

语音基音周期检测方法

郑继明¹, 王劲松²

(1. 重庆邮电大学应用数学研究所, 重庆 400065; 2. 重庆邮电大学计算机科学与技术学院, 重庆 400065)

摘 要: 针对现有语音基音检测算法抗噪能力低的问题, 利用多尺度连续小波对语音信号进行预处理。基于短时平均幅度差函数和短时自相关函数, 提出一种语音基音周期检测方法, 在确保运算量较低的前提下, 获得准确结果。使用平滑算法消除倍频点、半频点和随机点产生的误差, 实现基音曲线平滑。

关键词: 连续小波; 多尺度分析; 基音周期检测; 自相关函数

Method for Voice Pitch Period Detection

ZHENG Ji-ming¹, WANG Jin-song²

(1. Institute of Applied Mathematics, Chongqing University of Posts and Telecommunications, Chongqing 400065;
2. College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065)

【Abstract】 Current pitch period detection algorithms can not work well under high noise environment. Aiming at this problem, this paper utilizes multi-scale continuous wavelet to pre-filtering the signal of voice. Based on the short-time autocorrelation function and the average magnitude difference function, a method for voice pitch period detection is proposed. This method needs little computation expense and can obtain the right pitch period. Smoothing algorithm is exploited to remove the errors of multiple frequency point, half frequency point and random point, and realize pitch curve smoothing.

【Key words】 continuous wavelet; multi-scale analysis; pitch period detection; autocorrelation function

1 概述

语音信号具有3个物理特征: (1)响度, 表示声音能量的强弱程度, 取决于声波振幅的大小; (2)音高, 又称音调, 取决于声波基频的高低, 其单位用赫兹(Hz)表示; (3)音色, 又称音品, 由声音波形的谐波频谱和包络决定。声音波形基频产生的能听得最清楚的音称为基音。每个基音都有固有频率, 用于区别于其他具有相同响度和音高的声音, 因此, 基音是声音的重要特征。语音信号是由气流冲击声带, 使声带震动而产生的, 其震动频率的倒数为基音周期, 即震动产生的浊音信号之间的时间间隔。

2 基音检测

目前有很多基音检测方法, 传统方法仅利用自相关函数或平均幅度差函数提取语音基音周期, 效果不理想。主要原因包括: (1)实际生活中的语音不是理想的纯语音, 往往夹杂着高频的噪声或者带有背景音乐; (2)短时自相关函数包含的很多信息是多余的, 能反映基音周期性的只是少数几个峰, 其他多数峰都是由声道的谐振特性引起的, 称为共振峰。

为解决上述问题, 研究者提出小波系数法、短时样本自相关函数法(Sample AutoCorrelation Function, SACF)、短时平均幅度差函数(Average Magnitude Differential Function, AMDF)、倒谱法等^[1], 但此类方法都有其局限性。近年来出现了很多改进方法, 文献[2]对加白噪声的语音信号进行3层小波包分解, 并用小波系数加权法和互相关函数结合的方法(Weighted Wavelet Coefficient-Normalized Cross Correlation Function, WWC-NCCF)提取语音的基音周期, 取得了较好的实验结果。文献[3]提出了改进的平均幅度差方法(M-AMDF), 对男女老少的基音周期检测都有较高的准确度。文献[4]采用

改进的自相关函数法(P-ACF)对语音进行基音周期检测, 其实验结果较好。文献[5]提出线性预测倒谱和谐波频谱相结合的方法, 能大幅度减少基音频率提取中的错误。

本文提出一种基于多尺度连续小波滤波的方法, 先去除语音信号所夹杂的噪声, 并用AMDF算法算出AMDF的值。然后把每8位AMDF值修正为一位, 以避免计算ACF的乘法运算, 减少计算量。经过中心消波后, 用SACF方法提取语音的基音周期。最后用平滑算法修正每帧的基音曲线。本文方法具有如下优点: (1)因为连续小波具有紧支集且是光滑的, 所以小波函数和信号做卷积运算可以得到光滑的函数, 能降低滤波产生的误差。选取该光滑函数的低频部分(语音部分)作为信号源, 用于下一步平均幅度差函数的计算。(2)直接使用ACF时, 计算量较大, 而只用AMDF提取语音的基音周期时, 准确率较低。为了提高基音提取的准确率并避免乘法运算, 本文将AMDF与ACF方法相结合, 先用AMDF算法算出每帧信号的AMDF值, 把每帧语音信号AMDF值的每8位数据转换成一位数据信号, 再用修正后的数据做SACF运算。这2个函数结合在一起极大降低了运算量且消除了大部分共振峰, 提高了基音周期检测的准确率。(3)使用SACF前, 为克服剩下的共振峰对基音周期检测造成的影响, 用中心消波的方法去除一些多余的峰, 这样做的目的是进一步减少计算量, 并提高基音周期提取的准确率。(4)为了修正短时

基金项目: 重庆市教育委员会科学技术研究基金资助项目(KJ080524)

作者简介: 郑继明(1963—), 男, 副教授, 主研方向: 小波分析及其应用; 王劲松, 硕士研究生

收稿日期: 2009-12-14 **E-mail:** zhengjm@cqupt.edu.cn

自相关函数在基音周期检测中出现的跳跃点、半频点、倍频点和随机错误点,用平滑算法对得到的基音曲线进行平滑处理,从而得到准确完整的基音曲线。

3 多尺度连续小波滤波原理

墨西哥帽小波本质上是高斯函数的二阶导数,它的作用相当于一个低通滤波器,把信号高频部分过滤掉(高频部分往往是噪音和杂音),并用低频部分(语音部分)的信号源求自相关函数。

墨西哥帽小波函数形式如下:

$$\psi(x) = \frac{2}{\sqrt{3}} \pi^{-\frac{1}{4}} (1-x^2) e^{-x^2/2} \quad (1)$$

把构造的具有紧支集的墨西哥帽小波函数 $\psi(x)$ 和原始信号 s 做卷积变换。卷积变换公式如下:

$$w(k) = \sum_j s(j) \psi(k+1-j) \quad (2)$$

其中, j 是信号 s 每帧的长度; ψ 是墨西哥帽小波函数。由卷积得到的函数 $s * \psi$ 一般比 s 和 ψ 光滑,特别当 ψ 为具有紧支集的光滑函数时,它们的卷积 $s * \psi$ 也是光滑(正则)函数。上述处理能有效地分开低频信号和高频信号,降低了滤波产生的误差。

4 基于 AMDF 和 SACF 的语音基音周期检测方法

设信号 $w(n)$ 的周期为 T , 如果 m 为 0, 则语音信号在周期为 $\pm T, \pm 2T, \dots$ 时, 函数 $x(m) = w(m+n) - w(n)$ 的差异近似为 0, 因为语音信号是准周期的, 函数 $x(m)$ 的差异很小。基于该原理, 短时平均幅度差函数定义如下:

$$R(m) = \frac{1}{N-m-1} \sum_{n=0}^{N-m-1} |w(n+m) - w(n)| \quad (3)$$

其中, w 是经过小波滤波后的语音信号; N 是一帧语音信号的长度; m 的取值范围为 $0 \sim N$ 。若一帧语音信号是准周期的, 则每帧信号 AMDF 的值也是准周期的^[3], 如图 1(a)所示。

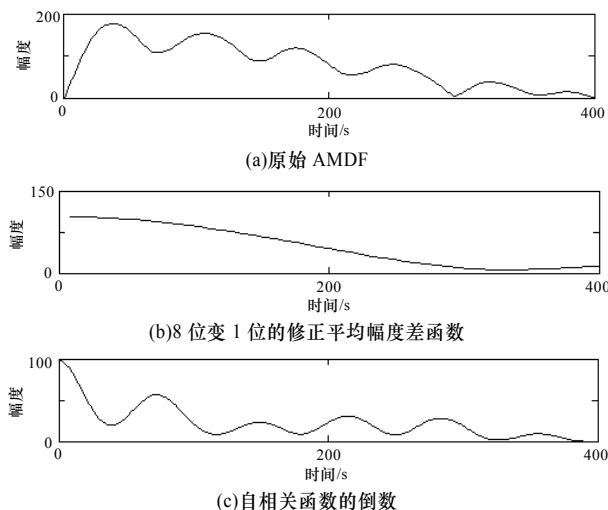


图1 基音检测算法波形

为了进一步减少计算量, 本文提出基于 AMDF 和 SACF 的语音基音周期检测算法, 其流程如图 2 所示。

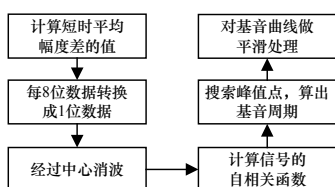


图2 本文算法流程

由每帧语音信号得出 AMDF 值。基音周期的值可以由 SACF 算出, 但是 SACF 需要多次乘法计算, 计算量比较复杂, 为了避免用乘法计算, 每 8 位的 AMDF 值转变为一位信号, 根据每帧语音信号的 AMDF 最大值 R_{\max} 和最小值 R_{\min} 计算该位数据:

$$\theta = a(R_{\min} + R_{\max}) \quad (4)$$

其中, 因子 a 设定为 0.5。

当 AMDF 的值小于 θ 时, a 为 1, 当 AMDF 的值大于 θ 时, a 为 0, 此时, 每 8 位的 AMDF 信号转变为一位信号。如图 2(b)所示。用 SACF 计算语音的基音周期。语音信号虽然是非平稳的随机过程, 但是在很短的时间里, 这种变化比较缓慢, 可以认为声音信号在较短的时间里是平稳的。选择合适的窗长可以算出基音周期^[6]。语音信号在一个长时的自相关函数中意义不大, 所以, 定义一个短时自相关函数较有效。AMDF 运算后的语音信号 θ 的短时自相关函数定义如下:

$$R_w(k) = \sum_{n=0}^{N'-1} \theta(n) \theta(n+k) \quad (5)$$

其中, $\theta(n)$ 是一位信号, 信号 $\theta(n)$ 只能是 1 或 0。当 $\theta(n+k)$ 和 $\theta(n)$ 都为 1 时, $\theta(n)\theta(n+k)$ 值为 1。当 $\theta(n)$ 或 $\theta(n+k)$ 为 0 时, $\theta(n)\theta(n+k)$ 为 0。8 位的自相关函数乘法运算被转换成一位的简单加法计算, 极大减少了计算量并降低了时间复杂度, 如图 1(c)所示。

5 基音周期的误差降低方法

5.1 中心消波

中心消波的目的是减小共振峰的影响, 由于语音信号的变换和清浊音的边界具有动态变化性和潜在损失性, 因此消波电平具有重要作用, 其定义如下:

$$p_b(t) = \begin{cases} p(t) + c_i & p(t) < -c_i \\ 0 & |p(t)| \leq c_i \\ p(t) - c_i & p(t) > c_i \end{cases} \quad (6)$$

其中, $p_b(t)$ 是中心消波函数; $p(t)$ 是信号 n 的周期; c_i 表示消波电平, 它的阈值凭经验设定, 一般取本帧语音信号最大幅度值的 70%, 因为共振峰的幅度通常比语音幅度小 30%。削波后信号的自相关函数在基音周期位置的峰值会更加尖锐, 可以有效减少倍频或半频错误。消波后的信号见图 3(b)。

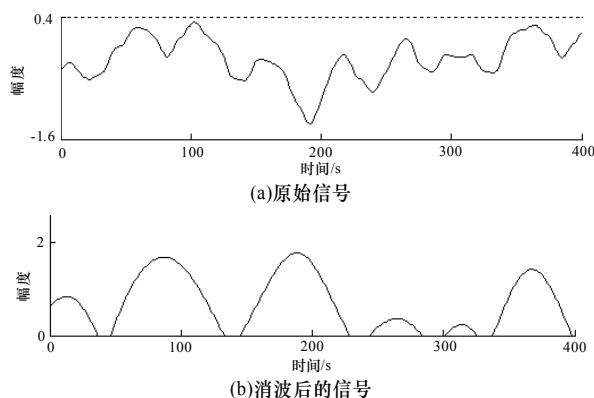


图3 语音信号中心消波

5.2 基音平滑方法

设 k_1, k_2, \dots, k_n 代表每帧的基音频率。在处理第 i 帧基音频率 f_i 时, 先处理基音提取过程中出现的倍频和半频问题。

如果^[5] $|f_i/2 - f_{i-1}| < C_1$, 则 $f'_i = f_i/2$ 。若 $|2f_i - f_{i-1}| < C_1$, 则 $f'_i = 2f_i$ 。

处理随机错误点时, 如果 $|f_i - f_{i-1}| > C_1$ 且 $|f_{i+1} - f_{i-1}| > C_2$,

则令 $f'_i = 2f_{i-1} - f_{i-2}$ ，若 $|f_i - f_{i-1}| > C_1$ 且 $|f_{i+1} - f_{i-1}| < C_2$ ，则令 $f'_i = (f_{i-1} + f_{i+1})$ ，其中， f'_i 代表第 i 帧平滑后得到的基音频率； C_1 和 C_2 是由实验决定的阈值， C_1 一般是连续 2 帧间的频率差的阈值， C_2 是间隔一帧的连续 2 帧间的频率差的阈值^[6]。

用基音平滑技术消除由基音曲线错误点引起的识别误差。该方法简单可靠、快速高效。平滑后的基音曲线见图 4。

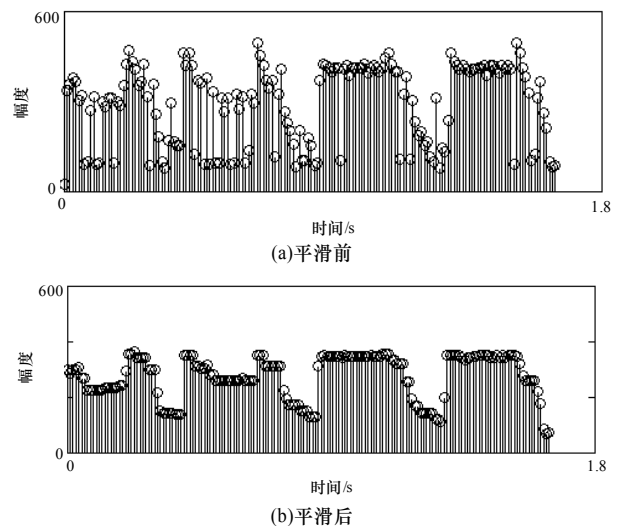


图 4 基音曲线

6 实验结果分析

本文选取一组具有代表性的样本(156 个纯语音和 320 个带有背景音乐的语音)，其频率为 22 kHz。分别用本文方法、小波加权系数法^[2]、修正的平均幅度差函数法^[3]和改进的自相关函数法^[4]对样本进行基音周期检测，结果如表 1 所示。可以看出，在纯语音情况下，4 种方法的检测准确率相近。而在带有背景音乐的情况下，本文方法的准确率高出其他方法。

表 1 实验结果

方法	纯语音样本 的准确识别数	带背景音乐的语音样本 的准确识别数	平均准确 识别率/(%)
WWC-NCCF	138	211	73.32
M-AMDF	121	155	57.98
P-ACF	125	201	68.50
本文方法	131	255	81.70

7 结束语

为了减少实验误差，本文通过中心消波去除共振峰造成的误差，并用基音平滑技术避免基音曲线出现错误点。实验结果表明，在带噪声的环境下，本文方法提取的语音基音周期准确率高于其他同类方法。

参考文献

[1] 李 辉, 戴蓓蓓, 陆 伟. 基于前置滤波和小波变换的带噪声语音基音周期检测方法[J]. 数据采集与处理, 2005, 20(1): 99-104.

[2] Li Ruwei, Bao Changchun, Dou Huijing. Pitch Detection Method for Noisy Speech Signals Based on Pre-filter and Weighted Wavelet Coefficients[C]//Proc. of the 9th International Conference on Signal Processing. Beijing, China: [s. n.], 2008: 530-533.

[3] Zeng Yumin, Wu Zhenyang, Liu Haibin, et al. Modified AMDF Pitch Detection Algorithm[C]//Proc. of the 2nd International Conference on Machine Learning and Cybernetics. Xi'an, China: [s. n.]: 2003: 470-473.

[4] 杨森斌, 陈砚圃, 李 真. 一种改进的自相关函数基音检测算法[J]. 现代电子技术, 2008, 31(9): 135-137.

[5] Ding Hui, Qian Bo, Li Yanping, et al. A Method Combining LPC-based Cepstrum and Harmonic Product Spectrum for Pitch Detection[C]//Proc. of Int'l Conference on Information Hiding and Multimedia Signal Processing. Pasadena, California, USA: [s. n.], 2006: 537-540.

[6] 朱小燕, 王 昱, 刘 俊. 汉语声调识别中的基音平滑新方法[J]. 中文信息学报, 2001, 15(2): 45-50.

编辑 陈 晖

(上接第 272 页)

5 结束语

本体明确并形式化地表达了领域知识的基本概念、属性和内在关系，提供了领域知识的处理方法，有很强的知识表示和推理能力，从知识工程的角度来看是构建知识库的理想技术。矿井电机故障知识库可以帮助维修人员准确的查找故障产生的原因和提高故障知识的复用性。本文分析了目前矿井提升机故障诊断的需求和现状之后，指出了知识库在矿井电机故障表示中的作用和地位。在研究知识库与矿井电机故障模型的基础上，建立了一个基于本体和规则的知识库模型。并对知识库的解析进行了实际的测试研究。但是就目前所做的工作而言，有许多问题值得进一步研究，进一步对知识库系统的完备性进行验证，系统的细节也需要在实际应用中得到更好的完善。

参考文献

[1] 周 瑾, 肖兴明. 基于故障树和神经网络模型的提升机故障诊断专家系统[J]. 矿山机械, 2004, (5): 39-41.

[2] 李吉宝, 李树涛. 提升机故障诊断专家系统研究[J]. 煤炭技术, 2005, 24(5): 14-16.

[3] 刘志海, 徐文尚, 鲁 青. 矿井提升机故障诊断专家系统的研究[J]. 煤矿机械, 2006, 27(5): 900-902.

[4] 蒋 维, 郝文宁, 杨晓恕. 军事训练领域核心本体的构建[J]. 计算机工程, 2008, 34(5): 191-192.

[5] Xu Binfeng, Luo Xiaogang, Peng Chenglin, et al. Based on Ontology: Construction and Application of Medical Knowledge Base[C]//Proc. of IEEE/ICME International Conference on Complex Medical Engineering. Beijing, China: IEEE Press, 2007.

[6] 吴振生, 孙秀迪, 李新云, 等. 基于本体的推理在行业信息化知识库中的应用[J]. 计算机工程, 2008, 34(18): 59-61.

编辑 金胡考