

# 基于经验模态分解和 Mel 倒谱系数的语音端点检测

陈蔚,熊卫华,施巍巍

(浙江理工大学机械与自动控制学院,杭州 310018)

**摘要:**提出了一种基于经验模态分解和 Mel 倒谱系数的语音端点检测方法。对语音信号进行分解得到一组 IMF 分量,将集中在低层 IMF 分量中的噪声信号滤除,重构剩余的 IMF 分量成语音信号,提取重构信号的 Mel 频率倒谱系数来检测语音信号的端点。实验结果表明,提出的方法可以较好地消除噪声对语音信号端点检测带来的影响,能在不同信噪比的环境下正确地对语音信号进行端点检测,具有一定的鲁棒性。

**关键词:**语音端点检测; EMD; IMF; MFCC; 信噪比

**中图分类号:** TP391.4 **文献标志码:** A

## 0 引言

在语音信号处理如语音增强、语音识别、语音合成、语音编码等过程中,都会涉及语音端点检测这个预处理步骤<sup>[1-2]</sup>,准确地检测语音端点对语音信号处理具有重要意义,也是语音信号处理中的一个热门研究方向。目前,常用的语音端点检测方法有短时能量、短时平均过零率、双门限法等<sup>[3]</sup>。上述方法均假设语音信号在短时间内是平稳的,而实际上,语音信号是一种非平稳非线性的信号,通常包含着噪声信号,所以上述方法处理语音信号的效果并不理想<sup>[4]</sup>。

经验模态分解(empirical mode decomposition, EMD)算法可以用于处理非线性非平稳的信号,而语音信号就是一种典型的非平稳非线性信号,因此可以将 EMD 变换用于语音信号处理。本文首先利用 EMD 分解原始的带噪语音信号,得到一组固有模态函数(intrinsic mode function, IMF)和一个剩余量,得到的 IMF 分量中包含了从高到低不同频率段的成分;然后根据 EMD 算法的滤波特性,取合适的 IMF 分量重构语音信号,得到的语音信号消去了部分噪声,对重构语音信号提取 Mel 频率倒谱系数(Mel frequency cepstrum coefficient, MFCC);最后根据此特征参数来判断语音端点。

## 1 经验模态分解和 Mel 频率倒谱系数

### 1.1 EMD 分解原理

EMD 分解是 1998 年由 Huang 等提出的分析和处理连续的非线性非平稳信号的方法<sup>[5-6]</sup>。EMD 方法将信号分解成一组 IMF 分量,IMF 分量的频率呈现逐渐减小的趋势,体现信号的局部特征。由于 IMF 分量是根据信号的特征时间尺度来获取的<sup>[7-8]</sup>,故其可以将信号的特征保留下来。与 Fourier 变换、小波变换等方法相比,EMD 方法分解所用的基本就是原始信号本身,故具有自适应性。每个 IMF 分量均需要满足 2 个约束条件:a)在整个信号中,信号的极值点个数与过零点个数最多相差一个;b)在每一个时间点上,由局部极大值定义的上包络线和局部极小值定义的下包络线的平均值必须为零,即信号的局部极值关于时间轴局部对称。

EMD 算法的具体步骤如下:

a) 找出原始离散信号  $S(t)$  所有的局部极大值和局部极小值点,并用三次样条函数拟合原始数据数列的上下包络线。

b) 计算上下包络线的均值,记为  $m_1(t)$ ;并将  $S(t)$  减去该均值,得到一个去掉低频的数据序列  $h_1(t)$ ,即

$$h_1(t) = S(t) - m_1(t) \quad (1)$$

c) 判断  $h_1(t)$  是否满足 IMF 条件, 如果不满足, 则对  $h_1(t)$  信号重复 a)、b) 两个过程, 直到均值趋近于零为止, 这样得到第一个 IMF 分量  $C_1(t)$ , 它代表  $S(t)$  中最高频率的分量。

d) 将  $S(t)$  与  $C_1(t)$  相减, 得到原始信号中不包含最高频率分量的剩余信号  $r_1(t)$ , 即:

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

e) 将  $r_1(t)$  作为原始信号重复上述 a)—d) 4 个过程, 得到其余 IMF 分量  $C_i(t)$ ,  $i=2, 3, \dots, n$ , 直到残余量  $r_n(t)$  为单调函数时停止。重复过程如下式(3)—式(5)所示:

$$r_2(t) = r_1(t) - C_2(t) \quad (3)$$

$$r_3(t) = r_2(t) - C_3(t) \quad (4)$$

$$r_n(t) = r_{n-1}(t) - C_n(t) \quad (5)$$

f) 经过有限次的分解之后, 可以得到原始信号所包含的不同频率分量的信号, 则原始信号  $S(t)$  可以用式(6)表示:

$$S(t) = \sum_{i=1}^n C_i(t) + r_n(t) \quad (6)$$

其中:  $r_n(t)$  为剩余量函数, 表示信号的趋势,  $C_i(t)$  为各个 IMF 分量, 其分别包含了信号从高到低不同频率段的成分。

## 1.2 Mel 频率倒谱系数

人耳对不同频率语音信号的感知能力是不同的, 小于 1 000 Hz 时, 感知能力与频率大致呈线性关系; 当大于 1 000 Hz 时, 感知能力与频率呈对数关系增长<sup>[9]</sup>。Davies 等<sup>[10]</sup>提出了 Mel 频率倒谱系数这个概念, 将频谱转换为基于 Mel 频标的非线性频谱, 再转换到频谱域中。MFCC 参数的计算以 Mel 频率为基准, 它和实际频率的转换关系可用式(7)表示。

$D(i) =$

$$\frac{1}{2N+1} \sqrt{(C(i+N) - C_{\text{MFCC-noise}})^2 + (C(i+N-1) - C_{\text{MFCC-noise}})^2 + \dots + (C(i) - C_{\text{MFCC-noise}})^2 + \dots + (C(i-N) - C_{\text{MFCC-noise}})^2} \quad (10)$$

其中,  $D(i)$  为第  $i$  帧的 MFCC 距离噪声信号的 MFCC 系数的距离。设  $T$  代表一个阈值, 由参考文献[9]中 MFCC 相关系数距离曲线可知, 取  $T=2$ 。当  $D(i)$  大于等于 2 且  $D(i+1)$  也大于等于 2 时, 就认为第  $i$  帧是一个语音信号。

## 3 仿真实验与结果

在仿真实验中, 使用 PC 机在安静的实验室来录制纯净的语音信号, 其中信号的采样率为 22 kHz, 量化级为 16bit, 每帧 128 个采样点, 帧移为 64 个采样点。图 1 为纯净语音信号“西湖”的时域波形图。

$$f_{\text{Mel}} = 2595 \lg \left( 1 + \frac{f}{700} \right) \quad (7)$$

其中:  $f$  表示实际频率,  $f_{\text{Mel}}$  表示 Mel 频率。

## 2 基于 EMD 和 MFCC 的语音端点检测的算法

基于 EMD 和 MFCC 的语音端点检测算法具体步骤如下:

a) 首先, 对语音信号用 EMD 方法进行分解, 得到一组 IMF 函数和一个剩余量。各阶 IMF 的能量密度和平均周期的乘积近视为常数, 关系如下公式所示:

$$E_n \bar{T}_n = \text{const} \quad (8)$$

其中:  $E_n$  为第  $n$  阶 IMF 能量密度,  $\bar{T}_n$  为第  $n$  阶 IMF 平均周期, const 为常数。

b) 其次根据二进滤波器的性质以及公式(8)可知: 白噪声的各阶 IMF 的能量密度按照 2 倍关系逐渐减小, 白噪声信号主要集中在前几层 IMF 中。通过大量实验发现, 白噪声信号主要集中在前面 1 阶 IMF 中, 故重构信号时去除前面 2 阶 IMF; 又由于语音信号主要集中在前面的 IMF 中, 后面的 IMF 包含的语音信号幅度趋于平坦, 包含的语音信号不断减小, 所以, 取 IMF3—IMF10 来重构语音信号。

c) 然后假定重构的语音信号的前 5 帧为噪声信号, 提取其的 MFCC 并且加权求平均值  $C_{\text{MFCC-noise}}$ 。其加权求平均值的计算公式如下:

$$C_{\text{MFCC-noise}} = \frac{1}{5} \sum_{i=1}^5 C(i) \quad (9)$$

d) 最后利用滑动平均求取每帧语音的  $C_{\text{MFCC}}$  离  $C_{\text{MFCC-noise}}$  的距离。设 MFCC 系数的序列为  $C_{\text{MFCC}}(i)$ ,  $i=1, 2, \dots, n$ , 则滑动平均的公式为:

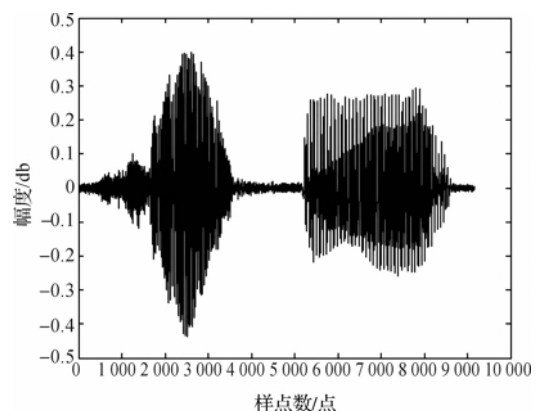


图 1 语音“西湖”的时域波形

首先,对纯净的语音信号进行加噪处理,噪声信号为高斯白噪声。然后利用 EMD 算法对加噪后的语音信号进行分解,根据本文提出的准则去选取 IMF 分量重构语音信号。图 2 为纯净语音信号“西湖”经过 EMD 分解后的前 10 阶 IMF 分量,图 3 为含噪语音信号“西

湖”经过 EMD 分解后的前 10 阶 IMF 分量。最后对重构后的语音信号提取 MFCC 系数进行端点检测。图 4 分别为纯净语音信号“西湖”时域波形端点检测结果和 MFCC 检测结果。在 SNR=20 的情况下,图 5 为语音“西湖”时域波形端点检测结果和 MFCC 检测结果。

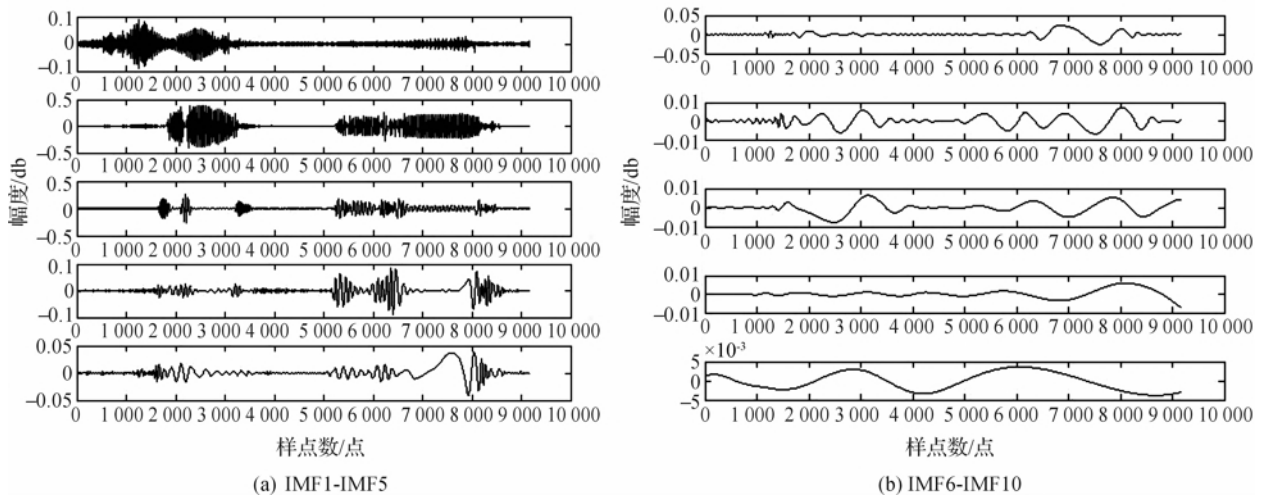


图 2 语音“西湖”的 IMF 分量

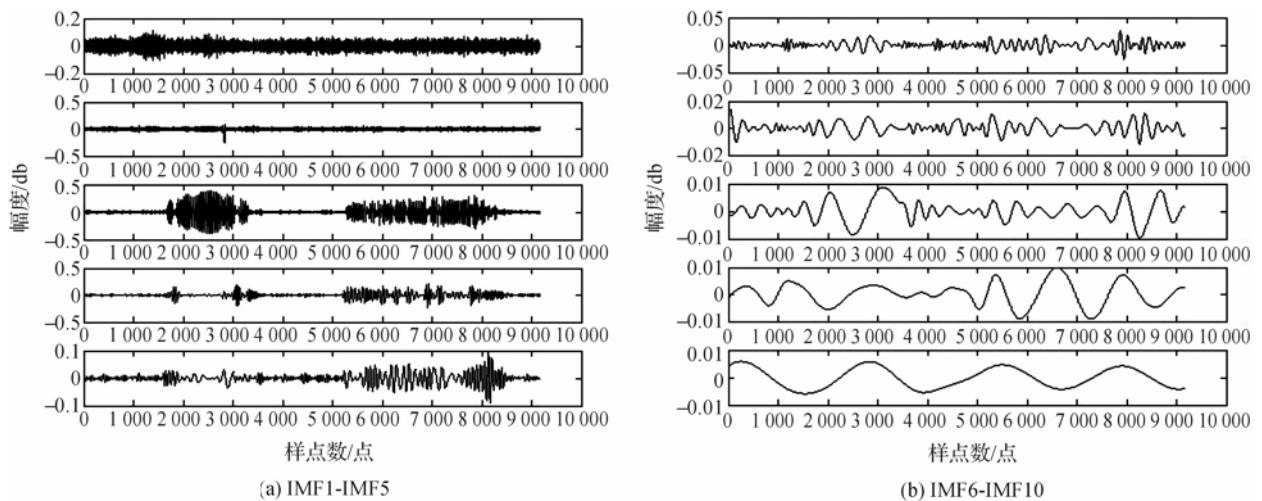


图 3 语音“西湖”加噪后的 IMF 分量

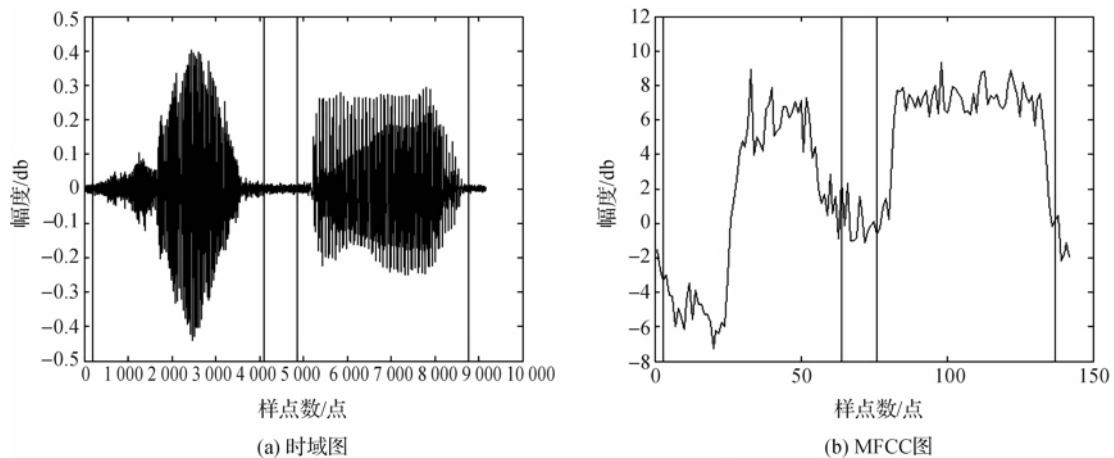


图 4 “西湖”端点检测

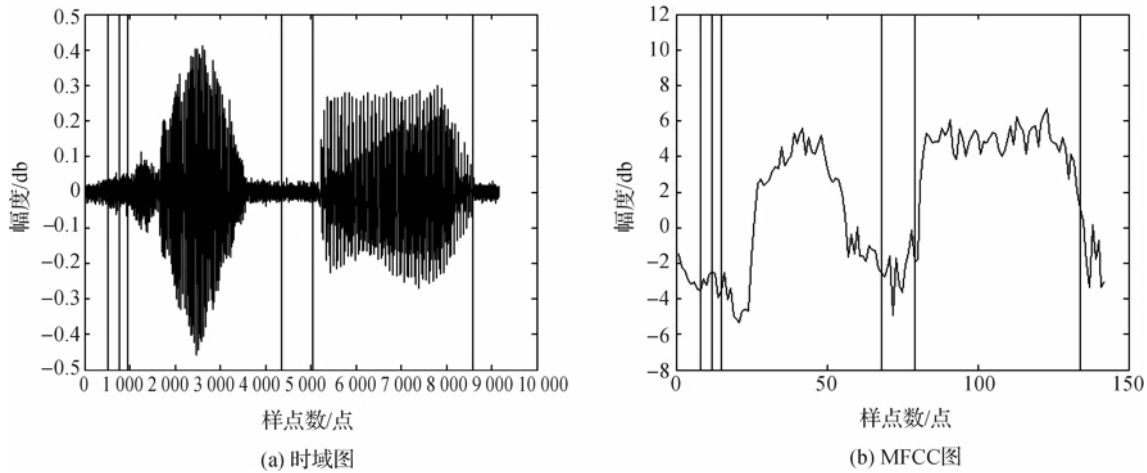


图 5 “西湖”端点检测(SNR=20)

从图 4 中可以看出,对于纯净的语音信号,本文提出的算法能正确地检测出语音信号的端点;从图 5 中可以看出,在加入高斯白噪声以后(其中信噪比为 20),本文提出的算法也能很好地检测出语音端点。本文对另外 5 个纯净语音信号和加噪语音信号分别进行检测,准确率如表 1 所示。

表 1 纯净语音信号与加噪语音信号 (SNR=20)的端点检测结果

纯净信号	准确率/%	加噪信号	准确率/%
杭州	95	杭州	90
宁波	96	宁波	89
衢州	94	衢州	92
金华	95	金华	89
绍兴	96	绍兴	90

从表 1 中可以看出,虽然检测结果存在一定的漏检、误检,但在低信噪比情况下,本算法依然能很好的检测出语音信号的端点,说明该算法对含噪信号具有自适应性。

## 4 结 论

本文研究了一种算法,该算法通过 EMD 分解获得一组 IMF 分量,利用 EMD 方法的滤波特性取合适的 IMF 分量重构语音信号,再对重构后的语音信号提取 Mel 频率倒谱系数进行端点检测。针对噪声对语音端点检测的影响,对 IMF 分量进行合理选取;根据噪声的 MFCC 系数,自适应的设置距离阈值,检测语音信号的端点。实验结果表明,本文提出的方法在低信噪比情况下,能很好地检测出语音端点,对噪声信号具有一定的鲁棒性。

## 参考文献:

- [1] 孙科伟,李建海,杨海东,等. 一种强噪声环境中的语音端点检测算法[J]. 计算机计算与应用, 2013, 39(12): 135-137.
- [2] 王 辉,袁淑丹. 基于 EMD 与功率谱熵的语音端点检测[J]. 电声技术, 2013, 37(11): 40-44.
- [3] 徐耀华,郭 英,王 刚. 基于负熵的语音端点检测算法[J]. 信号处理, 2009, 25(2): 307-312.
- [4] 张德祥,吴小培,吕 钊,等. 基于经验模态分解和 Teager 峭度的语音端点检测[J]. 仪器仪表学报, 2010, 31(3): 4-7.
- [5] 刘 潮,李政杰,童宁宁,等. 改进的 EMD 算法及仿真分析[J]. 航天电子对抗, 2011, 27(2): 58-61.
- [6] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903-995.
- [7] Rilling G, Flandrin P, Goncalves P. On empirical mode decomposition and its algorithms[C]//IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing. NSIP-03, Grado (I), 2003, 3: 8-11.
- [8] Huang N E, Shen Z, Long S R. A new view of nonlinear water waves: the Hilbert spectrum 1 [J]. Annual Review of Fluid Mechanics, 1999, 31(1): 417-457.
- [9] 王宏志,徐玉超,李美静. 基于 Mel 频率倒谱参数相似度的语音端点检测算法[J]. 吉林大学学报: 工学版, 2012, 42(5): 1331-1335.
- [10] Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1980, 28(4): 357-366.

## Speech Endpoint Detection Based on the Empirical Mode Decomposition and Mel Cepstrum Coefficient

CHEN Wei, XIONG Wei-hua, SHI Wei-wei

(School of Mechanical Engineering & Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** This paper puts forward speech endpoint detection method based on empirical mode decomposition and Mel cepstrum coefficient. A group of IMF components were gained through decomposing speech signal. Noise signals concentrating at low-layer IMF components were filtered. The surplus IMF components were reconstructed to speech signal. Mel frequency cepstrum coefficient of reconstructed signal was extracted to detect endpoint of speech signal. The results show that this method can well eliminate effects of noise on endpoint detection of speech signal, and correctly detect endpoint of speech signal under different signal-to-noise ratios. This method has certain robustness.

**Key words:** speech endpoint detection; empirical mode decomposition; intrinsic mode function; Mel frequency cepstrum coefficient; signal-to-noise ratio

(责任编辑: 康 锋)

---

(上接第 569 页)

## Test Case Prioritization Techniques Based on Dependence Structure

ZHANG Na, LIU Yang-qiu, BAO Xiao-an, YU Cheng-hai, XU Li-cheng, DAI Yan-yun

(School of Information Science & Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** The prioritization technique can improve the performance of test case set without reducing the size of the test cases. In order to reduce testing workload, many literatures generally assume that the test case is independent, but in the actual test, there is dependence relationship among test cases. Only the test which considers dependence relationship can run properly. Aiming at this problem, this paper introduce the dependence relationship into prioritization technique and propose a kind of test case prioritization technique based on dependence structure. Firstly, the dependence relationship is transformed to quantitative index dependence depth. Then, this index serves as weighting function, and weighted depth-first search algorithm is applied to sort test cases. The results show that compared with the random sorting and prioritization techniques based on functional coverage, our strategy can reduce testing cost and improve defect detection rate.

**Key words:** software testing; software reliability; test case prioritization; dependency structure; depth-first search

(责任编辑: 陈和榜)