方法：短时信号围绕面积法，根据非静音段的短时信号围绕面积门限找到非静音段的端点。

排除非语音信号方法：语音信号频率在 50~450 Hz，18 ms 为 55.6 Hz，当 18 ms 中都没有 2 个以上正数最大极大值（两个负数极小值之间所有正数极大值点中的最大值点）的周期端点，那么其频率一定低于 55.6 Hz，即为非语音段。

排除有周期性的噪声方法：根据人类语音浊音频率范围是 50 Hz 到 450 Hz，在基音周期端点检测后，根据频率大小可判断该非静音语音是否为语音信号。本文使用有效翻转率和频谱振幅均值方法来划分清浊语音信号。

定义 1（短时信号围绕面积）：对一帧信号作一次拟合，拟合后的曲线和横轴围绕的平均面积就是短时信号围绕面积。它是短时平均幅度函数^[11]的变化。

定义 2（有效翻转率）：一个拐点和其两侧拐点的振幅差值大小比例都超过一个有效翻转门限（本文用的是 0.25）的时候，称为一次有效翻转。一帧语音信号中有效翻转的个数就是有效翻转率。它是短时过零率算法^[12]的变化。

1.2 检测基音周期

首先将浊音信号划分为周期特征明显段和周期特征不明显段，然后对周期特征明显段进行基音周期检测，最后将倍、半频划分进行检测、修复和排除。

周期特征明显段和周期特征不明显段划分方法：如果某帧的第一周期和其整帧的周期预划分成功，则该帧确定为周期特征明显段，否则为周期特征不明显段。

倍频和半频划分检测、修复和排除方法：1) 对周期特征明显段语音帧按周期长度分类，对每类中任意一帧，寻找其最优第一周期，如果最优第一周期比历史划分中的第一周期相差超过倍频或者半频门限，就对该种类的语音帧进行周期端点修复，修复时，用最优第一周期确定 LVAMDF 算法中周期长度变量的变化范围，确保再次划分的正确性；2) 由于倍频和半频划分非常少，所占的比例非常少，所以本文将语音段按照周期长度分类，根据分类的统计结果，将所占比例非常少的部分视为倍频或半频，并将其排除。

1.3 LVAMDF 算法

传统的 AMDF 算法^[5]为公式：

$$D_A(\tau) = \frac{1}{L} \sum_{j=1}^L |s_j - s_{j+\tau}|, \tau = \tau_{\min} \sim \tau_{\max} \quad (1)$$

式 (1) 中，语音采样序列 $(s_j) = (s_1, s_2, \dots, s_N)$ ， N 为样本点总数， τ_{\min} 、 τ_{\max} 确定了基音周期范围，一般为 2~20 ms， L 为样本幅度差的计算范围。

AMDF 算法的基音周期定义为公式 (2)：

$$T_A = \operatorname{argmin}_{\tau} \{D_A(\tau), \tau_{\min} < \tau < \tau_{\max}\} \quad (2)$$

归一化的 LVAMDF 算法^[6]，其定义为公式 (3)：

$$D_L(\tau) = \frac{\frac{1}{\tau} \sum_{j=1}^{\tau} |s_j - s_{j+\tau}|}{\frac{1}{2} \sum_{j=1}^{2\tau} |s_j|}, \tau = \tau_{\min} \sim \tau_{\max} \quad (3)$$

归一化的 LVAMDF 算法的基音周期定义为公式 (4)：

$$T_L = \operatorname{argmin}_{\tau} \{D_L(\tau), \tau_{\min} < \tau < \tau_{\max}\} \quad (4)$$

式 (3) 中， $\frac{1}{2} \sum_{j=1}^{2\tau} s_j$ 的作用为：相对于能量归一化，保证 $D_L(\tau)$ 在基音周期处的准确性和连续性。

1.4 LVAMDF 算法改进

在本文语音基音周期端点检测过程中，改进 LVAMDF 算法在语音处理两个不同阶段进行，两部分整合到一起，使得检测结果非常精确。这两个部分分别为：周期端点检测部分，倍频和半频识别修正部分。

1.4.1 在周期端点检测部分 LVAMDF 的优化

在语音信号基音周期检测中，识别精度关键在于基音周期之间相似度 (similarity)^[13] 的量化评估上，其定义如公式 (6) 所示。本文使用的基音周期之间相似度量化方法是改进的 LVAMDF 算法。

由于语音信号有振幅的高低起伏变化，相邻周期信号的对比差异就不能在同一个视角下观察，为了将所有相邻周期信号的对比运算大概放到同一个视角下进行，本文提出，用两帧信号中最大的振幅和 1 的比例，对两帧信号进行缩放，然后再进行对比运算。所以本文在式 (3) 分母中运用的是 $\operatorname{MAX}(s_j, s_{j+\tau}) * \tau$ ，保证所有相邻周期对比运算在同一个视角上进行。如式 (5) 所示：

$$D_L(\tau) = \frac{\sum_{j=1}^{\tau} |s_{j+\tau}|}{\operatorname{MAX}(s_j, s_{j+\tau}) * \tau}, \tau = \tau_{\min} \sim \tau_{\max} \quad (5)$$

$$\operatorname{similarity} = D_L(\tau) \quad (6)$$

此外，文献 [5] 中提出 LVAMDF 中 τ_{\min} 和 τ_{\max} 分别为相邻帧基音周期的 $\pm 0.1 \sim 0.2$ 倍。为了更加精确地估算 τ_{\min} 和 τ_{\max} 值，避免帧和帧之间周期长度差异超过估算范围（相距比较远的帧，周期长度相差比较大），本文将相邻周期基音周期长度的 $\pm 0.1 \sim 0.2$ 倍分别作为 τ_{\min} 和 τ_{\max} 。即定义为式 (7) 和 (8)：

$$\tau_{\max} = L_{\text{相邻已经确定的周期}} * 1.2 \quad (7)$$

$$\tau_{\min} = L_{\text{相邻已经确定的周期}} * 0.8 \quad (8)$$

因此，准确找到一帧中的第一个周期至关重要，本文提出了和文献 [5] 不同的第一周期寻找方法，其流程如图 2 所示。

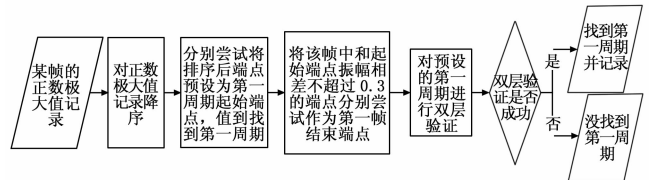


图 2 寻找某帧第一个基音周期流程图

对图 2 的解释和补充：

- 1) 周期端点都在正数极大值上寻找。
- 2) 第一周期的起始端点设定在尽可能大的正数极大值点上，第一周期的结束端点在起始端点的附近；
- 3) 第一周期起、止端点振幅大小比值允许范围为 0.7~1.3，大小比值公式如式 (9) 所示：

$$\frac{\operatorname{AmplitudeValue}_{\text{第一周期两端点中较小者}}}{\operatorname{AmplitudeValue}_{\text{第一周期两端点中较大者}}} \quad (9)$$

4) 本文在选择第一周期结束端点的时候，是从符合 1)、2) 和 3) 条的预选端点中，先向后面由近及远尝试划分，如果没有找到，再向前由近及远尝试划分，运用这种方法，当找到第一个通过第一周期验证的周期，就是第一周期。

5) 第一周期的验证方法为双层验证法。第一层验证为第一周期和第二周期相似度大小高标准验证（由于相邻周期的周

期长度并非完全相等, 所以在对比时有左对齐对比和右对齐对比, 对于第一层验证, 两种对齐方式的对比都要作, 以保证第一周期的起、止位置最合适), 即相似度量值门限设定为 0.172 9 (这个门限是由语音信号统计出来的), 并且在满足该门限的同时选择相似度量值最小值; 第二层验证为该帧中其他相邻周期之间相似度 (只作新端点方向对齐对比) 大小低标准通过并且能顺次将该帧完全划分, 其相似度量值门限设定为 0.327 5, 并且在两周期相似度的量化中, 实时改变 LVAMDF 算法中的窗长变化范围。

该方法大大减小了运算量, 降低了倍频的可能性。

1.4.2 在倍频和半频识别修正部分 LVAMDF 的优化

在周期端点检测部分完成后, 划分的端点中会有倍频或者半频划分, 引起原因为: 一帧中满足第一周期条件的第一周期和第二周期的相似度量值选择错误, 使得第一周期识别为倍频或者半频。

为解决这种问题, 本文对一帧所有满足第一周期条件的第一周期和第二周期的相似度量值进行优化选择, 使其尽可能减少倍频和半频划分。倍频和半频检测流程如图 3 所示。

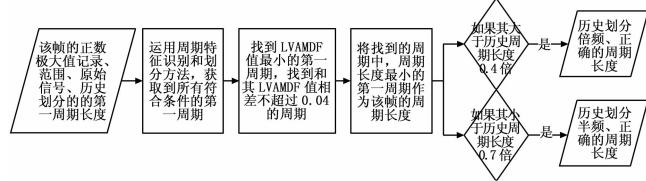


图 3 倍频和半频检测流程图

本文首先对倍频或半频划分帧中所有符合条件的第一周期进行统计, 在统计中发现, 正确的第一周期总是分布在相似度量值最小值附近, 所以在相似度量值最小值附近找第一周期端点。

在文献 [14] 中, 运用预判谷值范围方法的作用对象为归一化 LVAMDF 分析曲线, 目的是滤除归一化 LVAMDF 分析曲线的毛刺谷值 (倍频或半频谷值), 如图 4 第一条虚线所示, 而本文运用预判谷值范围作用对象为一帧中所有满足第一周期条件的相似度量值集合, 目的是确定正确的周期端点, 并且不需要计算归一化 LVAMDF 分析曲线, 只是计算第一周期相似度量值, 而且计算的帧数只有几个 (一段语音中按照周期长度差异大小, 将周期长度分为不同的种类, 只确定同一类中任意一帧的第一周期划分是否正确即可), 这样极大的减少了运算量。

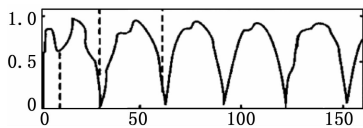


图 4 周期性强的语音段的 LVAMDF

本文统计计算出来的预测谷值范围如式 (10)、(11):

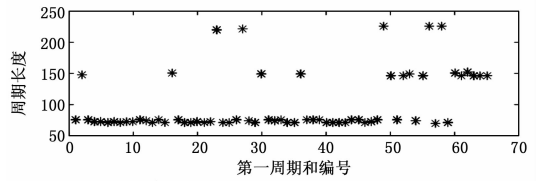
$$[similarity_{min}, similarity_{min} + 0.04] \quad (10)$$

$$similarity_{min} = \min_{\text{某帧}} (similarity(1 \cdots n_{\text{某帧第一周期}})) \quad (11)$$

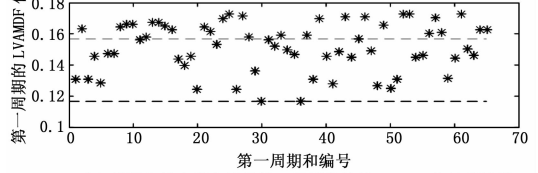
$similarity(1 \cdots n_{\text{某帧第一周期}})$ 是某帧中所有满足第一周期条件的相似度量值集合。

那么, 在该范围内所有预选周期中长度最小的为第一周期。通过和历史划分周期长度对比, 就可以判断出历史划分是

倍频还是半频, 如图 5 所示。



(a) 某语音帧中所有满足第一周期条件的周期长度及其编号



(b) 某语音帧中所有满足第一周期条件的 LVAMDF 值及其编号

图 5 第一周期准确获取图例

通过对 LVAMDF 值的优化选择, 识别和修复了倍频和半频的划分。

2 实验结果

本文改进 LVAMDF 算法及其综合调整周期长度、端点振幅和最优相似度选取因素检测方法通过 Matlab 实现, 实验对象为基本涵盖 4 个音调不同字母发音的 189 个语音信号, 进行处理, 其结果如下所述。

图 6~图 9 分别是 ge1, ge2, ge3, ge4 的原始信号时域图和本文方法、AMDF 算法、ACF 算法的基音周期时间轨迹变化图。

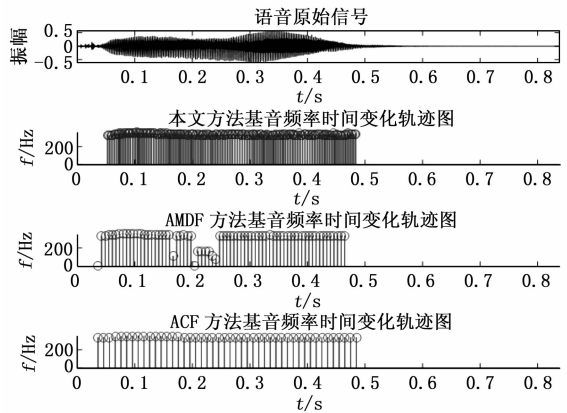


图 6 ge1 基音周期时间轨迹变化对比图

从图 6~图 8 中, 可以看到 AMDF 算法有较多的错误或者半频划分, 但是本文算法和 ACF 算法没有错误的划分, 本文算法有自动识别不适合 LVAMDF 算法语音段的能力, 并且将这部分语音段空出来, 如图 8 所示。

本文算法可将周期特征明显段的所有基音周期精确检测出来, 并且和 AMDF 算法检测一个基音周期的运算量相当。而 AMDF 和 ACF 算法只在一帧中检测一个基音周期, 如果用 AMDF 和 ACF 算法将所有基音周期都检测出来运算量会成倍的增加。

在实际应用中, 语音信号中会有噪声混杂, 为了检测本算法的鲁棒性, 在原始语音中加入高斯白噪声, 分别采用 AMDF 法、ACF 法和本方法在不同信噪比下对语音信号进行基音检测, 检测结果如表 1 所示。

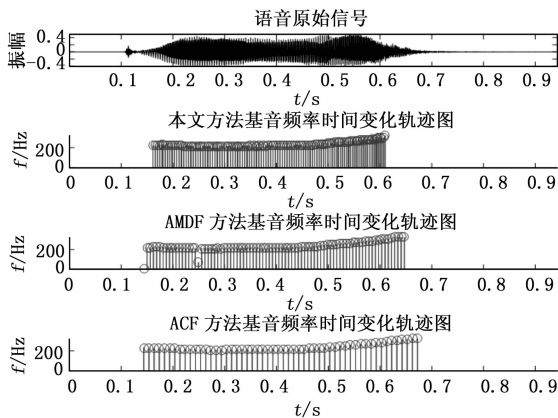


图 7 ge2 基音周期时间轨迹变化对比图

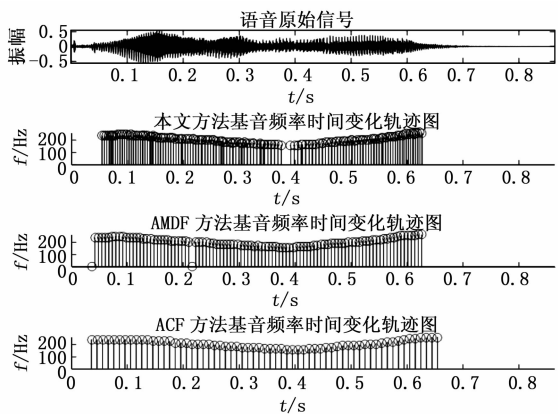


图 8 ge3 基音周期时间轨迹变化对比图

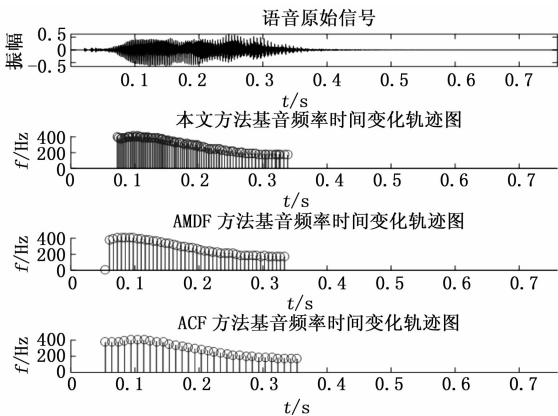


图 9 ge4 基音周期时间轨迹变化对比图

表 1 3 种方法在不同信噪比下的基音检测结果

检测方法	基音检测正确率				
	无噪	20 dB	15 dB	10 dB	5 dB
AMDF	0.920	0.911	0.887	0.842	0.752
ACF	0.965	0.968	0.967	0.967	0.964
本文算法	1	0.999	0.998	0.999	0.999

由表 1 可知, AMDF 算法在信噪比降低的时候, 基音检测正确率急速下降, ACF 算法和本文算法基本保持平衡, 本文算法正确率相对于 ACF 算法有所提高。

由于 ACF 算法中主要运用的是乘法运算, 而本文算法中只是运用的加减运算, 所以本文算法运算速度快, 和 AMDF

算法运算速度相当。

从算法的精确度、运算量、鲁棒性来看, 本算法适合用于实时语音信号处理中。

3 结束语

根据语音信号的特点, 本文分别用有效翻转率、信号和横轴围绕面积、语音频谱幅值均值, 将静音语音段、非语音段、清音段排除。又将浊音信号划分为周期特征明显段和周期特征不明显段, 在 LVAMDF 传统算法基础上, 加上信号同视角变化相似度比较、第一周期的精确估计、LVAMDF 窗长根据相邻周期长度实时变化, 精确的获取到基音周期端点, 对第一周期判定中少许的倍频和半频划分, 进行识别修复的处理, 在修复过程中, 通过分析整帧中所有符合第一周期的相似度量化值, 找到最优的第一周期, 大大地降低了倍频和半频划分, 最后通过统计方法将未能修复的倍频和半频划分(很少的可能)去除掉。在接下来的研究中, 要对浊音中少部分周期特征不明显段进行精确的基音周期端点检测。

参考文献:

- [1] Jong soon J, Jeong Jin K, Myung J B. Pitch alteration technique in speech synthesis system [J]. IEEE Transactions on Consumer Electronics, 2001, 47 (1): 163-167.
- [2] 马莎莎, 戴曙光, 穆平安. 基于短时能量的循环 AMDF 基音检测算法 [J]. 计算机仿真, 2014, 31 (7): 278-282.
- [3] 郑继明, 王劲松. 语音基音周期检测方法 [J]. 计算机工程, 2010, 36 (10): 273-275.
- [4] 潘峥嵘, 戴 芮, 张 宁. 改进的基音周期检测算法研究 [J]. 计算机工程与应用, 2015, 51 (10): 223-226.
- [5] 顾 良, 刘润生. 利用声调判别提高汉语数码语音识别性能 [J]. 清华大学学报: 自然科学版, 1998, 38 (9): 36-39.
- [6] 张康杰, 赵 欢, 饶居华. 基于 LV-AMDF 的自适应基音检测算法 [J]. 计算机应用, 2007, 27 (7): 1674-1679.
- [7] Chrjstensen O. Localization and navigation of a mobile robot using natural point landmarks extracted from data [J]. Robotics and Autonomous Systems, 2000, 33 (3): 131-142.
- [8] Krubsack D A, Niedeohn R J. An autocorrelation pitch detector and voicing decision with confidence in ensures developed for noise corrupted speech [J]. IEEE Trans. on Acoustics, Speech, Signal Processing, 1991, 39 (2): 319-329.
- [9] 焦 蓓, 曾以成, 毛燕湖. 基音周期检测的希尔伯特-黄变换方法 [J]. 计算机工程与应用, 2015, 51 (1): 204-207.
- [10] Kadame S, Broudreux B G F. Application of wavelet transform for pitch detection [J]. IEEE Trans. on IT, 1992, 38: 917-924.
- [11] 吴恬盈. 一种新的清浊音在线辨识法 [J]. 计算机仿真, 2007, 24 (4): 72-74.
- [12] Jalil M, Butt F A, Malik A. Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals [A]. 2013 International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE) [C]. IEEE, 2013: 208-212.
- [13] 陈雪勤, 刘 正, 赵鹤鸣. 基于相似度的高精度基音检测算法 [J]. 声学技术, 2008, 27 (5): 704-707.
- [14] 成新民, 曾毓敏, 赵 力. 一种改进的 AMDF 求取语音基音的方法 [J]. 微电子学与计算机, 2005, 22 (11): 162-164.