

基于 EMD 的 AMDF 基音检测改进算法

宗 源,曾毓敏,孙永熙,郑 瑞

(南京师范大学物理科学与技术学院,江苏 南京 210023)

[摘要] 提出了一种基于经验模式分解的平均幅度差函数基音检测改进算法. 该算法首先求取每帧浊音语音的 AMDF;而后利用 EMD 对 AMDF 进行分解,得到有限个本征模式函数和残余分量,利用所有的 IMF 重构组成新的 EMDAMDF;最后利用 EMDAMDF 检测出该语音帧的基音. 仿真结果表明,由于能够有效地去除传统 AMDF 的均值下降趋势,该算法的性能明显优于传统的 AMDF 和 CAMDF.

[关键词] 基音,平均幅度差函数,经验模式分解,本征模式函数

[中图分类号] TN912 [文献标志码] A [文章编号] 1672-1292(2013)01-0062-06

A Modified AMDF Pitch Detection Algorithm Based on EMD

Zong Yuan, Zeng Yumin, Sun Yongxi, Zheng Rui

(School of Physics and Technology, Nanjing Normal University, Nanjing 210023, China)

Abstract: This paper presents a modified Average Magnitude Difference Function (AMDF) algorithm based on Empirical Mode Decomposition (EMD) for pitch detection. Firstly, AMDF of the voiced speech frame is computed. Then AMDF is decomposed into a residue component and a finite set of band-limited signals termed as Intrinsic Mode Functions (IMFs) by using EMD. A modified AMDF called EMDAMDF is reconstructed by all IMFs. Finally, the true pitch can be detected from EMDAMDF. The simulated results show that the performance of the proposed algorithm is much better than that of the conventional AMDF and CAMDF as the falling trend is eliminated effectively.

Key words: pitch, average magnitude difference function (AMDF), empirical mode decomposition (EMD), intrinsic mode functions (IMFs)

基音(基音频率, Pitch)是语音信号处理领域中的重要参数之一,它广泛地应用于语音编码、语音合成、语音增强以及语音识别等方面. 因此,准确地检测基音具有非常重要的意义. 目前使用比较广泛的经典方法有自相关函数法^[1]、平均幅度差函数法(Average Magnitude Difference Function, AMDF)^[2]、倒谱法^[3]、小波变换法^[4]等以及在它们的基础上相互结合和改进的方法^[5-9]. 其中 AMDF 以其算法复杂度小、精度高的优点得到了广泛的应用. 但是,当浊音语音信号的周期性和平稳性不太好时,传统的 AMDF 常常会出现除零点外最低谷值点并不是实际基音点,从而检测结果为实际基音的 N 倍或者 $1/N$ 的情况. 出现这种情况是因为随着延迟增加,用于计算 AMDF 的语音帧的重叠部分减少,从而导致 AMDF 的值越来越小,这在 AMDF 的波形上体现为一种均值下降趋势^[10].

为了克服这种下降趋势,研究人员提出了许多 AMDF 的改进算法,其中比较具有代表性的有: Muhammad 提出的 EAMDF^[11],该方法将每个用于计算 AMDF 的语音帧前后各补充 $1/2$ 帧长的采样点,通过增加计算 AMDF 语音帧的长度达到去除均值下降趋势的效果;张文耀提出 CAMDF^[10],改变了传统 AMDF 的计算方法,以采样点索引对帧长的取模运算克服计算 AMDF 时随着延迟增加而语音帧重叠部分的减少,因此很好地修正了均值下降趋势,具有较好的检测效果. 这些改进方法虽然克服了均值下降趋势,但由于改变了 AMDF 的计算方法或 AMDF 的语音帧的帧长,并未从根本上解决 AMDF 的均值下降趋势所带来的基音点误判的问题,因此依然会产生错误.

收稿日期:2012-12-09.
基金项目:江苏省自然科学基金(BK2010546).
通讯联系人:曾毓敏,博士,教授,研究方向:语音信号处理. E-mail: zengyumin@njnu.edu.cn

针对 AMDF 存在的这种问题,本文提出一种基于经验模式分解(Empirical Mode Decomposition, EMD)的 AMDF 基音检测改进算法,该算法对每帧浊音语音的 AMDF 进行 EMD 处理,而后去除表示 AMDF 均值下降趋势的残余分量,通过分解得到的各个 IMF 重构组成新的改进 AMDF(本文定义为 EMDAMDF),最后利用 EMDAMDF 检测出该浊音语音帧的基音频率. 仿真结果表明,与传统 AMDF 法和 CAMDF 法相比,本文方法由于有效地去除了 AMDF 的均值下降趋势,其检测效果明显优于传统 AMDF;对于性能优良的 CAMDF 法,其检测效果也有了显著的提升.

1 经验模式分解

EMD 是 Norden E Huang 等人于 1998 年提出的一种新的信号分解方法^[12]. EMD 的本质是通过信号本身的特征尺度来将信号分解成本征模式函数(Intrinsic Mode Function, IMF),它一方面具有小波变换多分辨率的优点,另一方面又克服了小波变换中选择合适的基的困难. 因此,EMD 可有效地处理非平稳信号,且具有良好的自适应性.

IMF 分量必须满足两个条件:整个信号中极值点的个数与过零点的个数相等或最多相差 1;信号上的任意点处,由所有局部极大值点确定的上包络和所有局部极小值所确定的下包络的均值为零,即上下包络线对称于零均线.

EMD 的具体算法^[13]为:

- (1) 令 $g_1(t) = s(t)$ ($s(t)$ 为待分解信号);
- (2) 找出 $g_1(t)$ 所有的极值点(包括极大值和极小值);
- (3) 利用三次样条插值分别将所有极大值点和极小值点拟合成上包络 $h(t)$ 和下包络 $l(t)$;
- (4) 利用上包络和下包络算出局部均值:

$$u_1(t) = [h(t) + l(t)] / 2; \quad (1)$$

- (5) 从 $g_1(t)$ 中减去 $u_1(t)$: $g(t) = g(t) - u_1(t)$;
- (6) 根据上文所提的 IMF 必须满足的两个条件判断 $g_1(t)$ 是否为 IMF;
- (7) 重复(2) ~ (6) 直到 $g_1(t)$ 满足 IMF 的条件.

$C_1(t) = g_1(t)$ 即为第一个 IMF, 记为 IMF1, 利用上述算法对残余分量

$$r_1(t) = s(t) - C_1(t) \quad (2)$$

进行处理, 提取出第二个 IMF, 如此循环, 直至最后一个 IMF 即 $C_n(t)$ 被提取出来. 最后的残余分量 $r_N(t)$ 通常为一常数或者趋势项且不可以再分解. 各个 IMF 和残余分量可以重构信号 $s(t)$:

$$s(t) = \sum_{n=1}^N C_n(t) + r_N(t). \quad (3)$$

EMD 将信号分解成若干个频率从高至低的 IMF, 整个过程体现了多尺度的自适应滤波特性. 根据这一特点, 可以根据信号的要求, 有效地对某一频率范围内的信号进行处理^[14]. 此外, 由于 EMD 是根据信号的局部时间尺度进行分解的, 其基函数的选择来自于信号本身, 因此减少了主观因素的影响.

2 基于 EMD 的 AMDF 改进算法

语音信号的短时平均幅度差函数 $D(\tau)$ 定义^[2]为:

$$D(\tau) = \sum_{n=0}^{N-\tau-1} |x(n) - x(n+\tau)|, \quad (4)$$

式中, $x(n)$ 为加上窗长为 N 的矩形窗的浊音语音帧, τ 为延迟.

$D(\tau)$ 拥有与浊音语音周期相一致的周期特性, 并在 $\tau = F_s / F_p$ 的地方出现最低谷值(F_p 为基音频率, F_s 为采样频率), 因此通过计算 $D(\tau)$ 找出除零点以外的最低谷值点就能够检测出浊音语音的基音频率. 但由于 $x(n)$ 为加窗过后的语音帧, 窗外的值全部为零, 随着延迟 τ 的增加, AMDF 不能完全显示出浊音语音帧的周期性的本质, 而呈现出均值下降趋势, 故而在实际应用中易出现表示真正的基音的点并不是最小谷值.

通过对 EMD 的分析可知, 信号经过 EMD 处理后, 最后一项残余分量为常数项或趋势项, 因此可利用 EMD 去除 AMDF 的均值下降趋势. 图 1 为一浊音语音帧的 AMDF 经过 EMD 处理之后的分解结果, 通过观

察可以发现残余分量的下降趋势正是 AMDF 的均值下降趋势. 故本文将残余分量滤除, 利用各个 IMF 重构组成一个新的 AMDF(本文定义为 EMDAMDF)即可自适应地去除传统 AMDF 的均值下降趋势. AMDF 的经验模式分解可表示为:

$$s_{\text{AMDF}}(t)=\sum_{n=1}^N c_n(t)+r_N(t),$$

(5)

式中, $c_n(t)$ 为 IMF, $r_N(t)$ 为残余分量. 则 EMDAMDF 可表示为:

$$s_{\text{EMDAMDF}}(t)=\sum_{n=1}^N c_n(t).$$

(6)

图 2 为一浊音语音帧的 AMDF、CAMDF 和 EMDAMDF 的仿真图. 由图 2 中 3 种方法的对比可以发现, AMDF 由于均值下降趋势, 表示基音频率的并不是全局最低谷值点, 不能有效地进行基音检测. CAMDF 虽然克服了均值下降趋势, 但由于其他谷值点也受到了影响, 故全局最低点出现在第一个谷值点, 造成倍频的错误检测结果, 也不能准确地检测出基音. 本文所提出的 EMDAMDF 由于去除了表示均值下降趋势的残余分量, 使 AMDF 的均值下降趋势得到了抑制, 在波形上表现为“AMDF 整体被拉平”, 可以看到表示基音的谷值点更加的突出, 而其他的谷值点相对于整体均值并没有改变, 明显高于基音谷值点, 因此, EMDAMDF 能够准确地检测出该语音帧的基音频率.

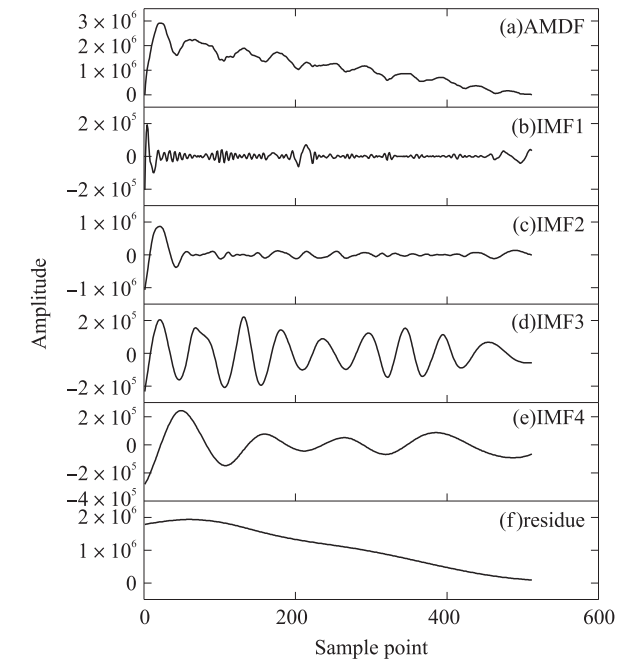


图 1 AMDF 的经验模式分解

Fig.1 The decomposition of AMDF by using EMD

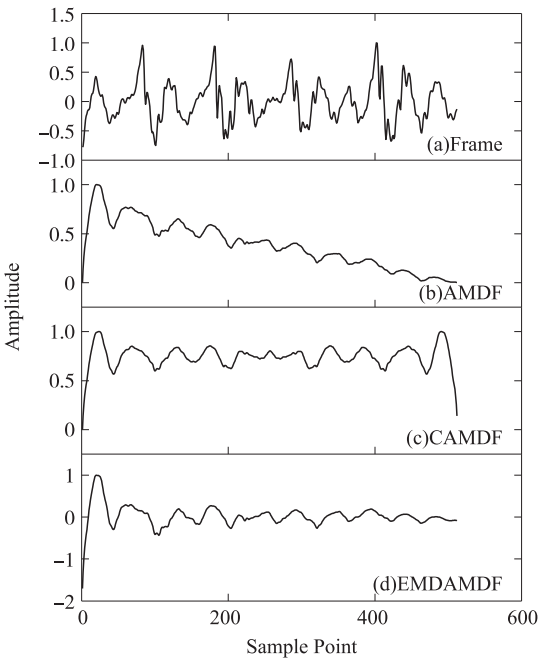


图 2 一帧浊音语音的 AMDF、CAMDF 和 EMDAMDF

Fig.2 AMDF、CAMDF and EMDAMDF of a voiced frame

基于此, 本文提出了如图 3 所示的基于 EMDAMDF 基音检测算法. 首先对语音信号分帧加窗; 然后对每帧语音进行清浊音判断, 清(静)音帧的基音频率判为 0, 同时对浊音帧求 AMDF; 再对 AMDF 进行 EMD 处理, 滤除最后的残余分量, 通过各个 IMF 重构组成 EMDAMDF; 最后通过 EMDAMDF 的除零点以外的最低谷值点检测出浊音帧的基音频率. 其中基音判定部分设定 50 ~ 500 Hz 为有效基音, 若检测结果小于 50 Hz 或大于 500 Hz, 则基音判为 0.

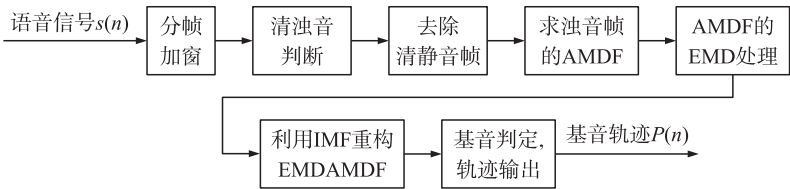


图 3 基于 EMDAMDF 的基音检测算法

Fig.3 A pitch detection algorithm based on EMDAMDF

3 仿真实验

实验所用语音为实验室录制的一名成年男子朗读“树上的落叶掉光了”的纯净语音,语音的采样率为 11 025 Hz,采样精度为 16 bits. 实验中语音分帧帧长设为 40 ms,帧移为 20 ms. 图 4 为分别使用 AMDF、CAMDF 和本文算法对该段语音进行基音检测的结果.

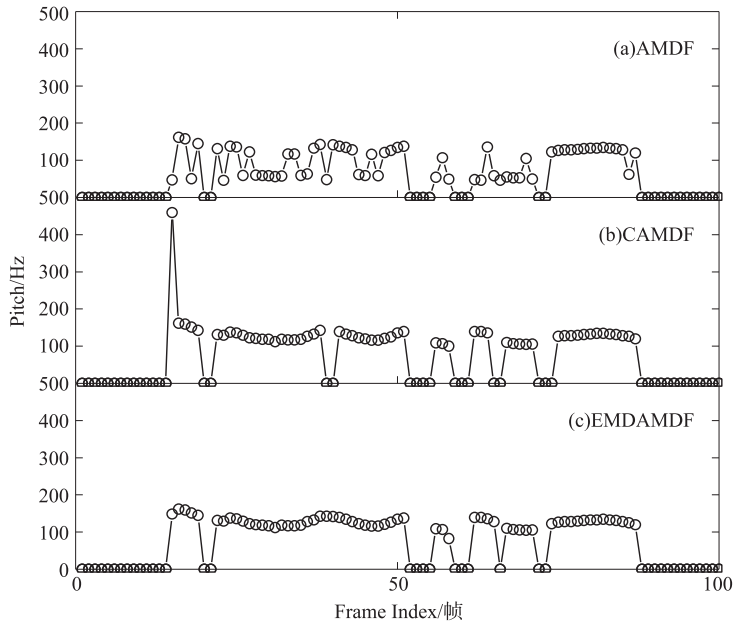


图 4 AMDF、CAMDF、EMDAMDF 检测一段语音的基音
Fig.4 Pitch detection of a speech signal by AMDF ,CAMDF and EMDAMDF

图 4(a)为 AMDF 的检测结果,从其基音曲线可以看出,AMDF 法在多处出现明显的半频错误,检测效果非常差,这是由于 AMDF 的均值下降趋势导致实际基音点之后的谷值点被误判为真实的基音点. 图 4 (b)为 CAMDF 的检测结果,它的基音曲线整体看上去光滑平稳,但两三处出现倍频错误或检测不出基音,说明 CAMDF 虽然很好地修正了均值下降趋势,但却有可能带来其他的检测错误. 图 4(c)为 EMDAMDF 的检测结果,它的基音曲线光滑规整,没有明显的倍频和半频错误,说明 AMDF 通过 EMD 处理滤除趋势项后,原有的检测错误从根本上得到了解决,因此,EMDAMDF 能够准确检测出浊音语音的基音频率.

为进一步评价本文所述算法,本文使用基尔基音检测参考数据库(Keele Pitch Extraction Reference Database)^[15]来评价各个算法的性能. 基尔基音数据库中所有语音均以 20 000 Hz 采样,16 bits 量化,数据库中提供以 512 点为帧长、100 点为帧移的所有浊音帧的参考基音信息. 使用 AMDF、CAMDF 和 EMDAMDF 对数据库中的 3 位男性(编号:M2 ~ M4)和 3 位女性(编号:F1 ~ F3)的语音进行基音检测,比较其检测误差. 检测结果与数据库中提供的参考基音的相对误差大于或等于 1 ms(以基音周期计算)则定义为基音粗差(Gross Pitch Error, GPE)^[16],本文以 % GPE 作为比较参量比较各方法.

表 1 不同信噪比下 3 种算法检测结果的比较(女性)

Table 1 Comparison of three algorithms at different SNR(The results are on Female)

| | | Clean | 10 dB | 5 dB | 0 dB | -5 dB | -10 dB |
|----|---------|-------|-------|-------|-------|-------|--------|
| F1 | AMDF | 22.66 | 34.23 | 45.00 | 57.22 | 71.85 | 84.65 |
| | CAMDF | 9.34 | 11.82 | 14.50 | 20.77 | 31.42 | 49.97 |
| | EMDAMDF | 6.07 | 7.71 | 10.25 | 11.43 | 16.53 | 27.89 |
| F2 | AMDF | 11.93 | 27.61 | 40.12 | 60.04 | 76.34 | 88.85 |
| | CAMDF | 5.73 | 8.78 | 12.46 | 17.67 | 29.39 | 47.58 |
| | EMDAMDF | 3.84 | 6.15 | 8.73 | 11.87 | 17.98 | 33.07 |
| F3 | AMDF | 13.11 | 24.11 | 35.96 | 50.93 | 69.34 | 82.65 |
| | CAMDF | 7.75 | 11.26 | 14.24 | 17.88 | 28.34 | 47.48 |
| | EMDAMDF | 4.63 | 6.56 | 9.47 | 11.58 | 17.88 | 27.02 |

表 2 不同信噪比下 3 种算法检测结果的比较(男性)
Table 2 Comparison of three algorithms at different SNR(The results are on Male)

| | | Clean | 10dB | 5dB | 0dB | -5 dB | -10 dB |
|----|---------|-------|-------|-------|-------|-------|--------|
| M2 | AMDF | 9.92 | 19.51 | 29.02 | 42.36 | 59.00 | 75.29 |
| | CAMDF | 7.32 | 10.06 | 11.29 | 17.11 | 30.87 | 55.37 |
| | EMDAMDF | 2.81 | 4.31 | 5.82 | 9.38 | 14.58 | 30.11 |
| M3 | AMDF | 21.31 | 23.15 | 26.91 | 33.62 | 45.81 | 60.22 |
| | CAMDF | 22.04 | 25.43 | 26.79 | 31.47 | 42.00 | 58.68 |
| | EMDAMDF | 9.10 | 10.04 | 11.76 | 15.76 | 25.49 | 41.63 |
| M4 | AMDF | 19.51 | 26.85 | 33.75 | 44.66 | 59.44 | 74.12 |
| | CAMDF | 17.87 | 19.65 | 23.23 | 30.66 | 43.46 | 59.92 |
| | EMDAMDF | 8.35 | 9.75 | 13.09 | 19.36 | 33.46 | 51.47 |

表 1 和表 2 分别给出了 3 位女性(F1 ~ F3)和 3 位男性(M2 ~ M4)各自的语音采用 AMDF、CAMDF 和本文所提出的 EMDAMDF 基音检测算法的% GPE. 可以发现,CAMDF 虽然很大程度上降低了 AMDF 的检测误差,但 CAMDF 修改了 AMDF 的定义,所以同时也会带来其他的检测错误,有时检测效果并不是非常好(如纯净语音情况下的 M3). 所以 CAMDF 并未从根本上解决 AMDF 的均值下降趋势所带来的检测问题. 而 EMDAMF 在 6 人的检测中均取得非常低的% GPE,这是因为 EMDAMDF 针对 AMDF 造成检测错误的原因,从根本上对 AMDF 进行了修正,即去除了引起 AMDF 误判的均值下降趋势,因此 EMDAMDF 相对于 AMDF 的性能有了很大的提升,同时又没有修改 AMDF 的定义引起其他未知的检测错误,所以 EMDAMDF 拥有较高的检测准确率. 图 5 所示为 3 种方法在低信噪比(SNR=0 dB,-5 dB,-10 dB)情况下平均检测性能的柱状图(即将 F1-F2-F3 和 M2-M3-M4 6 个小样本合为一个大样本进行统计),结合表 1 和表 2,从柱状图可以直观地看出,无论对于男性还是女性,纯净语音还是含噪语音,本文所提算法的检测效果均明显优于 AMDF;对于检测效果较好、使用广泛的 CAMDF,EMDAMDF 也依然有其一定的提升.

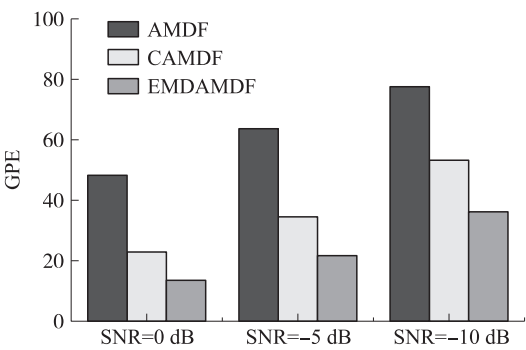


图 5 低信噪比下 3 种方法的平均性能对比图
Fig. 5 Average performance of three methods at low SNRs

4 结语

本文简要介绍了 EMD 的概念及其自适应多频分辨率的性质并且给出了 EMD 的分解算法;分析了 AMDF 基音检测算法并指出了其不足之处,即均值下降趋势引起其他谷值点被误判为基音点;接着针对 AMDF 的不足之处提出了基于 EMD 的改进算法;最后进行了仿真对比实验. 仿真实验结果表明,本文所提算法由于有效地去除了 AMDF 的均值下降趋势,新的 EMDAMDF 的基音检测性能得到了很大的提升,其性能明显优于传统 AMDF;相比应用广泛并且性能较好的 CAMDF 也有一定的提升.

[参考文献](References)

[1] Krubsack D A, Niederjohn R J. An autocerrelation pitch detector and voicing decision with confidence measures developed for noise corrupted speech[J]. IEEE Transactions on Signal Processing,1991,39(2) :319-329.

[2] Ross M, Shaffer H, Freudberg R, et al. Average magnitude difference function pitch extractor[J]. IEEE Transactions on Acoustics, Speech and Signal Processing,1974,22(5) :353-362.

[3] Ahmadi S, Spanias A S. Cepsrum-based pitch detection using a new statistical V/UV classification algorithm[J]. IEEE Transactions on Speech and Audio Processing,1999,7(3) :333-338.

[4] Kadame S, Broudreux-Bartels G F. Application of the wavelet transform for pitch detection of speech signals[J]. IEEE Transactions on Information Theory,1992,38(2) :917-924.

[5] Amado G. Pitch detection algorithms based on zero-cross rate and autocorrelation function for musical notes[C]//Proceedings of ICALIP. Shanghai:IEEE,2008:449-454.

- [6] Dong J. Pitch detection using circular average magnitude difference function based on wavelet transform[J]. International Journal of Innovative Computing, Information and Control, 2009, 5(9): 2717–2724.
- [7] Heckmann M. Combining rate and place information for robust pitch extraction[C]//Proceedings of Interspeech. Antwerp: ISCA, 2007: 1225–1228.
- [8] Lindh J. Robustness of long time measures of fundamental frequency[C]//Proceedings of Interspeech. Antwerp: ISCA, 2007: 2025–2028.
- [9] Pelle P. A robust pitch extraction system based on phase locked loops[C]//Proceedings of ICASSP. Toulouse: IEEE, 2006: 249–252.
- [10] Zhang W, Xu G, Wang Y. Pitch estimation based on circular AMDF[C]//Proceedings of ICASSP. Beijing: IEEE, 2002: 341–344.
- [11] Ghulam Muhammad. Noise robust pitch detection based on extended AMDF[C]//Proceedings of ISSPIT. Sarajevo: IEEE, 2008: 133–138.
- [12] Huang N E, Zheng S, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[C]//Proceedings of Royal Society A. London: Royal Society, 1998: 903–995.
- [13] Molla M, Khademul I, Hirose K, et al. Pitch estimation of noisy speech signals using empirical mode decomposition[C]//Proceedings of Interspeech. Antwerp: ISCA, 2007: 2117–2180.
- [14] 王婷. EMD 算法研究及其在信号去噪中的应用[D]. 哈尔滨: 哈尔滨工程大学信息与通信工程学院, 2010.
Wang Ting. Research on EMD algorithm and its application in signal denoising[D]. Harbin: College of Information and Communication Engineering, Harbin Engineering University, 2010. (in Chinese)
- [15] Rabiner L R, Cheng M J, McGonegal C A. A comparative performance study of several pitch detection algorithms[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, 24(5): 399–417.
- [16] Meyer G, Plante F, Ainsworth W A. A pitch extraction reference database[C]//Proceedings of EUROSPEECH. Madrid: ISCA, 1995: 827–840.

[责任编辑: 严海琳]