

含噪语音的基音周期提取算法

俞翠华¹ 行鸿彦¹ 李鹏¹

摘要

背景噪声对语音信号的基音周期提取有很大的影响. 结合能够对背景噪声进行有效降噪的 EMD (Empirical Mode Decomposition) 软阈值法, 提出了一种噪声背景下语音基音周期提取的方法. 首先将 EMD 软阈值降噪法和中心削波法相结合, 对信号进行预处理, 然后利用语音信号波形本身的信息, 进行波形之间的匹配估计, 从而获得准确的语音信号基音周期. 仿真实验表明: 与传统的基音周期提取方法相比较, 所提方法在基音周期提取的准确度方面有很大的提高.

关键词

基音周期; 降噪; 波形信息; 基音检测

中图分类号 TP391.42

文献标志码 A

收稿日期 2011-03-29

资助项目 江苏省科技创新与成果转化专项 (BE2008139); 江苏省“六大人才高峰”项目资助; 公益性行业科研专项 (GYHY200806014)

作者简介

俞翠华, 女, 硕士生, 主要研究方向为信号处理. ych-yucuihua@163.com

行鸿彦 (通讯作者), 男, 博士, 教授, 博导, 主要研究方向为生物医学信号、超声信号检测与处理等. xinghy@nuist.edu.cn

¹ 南京信息工程大学 电子与信息工程学院, 南京 210044

0 引言

在数字领域, 无论是语音分析与合成、语音压缩编码, 还是语音识别和说话者身份确认等, 都必须以准确可靠地检测基音周期为前提, 它严重影响着整个系统的性能. 传统的基音提取的方法大体上可以分为 3 类: 时域法、频域法和时频域法^[1]. 时域法主要包括短时自相关函数法、平均幅度差函数法等; 频域法包括倒谱法、简单逆滤波器跟踪法等; 时频域法包括小波函数法等多种方法^[2]. 这些方法都是通过利用语音信号的幅值信息或者频率信息来进行基音周期的提取的, 并没有充分利用信号本身的一些形状特征. 实际语音在产生的过程中, 不可避免地要受到背景噪声的影响, 因此, 在充分研究了语音信号波形信息的基础上, 本文采用结合经验模态分解^[3] (Empirical Mode Decomposition, EMD) 软阈值的去噪功能和语音信号时域波形^[4]相结合的语音信号基音周期检测算法, 并利用实验来验证所提方法的有效性.

本文将利用 EMD 软阈值法^[5]进行降噪处理, 抑制信号中的背景噪声成分; 然后对其进行中心削波预处理形成正负脉冲信号; 最后找出各个脉冲信号峰值的特征值, 按照本文规定的匹配处理方法进行峰与峰之间的匹配, 在得到初步的基音周期后, 对其进行中值滤波平滑, 从而得出最终的基音周期.

1 时域波形匹配法提取基音周期的原理

通常情况下, 语音信号总是以时域波形来描述和记录的, 通过人眼就可以大致地判断出语音信号的基音周期. 在这个估计的过程中, 主要是通过综合分析语音信号的波形特征来进行的, 而不是单单运用了语音信号的幅值信息. 文献 [5] 指出, 首先要对语音信号进行相关的预处理 (平滑、削波), 得到一系列基本正负脉冲信号, 并将处理过的信号作为新的被分析信号, 用波峰的正负关系 x_{i1} 、波峰的最大幅值 x_{i2} 、波峰的面积 x_{i3} 、波峰的宽度 x_{i4} 和波峰的位置 x_{i5} 作为各个波峰的特性参数, 并通过将各个波峰的前 4 个参数进行匹配对比, 找出间隔最小的相邻的类似波峰.

对于 2 个相匹配的波峰必须满足如下 2 个条件:

- 1) $x_{i1} = x_{j2}$;
- 2) $|x_{in} - x_{jn}| / |x_{in} + x_{jn}| < \beta, n = 2, 3, 4.$

条件 2) 中 β 是预先给定的门限阈值,用来筛选出最匹配的 2 个波峰。

如果 2 个波峰满足以上 2 个条件,则认为这 2 个波峰是匹配的,或者说是相似的,并且要求第 i 个波峰与第 j 个波峰之间的时间间隔是所有能匹配的波峰中最小的,此时 $x_{i5} - x_{j5}$ 则为相应的基音周期。

2 基于 EMD 软阈值降噪法和时域波形相结合的噪声背景下语音基音周期的提取方法

语音信号大体分为浊音和清音 2 类,浊音信号由声带振动在声门处产生的准周期脉冲序列激励声道而产生,它携带着语音中的大部分能量。人类语音的浊音段呈明显的准周期性,而清音段语音信号则类似于杂乱无章的白噪声^[1]。人在发浊音时,声门处于开合状态,声门闭合在语音信号中表现为突变,可以通过定位突变时脉冲信号的瞬时时间,得到 2 个相邻的时间间隔来确定基音周期。

为了在噪声情况下精确检测基音周期,可将 EMD 软阈值的降噪特性与基于波形特征的基音提取法相结合。首先利用 EMD 软阈值降噪法进行降噪处理,抑制信号中的背景噪声成分,使其接近于真实信号;将 EMD 软阈值降噪后的结果作为被处理信号,做中心削波处理,然后对其进行特征值提取,按照本文的处理方法,对各个波峰进行匹配,得出初期的基音周期;最后用中值平滑算法进行后处理,得到最终的基音周期。

2.1 EMD 软阈值降噪法的实现过程

EMD 分解方法是一种根据信号的极值特征尺度为度量的筛分过程^[3]。它依据信号本身的局部特征信息进行自适应地分解,得到一系列包含了从高到低不同频率成分,甚至是不等频带宽度的固有模式函数(Intrinsic Mode Function, IMF)分量 $c_i(t)$, 这些频率成分和带宽是随着信号的变化而变化的,所以这个过程也体现了多分辨率的滤波过程^[6]。因此有人提出了基于 EMD 算法的时空滤波方法^[7],即直接去掉包含噪声的 IMF 分量,将剩余的分量进行重构实现滤波,这是一个非常粗糙的降噪方法。当信号和噪声在各个 IMF 分量上混叠时,利用它进行降噪,会将一些有用的信号与噪声一起删掉,导致信号的严重失真。因此,本文结合软阈值降噪方法的特点,对由原始含噪信号分解出来的各个的 IMF 分量进行软阈值降噪,算法步骤为:

1) 对原始语音信号进行 EMD 分解,进而得到相应的各个 IMF 分量;

2) 对各个 IMF 分量使用 Stein 无偏似然估计准则进行阈值估计;

3) 选择合适的阈值函数^[8],其形式为

$$c_i'(t) = \begin{cases} \text{sgn}(c_i(t))(|c_i(t)| - \text{thr}(i)), & |c_i(t)| > \text{thr}(i), \\ 0, & |c_i(t)| < \text{thr}(i). \end{cases} \quad (1)$$

式(1)中 $c_i(t)$ 为降噪前的各 IMF 分量, $\text{thr}(i)$ 是第 i 层由阈值确定准则得出的阈值估计值,结合各个分量上的阈值估计值进行降噪计算,对各个降噪后的 IMF 分量进行重构,即可得到所期望的降噪信号 $x'(t)$:

$$x'(t) = \sum_{i=1}^n c_i'(t) + r_n'(t). \quad (2)$$

式(2)中 $x'(t)$ 为降噪后的信号, $c_i'(t)$ 表示经过降噪处理后的所有 IMF 分量, $r_n'(t)$ 表示经过降噪处理后残余分量。

2.2 中心削波法

中心削波处理主要用来消除共振峰的影响。通过利用中心削波函数 $c(t)$ 作用于输入函数 $x(t)$ 上,输出函数为 $y(t)$,实现减少共振峰对基音周期提取的影响。削波函数^[9]为

$$y(t) = c(t) = \begin{cases} x(t) - l, & x(t) > l, \\ 0, & |x(t)| < l, \\ x(t) + l, & x(t) < -l. \end{cases} \quad (3)$$

式(3)中 l 是由信号的最大幅度来决定的阈值门限,它是一个固定值,过大或过小都是不可取的,因为语音的基音信号是准周期的,每个周期上的最大值都不一样。如果 l 过大,有可能会削去信号的有用成分,使得基音周期估计结果为其实际值的双倍值,即半频现象;如果太小,会使得共振峰的抑制效果不明显,不利于实际应用。根据实验仿真,一般情况下,取最大幅度的 45%~65% 效果较好。

2.3 含噪语音基音提取算法的实现

首先对原始信号进行 EMD 软阈值降噪处理,滤除背景噪声;然后对其进行中心削波处理;最后通过匹配各个信号波峰信息来提取所处理信号的基音周期。下面给出了本文所提算法的具体的详细流程:

1) 用式(1)对原始信号进行降噪处理,并对降噪后的信号进行分帧处理;

2) 通过对每帧信号计算其短时能量和过零率 2 个参数,做清浊音判决,置清音部分的周期为 0;

- 3) 对得到的每帧浊音信号进行中心削波预处理;
- 4) 提取所得信号各个波峰的 5 个特征信息;
- 5) 根据匹配原则进行匹配 2 个相互匹配且最接近的波峰之间的间隔为本文所求的基音周期;
- 6) 通过计算时间间隔, 就可以求出初步的原始语音信号的基音周期;
- 7) 对所求得各个基音周期进行中值平滑处理, 去除一些误判的“野点”, 从而确定原语音信号的基音周期.

3 实验结果及分析

本文采用安徽科大讯飞信息科技股份有限公司的中文语音库为基音周期提取实验研究的语音信号源. 实验所用的原始语音是男声语音“普通话”, 采样频率为 8 kHz, 并以 16 位的形式进行存储, 波形如图 1 所示, 利用 Matlab7.1 进行信号处理和实验仿真.

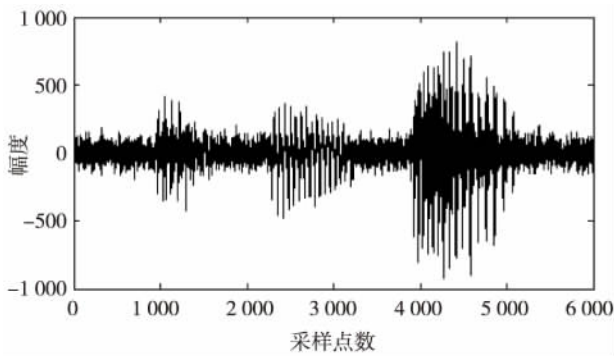


图 1 原始语音信号波形
Fig. 1 The wave of the clean speech signal

为了验证文中所提算法的有效性, 在原始语音中加入高斯白噪声, 形成信噪比 (Signal-Noise Rate, SNR) 分别为 5、0、-5 dB 的实验语音. 将本文所提算法分别与短时自相关算法 (200 个采样点为一帧) 和小波变换与时域波形相结合的算法^[3] 进行算法的准确性、抗噪性比较. 本文算法中, EMD 软阈值的函数如式 (1) 所示, 中心削波的阈值取语音信号最大幅值的 52%.

图 2 为人工标记得到的基音频率, 并以此作为参照标准值.

定义检测正确率为

$$\vartheta = \frac{n_1}{n_2} \times 100\% ,$$

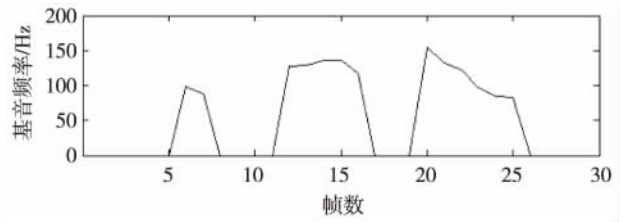


图 2 手工标记(标准)的基音频率

Fig. 2 The fundamental frequency of clean speech signal

其中 n_1 为正确帧数, n_2 为处理帧数. 当实际计算出的基音周期与参照基音周期相差超过 10 Hz 的则认为检测错误. ϑ 越大, 表示精确度越高, 算法性能越好.

对于在各个信噪比环境下, 3 种算法对基音周期检测的结果如图 3—5 所示, 表 1 给出内容为“普通话”的语音段的 3 种算法检测基音频率时的正确率.

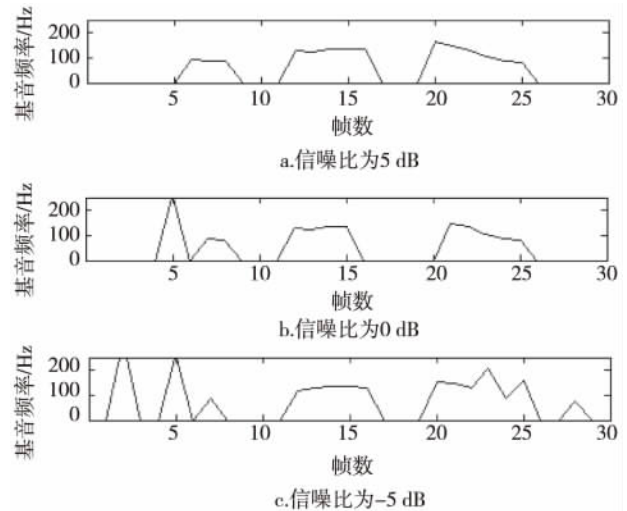


图 3 短时自相关法的基音频率

Fig. 3 The fundamental frequency of the traditional autocorrelation function method

表 1 3 种算法的正确率

Table 1 Accuracy rates of the three methods

信噪比/dB	正确率/%		
	自相关方法	小波结合时域波形算法	本文方法
5	83.3	90.0	93.3
0	73.3	80.0	86.6
-5	46.6	70.0	83.3

从表 1 可以看出: 在各个信噪比下, 本文方法都比其他方法具有更高的正确率.

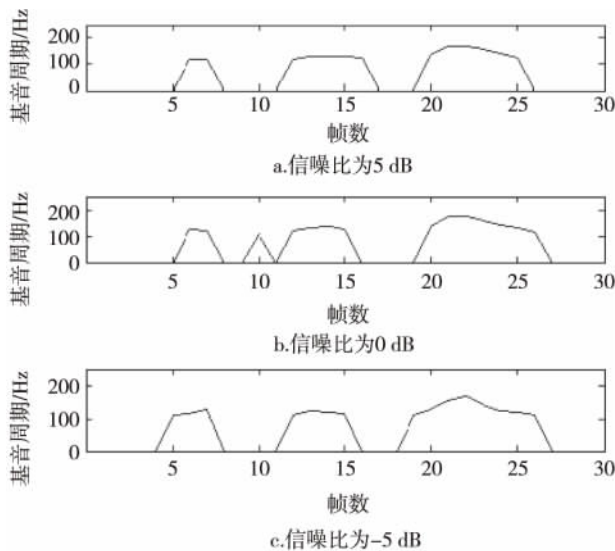


图4 小波变换与时域波形结合算法检测到的基音频率

Fig. 4 The fundamental frequency of the wavelet transform and signal time waveform method

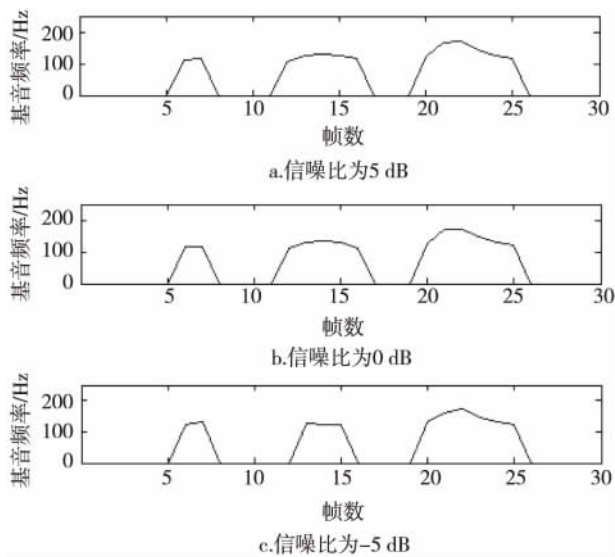


图5 本文算法的基音频率

Fig. 5 The experiment results of the proposed method

由图3可以看出:使用传统的短时自相关方法估计语音信号的基音频率时,由于共振峰和背景噪声的影响,半频和清浊音误判率随着信噪比的降低而明显上升。

图4表明:相对于传统自相关法,由于小波变换与时域波形结合的算法本身带有小波降噪功能,所以它的错误率有了一定的降低,但是在信噪比较低的情况下,它的性能还是会受到较大的影响,因为小波去噪并不能将背景噪声很好地去除,从而影响信号的波形信息。

从图5可以看出:本文算法采用的EMD软阈值降噪法,它在低信噪比的情况下,能够有效地去除背景噪声,因此在低信噪比环境下,其基音频率变化与人工标记的结果基本是吻合的,仍然能够较准确地检测出语音信号的基音周期,性能并没有随信噪比的降低而受到明显的影响,使得估算的基音周期准确性较高。

4 小结

本文采用的EMD分解方法,基函数直接从信号本身自适应的产生,不受人为因素影响,其分解是客观的,具有现实意义的。通过EMD软阈值降噪方法,对基音信号进行背景噪声处理,该滤波方法在滤除信号噪声的同时,不会改变信号的非线性与非平稳性,而且在低信噪比下具有良好的稳定性。中心削波削弱了共振峰对基音周期提取的影响,信号本身的波形特征匹配也是很好的基音提取技术。根据所提的匹配处理方法进行搜索处理,找出相互匹配的波峰,计算相应的时间间隔点,使得基音周期具有更高的正确率。

参考文献

References

[1] 胡瑛. 低信噪比下基音检测算法研究[D]. 长沙: 中南大学信息科学与工程学院 2007
HU Ying. Pitch detection algorithm in low signal to noise ratio [D]. Changsha: School of Information Science and Engineering, Central South University 2007

[2] 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
ZHENG Jiming, WANG Jingsong. Method for voice pitch period detection [J]. Computer Engineering, 2010, 36(10): 273-275

[3] 于德介, 程军圣, 杨宇. 机械故障诊断的 Hilbert-Huang 变换方法[M]. 北京: 科学出版社 2006: 24-46
YU Dejie, CHENG Junsheng, YANG Yu. The Hilbert-Huang transform in fault diagnosis of machinery [M]. Beijing: Science Press 2006: 24-46

[4] 陈小利, 徐金甫. 基于小波变换和时域波形的基音检测[J]. 现代电子技术 2011, 34(1): 77-79
CHEN Xiaoli, XU Jinfu. Pitch detection algorithm based on wavelet transform and waveform in time domain [J]. Modern Electronics Technique 2011, 34(1): 77-79

[5] Flandrin P, Rilling G, Goncalves P. Empirical mode decomposition as a filter bank [J]. IEEE Signal Processing Letters 2004, 11(2): 112-114

[6] 谭善文, 秦树人, 汤宝平. Hilbert-Huang 变换的滤波特性及其应用[J]. 重庆大学学报 2004, 27(2): 9-12
TAN Shanwen, QIN Shuren, TANG Baoping. The filtering character of Hilbert-Huang transform and its application

- [J]. Journal of Chongqing University 2004 27(2) :9-12
- [7] 刘维巍,张兴周,李春阳.基于 EMD 的基音检测预处理技术[J].应用科技 2010 37(7) :56-59
LIU Weiwei ,ZHANG Xingzhou ,LI Chunyang. Technology of pitch detection pretreatment based on EMD [J]. Applied Science and Technology 2010 37(7) :56-59
- [8] Donoho D L. De-noising by soft-thresholding [J]. IEEE Transaction on Information Theory ,1995 ,41 (3) : 613-627
- [9] 宋建华,朱勇,徐军,等.基于 Matlab 的一种基音周期检测算法[J].信息技术 2009(1) :50-54
SONG Jianhua ,ZHU Yong ,XU Jun ,et al. An algorithm for pitch detection based on Matlab [J]. Information Technology 2009(1) :50-54

Pitch detection algorithm for noisy speech

YU Cuihua¹ XING Hongyan¹ LI Peng¹

¹ School of Electronics and Information Engineering ,Nanjing University of Information Science & Technology ,Nanjing 210044

Abstract The background noise has a great influence on the accurate extraction of pitch period. For the EMD soft-threshold de-noising method can be effective against the background noise ,one novel method is proposed for pitch period extraction of noisy speech. Combined with the EMD soft-threshold de-noising method and center clipping ,this method preprocesses the speech signal ,then employs the signal waveform in time donation for matching estimation , in order to acquire the accurate pitch period. The implementing simulation results demonstrate that ,the proposed method has a higher accuracy compared with traditional methods.

Key words pitch period; de-noising; signal waveform; pitch extraction