

一种语音信号非周期性、周期性及基频检测的改进方法

杜 硕 杜利民

(沃克斯技术院 北京 100080)

摘 要: APP 方法可以准确检测语音信号中的非周期性、周期性和基频,是近年提出的一种先进检测新方法,对于语音基础研究和语音技术应用研究有重要作用。APP 方法的最大优点是可以同时检测语音信号的基频周期、周期成分和非周期成分的能量比例,而最大缺点是计算代价巨大,运行时间为 110 倍实时,成为实际应用的最大障碍。该文在深入剖析 APP 方法的基础上,从原理架构和技术实现两个方面消除不合理的冗余处理,提出新的改进途径,发展成为改进的 APP 方法,即 MAPP 方法。MAPP 方法不但加强了 APP 方法处理机制的合理性基础,改善基频检测的准确性和鲁棒性,而且提高计算效率约 1 个数量级,在 CPU 时钟频率为 1.70GHz 和内存为 512MB 的 Pentium 计算机上的运行时间加快到 12.3 倍实时。

关键词: 语音信号处理; 基频检测; 周期能量; 非周期能量; 平均幅值差分函数

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2008)04-0929-04

Modified Detection of Aperiodicity, Periodicity and Pitch in Speech

Du Shuo Du Li-min

(Voxeas Institute of Technology, Beijing 100080, China)

Abstract: The APP method is capable to provide excellent estimation of speech aperiodic / periodic measurement and pitch simultaneously which is useful in speech research and application. However, due to its heavy computational load, APP system is about 110 times real-time, being an extinct flaw for utilization. In this paper, a modified detection of aperiodicity, periodicity and pitch in speech (MAPP) method is presented, which maintains the merit of APP method and eliminates the redundancy of configuration and computation, rationalizing the methodology. Computer simulation shows that MAPP method maintains high accuracy and robustness and that the system is improved to 12.3 times real-time on Pentium processor with 1.70GHz CPU and 512MB RAM, speeding up about one order of magnitude.

Key words: Speech processing; Pitch detection; Periodicity energy; Aperiodicity energy; Average Magnitude Difference Function (AMDF)

1 引言

人说话时声道中存在多个产生声学能量的声源,其中非周期声源包括在唇、齿、喉、声道等处产生的送气声、摩擦声、爆破声,而周期声源则是在声门处由声带振动产生。这些声源通过声道的滤波和嘴唇处的辐射,产生在空气中传播的语音信号。语音信号中既包含周期成分,也包含非周期成分,二者的多寡,取决于人在发声时所涉及的声源种类和强弱。在语音基础研究和语音技术应用研究领域,相当多工作需要正确提取语音信号的基频周期^[1-6]以及周期成分与非周期成分的比例^[7-13]。

APP(Aperiodicity, Periodicity, Pitch)是近年提出的一种在时域同时估计语音信号的多种声学表征的有效方法^[14-16]。该方法的核心是对每一帧语音信号,先按照听觉

滤波器特性^[17]分解成 60 个频带或通道,接着对每一个频带或通道顺序提取希尔伯特包络(Hilbert Envelope)^[18],再估计希尔伯特包络的平均幅度差函数 AMDF(Average Magnitude Difference Function)^[19]及其凹点的强度,然后将 60 个通道平均幅度差函数的凹点强度进行汇总,构成自动测量语音信号的基频周期、周期成分比例和非周期成分比例的度量基础。最后通过一系列后处理,给出当前语音信号帧的基频周期、周期成分比例和非周期成分比例。

与现有其它基频检测的方法相比,APP 方法不但能够可靠检测语音信号中周期成分的基频周期,还能够同时计算语音信号中周期能量和周期能量的比例,因而对语音基础研究和语音技术应用研究有更大的作用。例如,APP 方法可以用于:(1)标注语音的浊音段和清音段;(2)识别周期成分和非周期成分同时存在的音段;(3)根据非周期度量的强弱区分声音的发音部位和浊音程度。但是,现有 APP 方法的最大缺

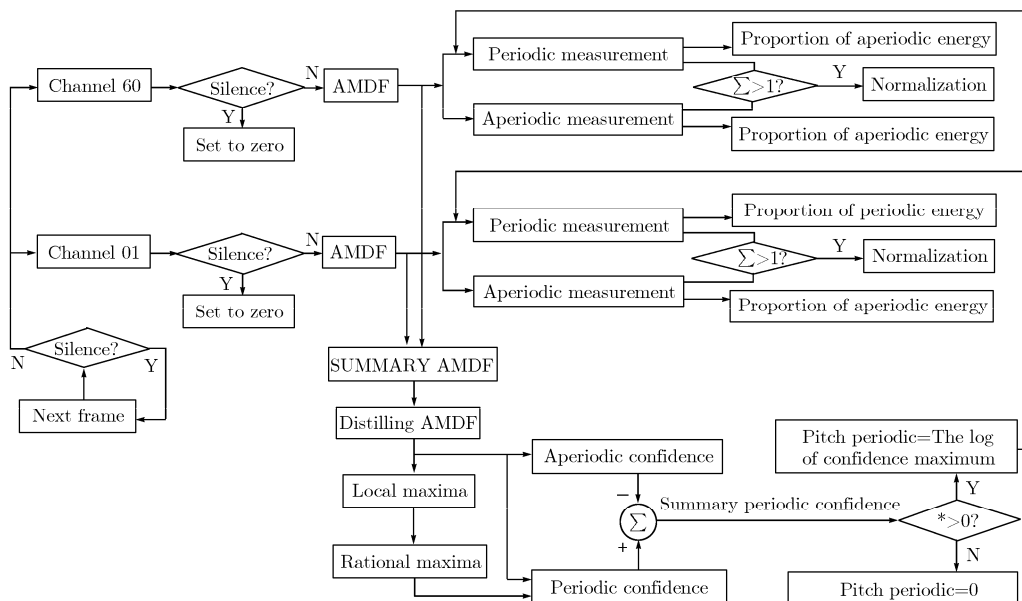


图1 MAPP 方法的原理架构

点是计算复杂度高, 处理信号的实时性低, 在 CPU 时钟频率为 1.7GHz 和内存为 1GB 的 Pentium 计算机系统中运行的速度为 110 倍实时, 成为应用于实际的最大障碍。

该文从 APP 方法的原理架构和技术实现方面进行改进提高, 目的在于获得较好的处理效果和更高的实时速度。

2 MAPP 方法的设计与实现

分析发现, APP 方法的计算复杂度主要源于其原理架构和技术实现的一些不合理配置。首先, APP 方法采用听觉滤波器将信号分解成多通道的动机与信号处理的实际效果并不一致。将信号分解成多通道处理的最大益处更可能在于两个方面^[20]: (1)各通道的信号带宽比原始信号的带宽更窄, 有利于在 AMDF 中展示信号的周期特征; (2)由于浊声语音信号具有丰富的谐波结构, 多个通道的 AMDF 在基频周期及其整数倍附近将比在其它位置更有可能共现凹陷特征, 因而提高检测基频周期的可靠性和鲁棒性。其次, 滤波器带通形状的影响在这里并不至关重要。按照人在语音感知中对各种频率成分加权的的关系, 特别是根据人对语音周期成分和非周期成分感知的相对强度的关系设计的滤波器形状在 APP 方法中的作用机理并不明确, 也不能对此进行直接或间接的试验佐证。因此, 突出多通道分解处理的实质, 但通过简化滤波器带通形状来降低计算复杂度的尝试是有实际意义的。

另外, 对每个通道提取希尔伯特包络的处理并非必要, 相反会降低多通道处理的积极效果, 还引入大量计算负担。关于希尔伯特包络处理对基频周期检测的负面影响主要是某些通道中基频周期及其整数倍位置附近的 AMDF 凹陷消失, 从而降低该通道的积极贡献^[20]。其原因在于: (1)浊语音信号经过多通道分解后, 低频窄带通道的输出接近正弦波, 其周期对应于语音的基频周期; (2)窄带信号的希尔伯特变换

结果为变换前信号的 90 相移, 将希尔伯特变换前后的信号进行复数相加, 然后取模得到的希尔伯特包络只保留了信号峰值的轮廓, 对应基频周期的信号细节完全消失了, 因此, 该通道的 AMDF 在后续处理中对基频周期检测的积极贡献也消失了。

最后, 先对每一个通道的 AMDF 进行凹点的强度检测, 然后再进行多通道汇总的处理机制不但要付出极大的计算代价, 还会降低多通道处理机制的积极作用。由于语音浊声信号是宽带的、非理想的周期信号, 各个通道信号成分的周期可能有一定程度的偏差, 即通道 AMDF 的凹陷峰的位置与信号基频周期及其整数倍位置都不可避免地发生一定偏移。先取凹点强度后汇总, 各个通道的 AMDF 仅仅对本通道的凹点所对应的周期分量进行投票, 放弃了对其他所有可能的周期分量进行投票的积极贡献, 从而降低了多通道处理的积极作用^[20]。因此, 消除对每一个通道的 AMDF 进行凹点强度检测的处理也是非常具有积极意义的。

基于上述信号处理机理的研究, 构成该文的改进 APP (即 MAPP)方法的原理架构如图 1 所示。

MAPP 方法中通道滤波器组的参数设计, 是根据人耳感知灵敏度与频率的大致关系, 按照式(1)先将线性刻度的频率 f 转换到美刻度频率 F_{mei} :

$$F_{mei} = 2595 \left(\lg \left(1 + \frac{f}{7000} \right) \right) \quad (1)$$

然后在美刻度频率空间进行等间隔划分, 并使相邻滤波器组的带通范围部分重叠, 构成美频率空间等带宽近邻重叠的带通滤波器 (Equal Band-Pass and Overlap Filterbank in Mei-Frequency Scale, MF-EBPOFB), 见图 2 所示。

为了在计算复杂度和带通性能之间取得较好的折衷, 所有滤波器的实现都采用 4 阶 IIR 结构。

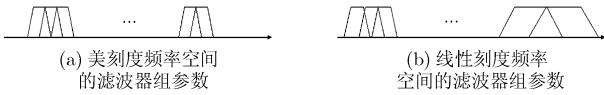


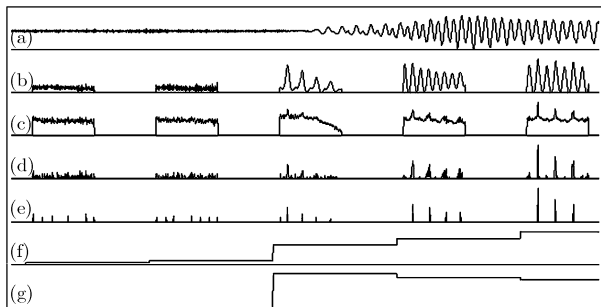
图 2 美刻度频率空间等带宽、带通范围重叠的滤波器组(MF-EBPOFB)

在 MAPP 方法中，对每一帧不为静音的语音信号，经过 MF-EBPOFB 滤波器组分解成 60 个通道信号，对于不为静音的通道信号直接计算 AMDF，然后按照式(2)进行取反归一处理。

$$r_i(k) = \frac{(r_o_max - r_o(k))}{r_o_max} \quad (2)$$

其中 k 为 AMDF 的延迟点， $r_o(k)$ 为 AMDF， r_o_max 为当前帧中 $r_o(k)$ 的最大值， $r_i(k)$ 为取反归一的 AMDF。在图 3 (c) 中，60 通道 AMDF 的汇总曲线由整体取向和叠加在整体取向之上的波纹两部分组成，前者反映 60 个通道对 AMDF 的每一个延迟点 k 的周期性感知的平均阈值，后者为超过平均阈值的周期性强度差异。由于整体取向因信号成分的不同呈现多样性变化，会直接影响到后续的自动处理。该文通过图 4 所示的局部均衡处理^[20]，即将图 3 (c) 曲线中的每一点用该点的值与该点邻域 δ_1 的均值的差来替换，结果如图 3 (d) 所示，很好地消除了图 3 (c) 曲线中各个处理帧的不同整体取向。

图 3(d)所示的曲线主要反映语音信号的周期性强度差异。信号周期性越强，曲线在周期及其整数倍附近的峰值越高；信号非周期性越强，曲线的峰值越低，峰值点分布范围越宽。为寻找周期最大值，该文用图 5 所示的局部最大值算法^[20]，即对图 3(d)的每一点搜索其邻域 δ_2 中的最大值，如果该点与其邻域的最大值相等，则该点为局部最大值，算法处理的结果见图 3 (e)。由于局部最大值估计算法的邻域是固定的，当语音信号的基频周期大于邻域宽度 δ_2 的时候，可能出现虚假的基频周期估计，还需要对每一个局部最大值进行如图 6 所示的基频周期的置信度检验^[20]。如果其中的最大置信度的值大于 0，则该帧为周期信号，基频周期为最大置信度的局部最大值所对应的时间延迟 T_0 ，基频 $f_0=1/T_0$ ；如果最



(a) 原始语音信号 (b)(a) 经过 800Hz 低通滤波后的 AMDF (c)60 通道 AMDF 的汇总 (d)(c) 经过均衡处理的结果 (e)(d) 中局部最大值 (f) 语音帧的周期置信度 (g) 语音帧的基频周期

图 3 MAPP 方法中各探测点信号波形

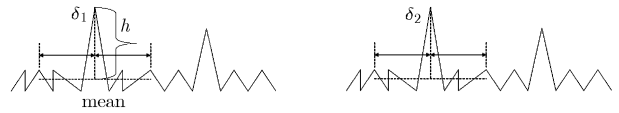


图 4 局部均衡处理

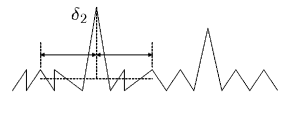


图 5 局部最大值算法

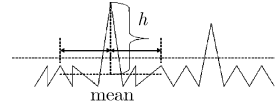
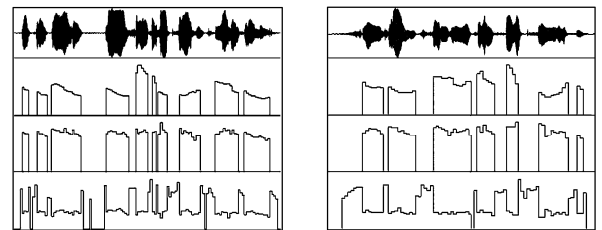


图 6 周期估计的置信度检验

大置信度的值小于等于 0，则该帧为非周期，令 $f_0=0$ 。图 3(f)和 3(g)分别为本方法检测到的语音信号的基频周期置信度和基频 f_0 。

3 试验与讨论

MAPP 方法采用 C 语言编程，语音信号的帧长和帧移可以改变，该文图示的试验结果为 30ms 帧长，30ms 帧移。语音信号的采样频率为 16kHz，每个样点的量化精度为 16bit。部分实验结果如图 7 所示。



(a) “那走道的, 马上就把手套脱了下来” (b) “身上穿着一件厚袍子”

图 7 为两个自然语句的基频和周期性与非周期性检测的试验结果(每一个图例从上往下, 第 1 个窗口为语音波形; 第 2 个窗口为检测的基频; 第 3 个窗口为检测的周期性能量; 第 4 个窗口为检测的非周期性能量)

首先，试验表明了 MAPP 方法对 1/2 基频和 2 倍基频有比较强的抑制作用，见图 3(c)，这主要受益于多通道 AMDF 的直接汇总处理。

对比分析原始语音信号经过 800Hz 低通滤波后的 AMDF，见图 3(b)，最后 2 帧在 1/2 基频周期及其整数倍的附近有很强而且幅值相近的凸峰，对后续处理干扰较大。

试验还表明 MAPP 方法的多通道 AMDF 直接汇总的结果比 APP 方法的多通道 AMDF 极值点汇总的结果具有更加明显的周期性凸现作用和更好的鲁棒性，这主要得益于取消对每个通道进行希尔伯特包络处理和对每个通道 AMDF 的凸点检测。

最后，对自然语句的基频周期、周期性与非周期性检测的试验结果表明，MAPP 方法在不进行帧间动态规划或帧间平滑等后处理的情况下都能获得如图 7 所示准确而平滑的整句基频曲线。

MAPP 方法的计算效率试验在与文献[14]相当的情况下

(帧长为 20ms, 帧移为 2ms, Pentium 计算机的 CPU 时钟频率为 1.70GHz、内存为 512MB)进行, MAPP 方法的运行时间为 12.3 倍实时, 计算效率比 APP 方法提高约 1 个数量级。

4 结论

该文简化滤波器形状, 采用在美刻度频率空间等带宽、相邻滤波器带通范围重叠的 4 阶 IIR 滤波器组(MF-EBPOFB), 大大降低了计算复杂度。

发现多通道滤波器组输出信号的希尔伯特包络处理有损低频窄带通道信号的周期特征的消极作用, 在 MAPP 原理架构中消除了不合理的希尔伯特包络处理, 进一步降低了计算复杂度。

该文提出的反向归一的 AMDF、多通道 AMDF 直接汇总、多通道 AMDF 汇总的均衡等信号处理方法, 对于提高基频周期的检测准确度和抑制 1/2 基频和 2 倍基频干扰有明显的积极作用。

在相同的试验条件下, MAPP 方法的计算效率比 APP 算法提高约 1 个数量级。

在更大规模标定周期/非周期性能量比率和基频周期的数据库上, 进一步研究检验 MAPP 方法的精确度和鲁棒性, 进行更合理性的精细化完善是该文后续的一项非常有意义和令人感兴趣的研究工作。

参考文献

- [1] 刘建, 郑方, 吴文虎. 基于幅度差平方和函数的基音周期提取算法[J]. 清华大学学报(自然科学版), 2006, 46(1): 44-77.
Liu Jian, Zheng Fang, and Wu Wen-hu. Real-time pitch tracking based on sum of magnitude difference square function[J]. *Journal of Tsinghua University(Science and Technology)*, 2006, 46(1): 44-77.
- [2] Luengo I, Saratxaga I, and Navas E, *et al.* Evaluation of pitch detection algorithms under real conditions[C]. ICASSP'07 Proc., Hawaii, USA, Apr. 15-20, 2007: 1057-1060.
- [3] Li Y and Wang D L. Pitch detection in polyphonic music using Instrument tone models[C]. ICASSP'07 Proc., Hawaii, USA, Apr. 15-20, 2007: 481-484.
- [4] Roa S, Bennewitz M, and Behnke S. Fundamental frequency estimation based on pitch-scaled harmonic filtering[C]. ICASSP'07 Proc., Hawaii, USA, Apr. 15-20, 2007: 397-400.
- [5] Joho D, Bennewitz M, and Behnke S. Pitch estimation using models of voiced speech on three levels[C]. ICASSP'07 Proc., Hawaii, USA, Apr. 15-20, 2007: 1077-1080.
- [6] Wohlmayr M. Joint position-pitch extraction from multichannel audio[C]. Interspeech2007 Proc., Antwerp, Belgium, August 27-31, 2007: 303-306.
- [7] Brown G and Cooke M. Computational auditory scene analysis[J]. *Computer Speech and Language*, 1994, (8): 297-336.
- [8] Ellis D P W. Using knowledge to organize sound: the prediction-driven approach to computational auditory scene analysis, and its application to speech/nonspeech mixtures[J]. *Speech Communications*, 1999, (27): 281-298.
- [9] Yegnanarayana B, d'Alessandro C, and Darsinos V. An iterative algorithm for decomposition of speech signals into periodic and aperiodic components[J]. *IEEE Trans. on Speech Audio Process.*, 1998, 6(1): 1-11.
- [10] d'Alessandro C, Darsinos V, and Yegnanarayana B. Effectiveness of aperiodic and periodic decomposition method for analysis of voice sources[J]. *IEEE Trans. on Speech Audio Process.*, 1998, 6(1): 12-23.
- [11] Fujimura O. Approximation to voice aperiodicity. *IEEE Trans. on Audio Electroacoust.*, 1968, AU-16(1): 68-73.
- [12] Jackson P and Shadle C. Frication noise modulated by voicing, as revealed by pitch-scaled decomposition[J]. *J. Acoust. Soc. Amer.*, 2000, 108(4): 1421-1434.
- [13] Serra X and Smith J. Spectral Modeling Synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition[J]. *Comput. Music J.*, 1990, 14(4): 12-24.
- [14] Deshmukh O, Espy-Wilson C Y, and Salomon A, *et al.* Use of temporal information: detection of periodicity, aperiodicity, and pitch in speech[J]. *IEEE Trans. on Speech and Audio Processing*, 2005, 13(5): 776-786.
- [15] Deshmukh O and Espy-Wilson C. Detection of periodicity and aperiodicity in speech signal based on temporal information[C]. 15th Int. Congr. Phonetic Sciences Proc., Barcelona, Spain, 2003: 1365-1368.
- [16] Deshmukh O and Espy-Wilson C. A measure of periodicity and aperiodicity in speech[C]. IEEE ICASSP Proc., Hong Kong, China, 2003: 448-451.
- [17] Glasberg B R and Moore B C J. Derivation of auditory filter shapes from notched-noise data[J]. *Hear. Res.*, 1990, 47 (1-2):103-138.
- [18] Johansson M. The Hilbert transform[D]. [Master thesis]. Vaxjo University, 1999.
- [19] Ross M, Shaffer H, and Cohen A, *et al.* Average magnitude difference function pitch extractor[J]. *IEEE Trans. on Signal Processing*, 1974, 22 (5): 353-362.
- [20] 杜硕. 语音信号的周期性、非周期性及基频的检测[D]. [学士论文]. 北京工业大学, 2007.

杜 硕: 女, 1985 年生, 学士, 研究方向为信号与信息处理。

杜利民: 男, 1957 年生, 研究员, 教授, 博士生导师, 研究方向为信号与信息处理。