·人工智能及识别技术· 文章编号:1000-3428(2016)06-0171-04 文献标志码:A 中图分类号:TP391

基于改进 HHT 的语音端点检测

章小兵,李燕萍,王双杰

(安徽工业大学 电气与信息工程学院,安徽 马鞍山 243002)

摘 要: 针对带噪语音在不同噪声环境下,尤其是非平稳噪声下难以判断语音段端点的问题,提出一种基于改进的希尔伯特-黄变换(HHT)瞬时能量的端点检测方法。对每帧带噪信号进行经验模式分解得到固有模态函数(IMF),选取分量 IMF1,将其作为本底噪声并进行 HHT 得到噪声的瞬时能量,设置门限阈值,结合分量 IMF3 对带噪信号进行端点检测。该方法可提取非平稳噪声下带噪语音中的噪声成分,避免传统方法中选取前几帧信号作为噪声的局限性,同时利用分量 IMF3 进行端点检测达到滤波的效果。实验结果表明,该方法在不同噪声环境和低信噪比条件下提高了带噪语音端点检测的准确率。

关键词: 带噪语音;端点检测;希尔伯特-黄变换;瞬时能量;本底噪声;门限阈值

中文引用格式: 章小兵,李燕萍,王双杰. 基于改进 HHT 的语音端点检测[J]. 计算机工程,2016,42(6):171-174.

英文引用格式: Zhang Xiaobing,Li Yanping,Wang Shuangjie. Speech Endpoint Detection Based on Improved HHT[J]. Computer Engineering,2016,42(6):171-174.

Speech Endpoint Detection Based on Improved HHT

ZHANG Xiaobing,LI Yanping,WANG Shuangjie

(School of Electrical and Information Engineering,Anhui University of Technology,Maanshan,Anhui 243002,China)

【Abstract】 Aiming at the problem that speech endpoint is difficult to detect in different noise background,especially in non-stationary noise,an effective endpoint detection method is proposed based on Hilbert-Huang Transform(HHT) of instantaneous energy in the low Signal-to-noise Ratio(SNR) environment. Every frame of signal is decomposed into finite Intrinsic Mode Functions(IMF) by Empirical Mode Decomposition(EMD),the instantaneous energy of IMF1 is calculated to get the energy value and combine with the IMF3 to detect the speech endpoint. The proposed method can effectively extract the noise component in the non-stationary noise,and avoid the limitation about the traditional method of selecting the former several frames as the noise,meanwhile selecting IMF3 to endpoint detection can achieve the effect of the filter. Experimental results show that this algorithm improves the endpoint detection accuracy in different noise background and the low SNR environment.

【Key words】 speech with noise; endpoint detection; Hilbert-Huang Transform(HHT); instantaneous energy; ground noise; threshold

DOI: 10.3969/j.issn.1000-3428.2016.06.030

1 概述

所谓语音的端点检测就是判断一段待处理信号是否为语音信号,并从该段信号中找到语音部分的起止点。有效的端点检测技术不仅能缩短系统处理时间,而且能够排除静音段的噪声干扰。在语音识别过程中,端点检测的准确与否直接决定着语音识别后期处理的正确率[1]。

近年来,各种改进的语音端点检测方法层出不穷,其中具有代表性的端点检测阈值法通常选取带噪语音段前几帧设置阈值,这种方法在复杂的噪声分布不均匀的环境下,阈值设置误差过大,检测效果急剧下降。语音是一种非平稳状态的信号,传统双门限法利用假设其在相对较短的时间内(10 ms ~ 30 ms)是平稳信号,从而对待处理信号进行端点检测;目前应用非常广泛的倒谱特征[2-3]、谱熵[4-5]、小波变换[6]、隐马尔可夫模型(HMM)[7]等端点检测方法虽然在处理非线性非平稳信号的能力上有了进

基金项目: 安徽工业大学产学研基金资助重大项目(RD14206003)。

作者简介: 章小兵(1972 -),男,教授、博士,主研方向为智能测控、人机交互、信号处理;李燕萍(通讯作者)、王双杰,硕士研究生。

收稿日期: 2015-05-29 **修回日期:** 2015-06-25 **E-mail:** liyanping_e@163.com

一步提高,但在稳定白噪声条件下信噪比下降到10 dB以下时,双门限检测法、倒谱特征法已经不能正常工作,0 dB 时谱熵法只有 57% 的准确率,小波变换、HMM 的准确率虽然可以达到 80% ,但是小波变换存在多种小波基的选择,HMM 需要事先训练,使检测复杂度增加,在信噪比下降到 0 dB 以下时,检测结果较差。

希尔伯特-黄变换(Hilbert-Huang Transform,HHT)是 N. E. Huang 等人于 1998 年提出的处理非线性、非平稳信号的时频分析方法[8-11],近年来在语音信号处理领域有了广泛的应用。因此,本文提出低信噪比下基于改进的 HHT 的语音端点检测方法,对传统的阈值设置进行改进,以提高在低信噪比下检测端点的准确性及鲁棒性。

2 希尔伯特-黄变换原理

HHT 主要包含两大部分:经验模式分解(EMD)和 Hilbert 谱分析(HSA)[12]。HHT 是将待分解信号经 EMD 分解成若干 IMF,然后对这些 IMF 分量进行 Hilbert 变换得到 Hilbert 谱,从而对信号进行时频分析。

2.1 经验模式分解

HHT 方法最关键的步骤就是 EMD。将带分解信号经 EMD 分解得到的 IMF 分量必须满足在整个数据序列中含有相同过零点数和极值点个数,或不相同时最多相差 1 个,以及要求局部均值为零的这样 2 个条件[13]。但是这 2 个条件是基于理论上的,实际中几乎达不到,因此,必须确定一个筛选终止准则。假设在筛选 IMF 分量过程中,$h_{k-1}(t)$ 和 $h_k(t)$ 是其中的 2 个时间序列,则该终止准则可以用以下公式实现:

$$Sd=\sum_{t=0}^{T}\frac{|h_{k-1}(t)-h_k(t)|^2}{h_k^2(t)} \tag{1}$$

其中,T 表示信号的时间长度;Sd 的取值范围为 0.2 ~ 0.3。这样经过 EMD 分解的任何语音信号 $x(t)$ 都可用 n 个 IMF 分量和残余函数 $r_n(t)$ 表示:

$$x(t)=\sum_{i=1}^{n}c_i(t)+r_n(t),i=1,2,\cdots,n \tag{2}$$

2.2 Hilbert 谱分析

对每个 IMF 进行 Hilbert 变换得到 $y(t)$[14],则得到相位函数 $\Phi(t)$、瞬时频率 $w(t)$、瞬时幅值 $a(t)$:

$$y(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{c(\tau)}{t-\tau}\mathrm{d}\tau \tag{3}$$

$$\Phi(t)=\arctan\frac{y(t)}{x(t)} \tag{4}$$

$$w(t)=\frac{\mathrm{d}\Phi(t)}{\mathrm{d}t} \tag{5}$$

$$a(t)=\sqrt{x^2(t)+y^2(t)} \tag{6}$$

最后语音信号经过 HHT 变换得到:

$$x(t)=\mathrm{Re}\sum_{i=1}^{n}a_i(t)\exp(\mathrm{j}\int w_i(t)\mathrm{d}t) \tag{7}$$

3 改进的 HHT 语音端点检测

对于具有非平稳、非线性特点的语音信号,噪声的能量明显低于语音段的能量[15],因此针对这个特点可以计算出噪声的瞬时能量并设置一个门限阈值进行两者的区分。门限阈值往往通过带噪语音信号的本底噪声计算得到,所以设置阈值的关键在于噪声特征的提取,只要正确估算出带噪语音段的本底噪声,结合其能量特点并设置合理的门限阈值便可以提高语音端点的检测效果。

传统的阈值端点检测方法直接提取带噪信号的前几帧设置门限阈值,经实验发现当噪声在不足以完全掩盖语音段的提前下,且噪声分布越不均匀时,此种方法检测端点效果明显下降。原因是前几帧噪声信号的特征不足以代表整个带噪语音段中的噪声特点,具有局限性,从而导致检测误差较大。实验中发现,对带噪语音进行 EMD 分解得到的分量 IMF1 含有绝大部分的噪声,而这些噪声就是带噪语音段中的噪声,因此门限阈值的大小可以由 IMF1 的能量特征值计算得到,同时经 EMD 分解得到的 IMF3 可以很好地表达出加噪前的语音段特征,因此,本文方法采用将分量 IMF1 作为本底噪声,经希尔伯特变换得到噪声的瞬时能量,将这些瞬时能量值加权平均得到的特征值作为区分噪声段和语音段的门限阈值,利用该门限阈值对分量 IMF3 进行端点检测。

语音段的瞬时能量可以通过以下步骤得到:

(1)将信号进行 HHT 变换得到式(7),再将式(7)表示成幅度谱 $H(w,t)$:

$$H(w,t)=\begin{cases}\mathrm{Re}\sum_{i=1}^{n}a_i(t)\exp(\mathrm{j}\int w_i(t)\mathrm{d}t) & w(i)=w\\ 0 & \text{其他}\end{cases} \tag{8}$$

(2)对式(8)平方,进行频率积分,便得到瞬时能量 $IE(t)$:

$$IE(t)=\int_{w}H^2(w,t)\mathrm{d}w \tag{9}$$

其中,$IE(t)$ 是以时间 t 为自变量的函数,表示不同时间的能量波动。

经实验,利用分量 IMF1 作为本底噪声比取带噪信号的前 5 帧噪声信号更具有代表性,在低信噪比环境下该端点检测方法具有很好的检测效果,在噪声分布不均匀的环境下也保持着很高的准确率,没有出现准确率骤降的现象,具有很好的鲁棒性。

假设带噪语音信号为 $x(t)$,端点检测具体步骤如下:

(1)将经过采样、分帧之后的 $x(t)$ 进行 EMD 分解,得到一定数量的 IMF;

(2)分别选取 IMF1,IMF3 并对其进行短时分帧处理;

(3)将 IMF1 分量作为本底噪声,分帧,进行 Hilbert 变换,由式(4)~式(9)求得瞬时能量 IE,将其作为特征参数,并设置门限阈值 T_1,T_2:

$$T_1 = \bar{E}_t = \frac{1}{n}\sum_{i=1}^{n} IE(i) \tag{10}$$

$$T_2 = T_1 + \frac{1}{n}\sum_{i=1}^{n} (E(i) - \bar{E}_t)^2 \tag{11}$$

其中,n 为 $IMF1$ 分帧后的帧数。

(4)求取 IMF3 分帧后的每帧瞬时能量,结合门限阈值进行端点检测,当能量大于 T_2 且持续一定帧数时判断为语音段。

(5)针对判断出的语音段进行前后搜索,当能量介于 T_1,T_2 之间的最小值时标记为语音端点。

4　实验结果与分析

4.1　实验环境

为了检测本文方法的可行性和有效性,选用安静环境下实验室录制的“安工大”作为目标语音信号,采样频率为 8 kHz,采样精度为 16 bit,帧长选择 256 ms,帧移为 128 ms,采用标准噪声库 NOISEX-92中的高斯白噪声、babble 以及 factory 作为复杂干扰噪声。

4.2　结果分析

图 1 所示为原始语音波形图以及加入信噪比 SNR = −8 dB 的高斯白噪声后波形图。

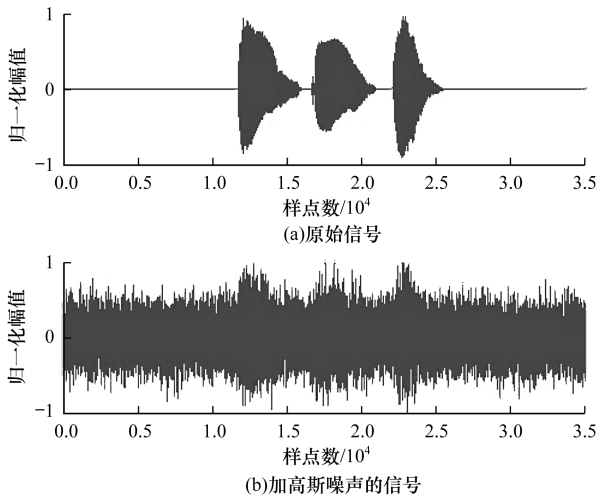

图 1　原始语音与加噪语音波形图

对该加入噪声的信号进行 EMD 分解。图 2 表示在 SNR = −5 dB 下分解后的 IMF 分量,分别为 IMF1,IMF3,IMF13。

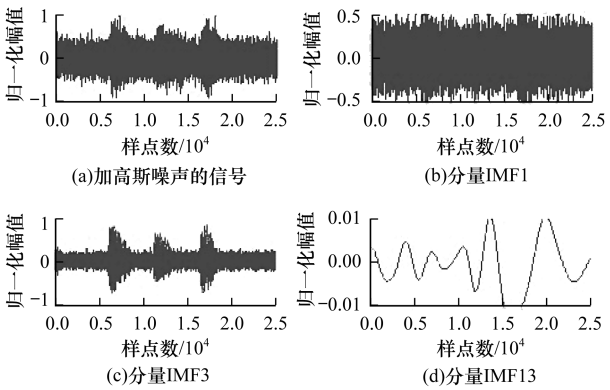

图 2　加噪语音和分解后的 IMF 分量

EMD 分解总是将频率较高的分量 IMF1 先提取出来,再依次由高到低,将各个 IMF 分量分离出来[16]。由上图可以看出加噪信号首尾段被已噪声掩盖,分量 IMF1 中语音信号几乎被噪声掩盖,其绝大部分都是噪声,低频分量 IMFn 又不能完全显示语音信号的特性,中高频分量 IMF3 较加噪后的语音信号而言更能表达出加噪前的语音特点。

因此,在低信噪比条件下,将 IMF1 分量作为本底噪声设置阈值具有可行性,尤其在噪声分布不均匀的环境下,对分量 IMF1 进行能量分析较选取带噪语音信号的前 5 帧更能反映出带噪语音中噪声的特点,从而提高了语音端点检测的准确率和鲁棒性。

图 3 所示为 SNR = −3 dB 时的检测结果。

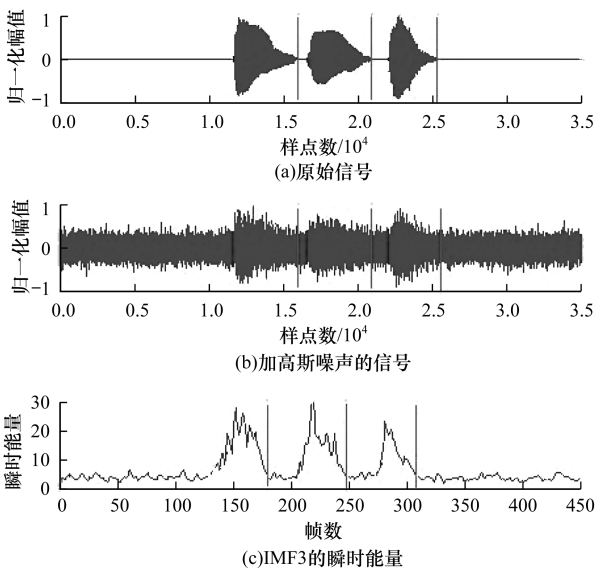

图 3　SNR 为 −3 dB 时的端点检测结果

小波变换、文献[6]中的方法(记为 NHHT 法)以及本文方法在 SNR = −8 dB 时的检测对比结果如图 4 所示。

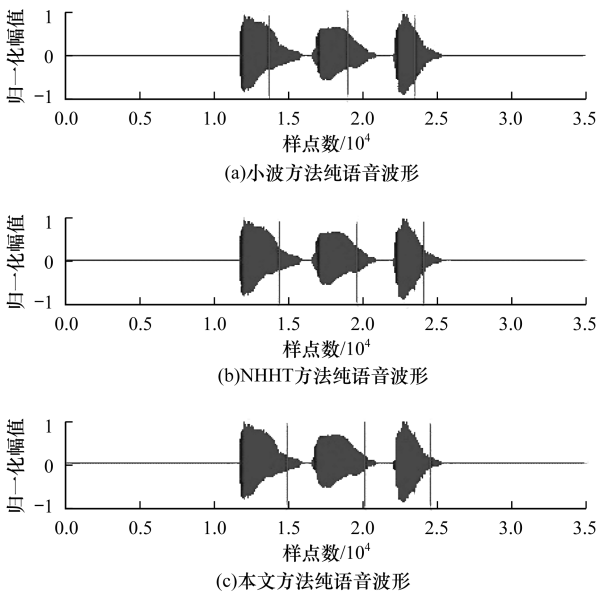

图 4 SNR 为 −8 dB 时的端点检测结果

由图 4 可以看出,本文方法检测的准确率明显高于小波变换方法以及 NHHT 方法。HHT 是处理非线性、非平稳信号的时频分析方法,所以在处理语音信号时采用 HHT 方法更能保存语音信号的特点,而且本文选取分解后的含有大部分噪声特点的 IMF1 分量作为本底噪声更能完整体现出带噪语音段中的噪声,因此计算出的阈值更精确,检测的准确率也明显提高。

文献[10]中已列出 NHHT 方法和小波等方法的比较结果,因此表 1 只记录了 NHHT 方法和本文方法在 SNR = −5 dB 时不同噪声环境下检测的准确率。

表 1 不同噪声下的端点检测结果 %

端点检测方法	白噪声	babbl 噪声	factory 噪声
NHHT 方法	83.32	66.12	72.73
本文方法	87.56	82.61	83.72

Babble 噪声和 factory 噪声均是分布不均匀的噪声,由表 1 可知本文方法在低信噪比环境下针对这 2 种噪声均有很好的检测效果,相对于 NHHT 方法,在 babble 噪声和 factory 噪声下,准确率分别提高了 19.96% 和 13.12%,由此可见在低信噪比条件下,本文方法在不同的噪声环境下都具有很好的检测效果。

5 结束语

端点检测对语音信号的后期处理起着至关重要的作用,实际处理中语音往往处于复杂的噪声环境,此时,判别语音段端点的问题主要归结为区别语音和噪声的问题。本文通过 HHT 变换中的 EMD 分解提取出带噪语音段中的本底噪声,并通过 Hilbert 变换计算出本底噪声的瞬时能量,根据噪声能量具有一定的规律性,与语音段有明显区别的特点,设置出区分噪声和语音段的阈值,最后将计算出的每帧信号的瞬时能量与阈值相比较,从而判断出语音段的端点。实验结果表明,在低信噪比下,该方法在不同的噪声环境下均可以有效地检测出端点,且检测效果理想。在实际应用中,如何在确保端点检测准确率的同时减少计算量、缩短处理时间将是下一步工作中研究的重点。

参考文献

[1] 鲁远耀,周 妮,肖 珂,等.强噪声环境下改进的语音端点检测算法[J].计算机应用,2014,34(5):1386-1390.
[2] 董 胡.倒谱距离和短时能量的语音端点检测方法研究[J].计算机技术与发展,2014,24(7):77-83.
[3] 王 帛,冯新喜.一种基于短时倒谱速变率的语音信号平滑端点检测方法[J].现代电子技术,2010(23):92-98.
[4] 刘华平,李 昕,郑 宇,等.一种改进的自适应子带谱熵语音端点检测方法[J].系统仿真学报,2008,20(5):1366-1371.
[5] Wu Di, Zhao Heming, Huang Chengwei, et al. Speech Endpoint Detection in Low-SNRs Environment Based on Perception Spectrogram Structure Boundary Parameter[J]. Chinese Journal of Acoustics,2014,33(4):428-440.
[6] 陈金龙,范影乐,倪红霞,等.基于小波包分解的含噪语音时频特性分析及端点检测[J].数据采集与处理,2014,29(2):293-297.
[7] Othman H, Abounasr T. A Semi-continuous State Transition Probability HMM-based Voice Activity Detection[C]//Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal, Canada: IEEE Press, 2004: 821-824.
[8] Huang N E, Shen Zheng, 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 A, 1998, 454(1): 903-995.
[9] Lu Zhimao, Liu Baisen, Shen Liran. Speech Endpoint Detection in Strong Noisy Environment Based on the Hilbert-Huang Transform [C]//Proceedings of International Conference on Mechareonics and Automation. Changchun, China: [s. n.], 2009: 4322-4326.
[10] 侯丽霞,曾以成,焦 蓓.强噪声环境下基于改进 HHT 的语音端点检测[J].计算机工程与应用,2012,48(28):139-142.
[11] Liu Baisen, Zhang Ye, Zhang Wulin. Speech Endpoint Detection with Low SNR Based on HHTSM [C]// Proceedings of the 11th IEEE International Conference on Electronic Measurement & Instruments. Washington D. C., USA: IEEE Press, 2013: 116-119.
[12] 卢志茂,金 辉,张春祥,等.基于 HHT 和 OSF 的复杂环境语音端点检测[J].电子与信息学报,2012,34(1):213-217.
[13] 申 涛,冯 刚.强背景噪声下基于 HHT 端点检测方法[J].电声技术,2014,38(1):69-72.
[14] 宋之用. Matlab 在语音信号分析与合成中的应用[M].北京:北京航空航天大学出版社,2013.
[15] Guo Yanmeng, Fu Qiang, Yan Yonghong. Speech Endpoint Detection in Real Noise Environments [J]. Chinese Journal of Acoustics, 2007, 26(1): 39-48.
[16] Zhang Dexiang, Wu Xiaopei, Zhao Lü. Speech Endpoint Detection in Noisy Environments Using EMD and Teager Energy Operator [J]. Journal of Electronic Science and Technology, 2010, 6(2): 183-186.

编辑 索书志