

基于先验信噪比和能零熵的语音端点检测算法

董 胡

(长沙师范学院 电子与信息工程系, 湖南 长沙 410100)

摘 要:端点检测技术是语音识别系统中的一项关键技术,其性能在某种程度上对整个语音识别系统有着较大的影响,传统的语音端点检测算法在低信噪比环境下存在端点检测正确率低、抗噪性能差等问题。针对传统端点检测算法在低信噪比环境下存在的上述问题,提出了一种基于先验信噪比估计和能零熵的语音端点检测算法。该算法通过改进的先验信噪比估计算法对含噪语音进行增强处理,并对增强后的语音信号设置自适应端点检测阈值,利用能零熵算法对增强后的语音信号进行端点检测,实现了低信噪比环境下的语音端点检测。仿真实验结果表明,与传统的能零积和谱熵端点检测算法相比,所提出的端点检测算法在不同的低信噪比环境下具有较好的鲁棒性与较高的端点检测正确率。

关键词:先验信噪比;鲁棒性;端点检测;谱熵;能零积

中图分类号: TN912.35

文献标识码: A

文章编号: 1673-629X(2017)07-0072-04

doi: 10.3969/j.issn.1673-629X.2017.07.017

Endpoint Detection Algorithm with Priori SNR and Energy-zero-spectral Entropy

DONG Hu

(Department of Electronic and Information Engineering, Changsha Normal University,
Changsha 410100, China)

Abstract: Speech endpoint detection is the key technology in voice recognition system, and its performance has a great influence for speech recognition system in some extent. However, traditional speech endpoint detection algorithms have problems of low accuracy and poor anti-noise under low SNR environment. In order to solve the problems above, a kind of speech endpoint detection algorithm based on prior SNR estimation and energy-zero-entropy has been proposed, in which the improved prior SNR estimation algorithm is employed to make speech enhancement processing of speech with noise and then the adaptive endpoint detection threshold is set for the enhanced speech signal. The energy-zero-spectral entropy algorithm is eventually adopted to make endpoint detection for enhanced speech signal and the speech endpoint detection under low noise environment is achieved. Simulation experiment results show that compared with traditional energy-zero-product and spectrum entropy endpoint detection algorithm, the proposed endpoint detection algorithm has better robustness and higher endpoint detection accuracy in different low SNR environment.

Key words: priori SNR; robustness; endpoint detection; spectral entropy; energy-zero-product

0 引 言

端点检测对语音识别、语音增强等语音信号处理都有较大影响。近年来,针对低信噪比环境下的语音端点检测成为研究热点。常见的语音端点检测算法有短时能量、过零率、谱熵^[1-2]、倒谱特征^[3-6]等,这些算法在信噪比较高时能有效区分语音与非语音的边界,但是伴随着信噪比的下降,这些算法的性能会逐渐降低。对涉及端点检测的三个特征参数—对数能量、过

零率和谱熵进行研究分析,提出了能零熵积的语音特征参数。通过改进的先验信噪比估计法对含噪语音信号进行增强处理,对增强后的语音信号采用能零熵法进行端点检测处理,实现了低信噪比环境下的语音端点检测。

1 改进的先验信噪比估计语音增强

含噪语音信号在时域可表示为:

收稿日期: 2016-08-20

修回日期: 2016-11-24

网络出版时间: 2017-06-05

基金项目: 湖南省教育厅科学研究优秀青年项目(17B025);湖南省教育厅科学基金项目(12C0952);长沙师范学院大学生研究性学习和创新性实验计划项目(DXVC201510);长沙师范学院院级科研项目(XYYB201517)

作者简介: 董 胡(1982-),男,硕士,讲师,研究方向为信号处理及嵌入式设计。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20170605.1509.062.html>

$$y(t) = x(t) + n(t) \quad (1)$$

其中, $y(t)$ 、 $x(t)$ 、 $n(t)$ 分别表示含有噪声的语音、纯净的语音及噪声信号。

对式(1)进行 FFT 变换可得:

$$Y(k, m) = X(k, m) + N(k, m) \quad (2)$$

其中, $X(k, m)$ 、 $N(k, m)$ 分别表示第 m 帧、第 k 谱分量对应干净语音和噪声的幅度谱。

经维纳滤波法得到的频谱增益函数为:

$$G_i(k, m) = \frac{\xi(k, m)}{1 + \xi(k, m)} \quad (3)$$

其中, $\xi(k, m)$ 表示第 m 帧第 k 个频点处的先验信噪比。

增强语音的幅度谱表示为:

$$\hat{X}(k, m) = G_i(k, m) \cdot Y(k, m) \quad (4)$$

对式(3)用到的 $\xi(k, m)$, 首先利用“直接判决”法^[7]进行计算:

$$\gamma(k, m) = \frac{Y^2(k, m)}{\lambda_d(k, m)} \quad (5)$$

$$\xi(k, m) = \alpha G_i^2(k, m - 1) \gamma(k, m - 1) + (1 - \alpha) \max(\gamma(k, m) - 1, 0) \quad (6)$$

其中, $\alpha = 0.98$ 为平滑系数; $\gamma(k, m)$ 为后验信噪比; $\lambda_d(k, m)$ 为第 m 帧背景噪声方差, 这里采用文献[8]中提出的噪声估计方法。

在“直接判决”法中, 式(6)中估计的 $\xi(k, m)$ 过分依赖前一帧语音的幅度谱估计, 有损语音增强性能。为得到更准确的 $\xi(k, m)$, 采用 MMSE 两步先验信噪比估计法作如下处理:

$$\hat{\xi}(k, m) = \sqrt{\frac{\xi(k, m)}{1 + \xi(k, m)}} (1 + \sqrt{\frac{\xi(k, m)}{1 + \xi(k, m)}} \gamma(k, m)) \quad (7)$$

其次, 研究表明先验信噪比估计中存在过估和欠估误差两种情况^[9], 尤其是低信噪比时大部分先验信噪比被过估, 会影响频谱增益函数, 进而影响增强语音的幅度谱, 最终影响语音增强可懂度^[10]。为降低过估和欠估误差对语音增强可懂度的影响, 通过文献[5]人工引入偏差修正频谱增益函数值:

$$\hat{G}_i(k, m) = (G_i(k, m) + C) / (1 + C), \quad 10 \log(\hat{\xi}(k, m)) < 0 \quad (8)$$

其中, $C = B / (1 - B)$, B 为先验信噪比小于 0 dB 的偏差修正因子, 通过实验取 $B = 0.2$ 效果较好。

此外, “直接判决”法存在一帧的延迟问题, 可利用文献[11]中的 TSNR 算法解决该问题。

2 能零熵端点检测

2.1 对数能量端点检测

对第 i 帧语音信号 $s(n)$, 文献[12]提出一种对数

能量特征 $LE(i)$, 如下:

$$LE(i) = \lg(E(i) + a) - \lg a \quad (9)$$

$$E(i) = \sum_{n=1}^N s^2(n) \quad (10)$$

其中, $E(i)$ 为第 i 帧信号的短时线性能量; a 为某常数。根据文献[12], 当 $a = 5 \times 10^5$ 时, 对数能量取得了较好的端点检测结果且端点检测效果优于短时能量。过零率端点检测算法见文献[13]。

2.2 谱熵端点检测

设含噪信号的时域波形是 $x(n)$, 经加窗分帧处理, 得到第 i 帧信号是 $x_i(m)$, 经 FFT 变换得到第 k 条谱线频率的分量 f_k 的能量谱 $Y_i(k)$, 定义每个信号频率分量的归一化谱概率密度函数为:

$$p_i(k) = \frac{Y_i(k)}{\sum_{l=0}^{N/2} Y(l)_i} \quad (11)$$

其中, $p_i(k)$ 为第 i 帧中第 k 个频率分量 f_k 的概率密度; N 为 FFT 长度。

语音帧短时谱熵定义如下:

$$H_i = - \sum_{k=0}^{N/2} p_i(k) \log p_i(k) \quad (12)$$

2.3 能零熵积特征计算

文献[14]提出的能零积端点检测算法简单、耗时少, 在高信噪比环境下具有良好的性能, 但在低信噪比环境下性能较差。而谱熵算法在相对较低的信噪比环境下可弥补能零积端点检测算法的不足, 结合文献[14]提出的能零积算法和文献[1]提出的谱熵端点检测算法, 提出能零熵积语音端点检测算法, 表示如下:

$$EZH(i) = \sqrt{1 + (LE(i) - A_{veE})(H^*(i) - A_{veH})(Z(i) - A_{veZ})} \quad (13)$$

其中, A_{veE} 表示语音信号前十帧对数能量均值; A_{veH} 表示语音信号前十帧短时谱熵均值; A_{veZ} 表示语音信号前十帧短时过零率均值。

能零熵积综合了时域与频域语音的特征。对数能量和过零率属于时域特征参数, 谱熵属于频域特征参数, 将时频域特征参数进行结合, 不仅可发挥它们各自的长处, 又能在一定程度上避免各自的缺陷, 进而可有效地应对不同的背景噪声, 其鲁棒性也得到增强。

2.4 端点检测自适应阈值选取

因为噪声环境随着时间在改变, 所以需要自适应地改变阈值用以区别语音部分与非语音部分。在较短的时间里, 谱熵的均值与方差可通过非语音部分进行估计, 通过局部噪声的统计来计算初始阈值, 表示为:

$$T = \delta_{new} + \alpha \cdot \sigma_{new} \quad (14)$$

其中, δ_{new} 与 σ_{new} 为噪声帧中能零熵的均值与方

差; α 为调整因子。

比较阈值和临近几帧语音的能零熵值, 当大于阈值时则判断为语音部分; 当小于阈值时, 则判断为非语音部分。语音部分的阈值保持不变, 而非语音部分的阈值则需根据噪声的统计值动态更新, 具体如下:

$$\delta_{\text{new}} = \beta \text{EZH}(i-1) + (1-\beta) \text{EZH}(i) \quad (15)$$

$$\sigma_{\text{new}} = \sqrt{|\text{EZH}_{\text{mean}}^2(i) - \delta_{\text{new}}|} \quad (16)$$

$$\text{EZH}_{\text{mean}}^2(i) = \beta \text{EZH}_{\text{mean}}^2(i-1) + (1-\beta) \text{EZH}^2(i) \quad (17)$$

$$\text{EZH}_{\text{mean}}^2(i-1) = \sum_{m=1}^3 \text{EZH}_{\text{mean}}^2(i-m)/3 \quad (18)$$

其中, β 为调整因子。通过大量实验取 $\alpha = 2.13$, $\beta = 0.42$ 。

3 仿真实验与分析

3.1 语音增强效果比较

实验所用语音样本采用 TIMIT 语音库及在实验室安静环境下录制的语音共计 760 条, 接着利用 NOISEX-92 噪声库中的白噪声和粉红色噪声, 分别添加至语音样本中, 生成 $-5 \sim 10$ dB 的含噪语音信号。进行 8 kHz 采样, 16 bit 量化处理, 采用 MATLAB 进行仿真。为检验所提出的改进算法的增强效果, 分别与常见的谱减法和维纳滤波进行比较。图 1、图 2 分别为 -5 dB 白噪声及粉红色噪声增强效果比较图。

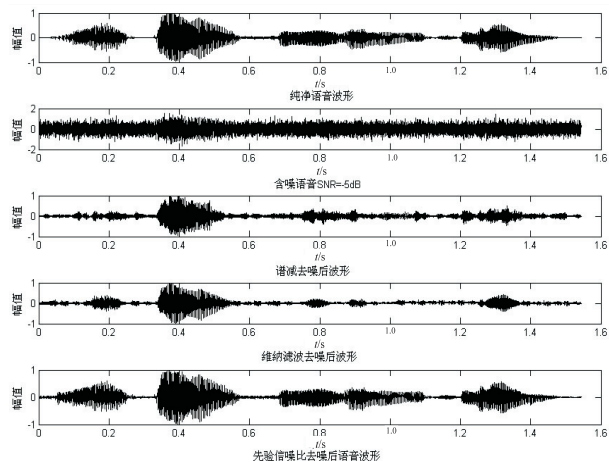


图 1 -5 dB 白噪声下三种增强方法比较

由图 1 和图 2 可知, 在 -5 dB 白噪声与粉红色噪声环境下, 常见的谱减法和维纳滤波法增强效果较差; 而所提出的先验信噪比法在低信噪比环境下能有效地提高语音输出信噪比, 同时经主观试听表明先验信噪比法能保持较好的语音可懂度。

3.2 语音端点检测效果比较

为进一步比较提出的能零熵端点检测算法的效果, 分别与能零积、谱熵端点检测算法在不同信噪比环境下的端点检测结果进行比较。图 3、图 4 分别为 -5

dB 白噪声及粉红色噪声环境下端点检测效果的比较, 图 5 为 0 dB babble 噪声环境下端点检测效果比较图。

由图 3 ~ 图 5 可知, 无论在 -5 dB 的白噪声和粉红色噪声环境下, 还是在 0 dB 的 babble 噪声环境下, 传统的能零积和谱熵端点检测算法几乎无法有效检测出语音信号的端起止端点, 而提出的端点检测算法却能较好地检测出语音起止端点位置。因而可知, 虽然传统的能零积和谱熵算法简单、易实现, 但抗噪性较差, 在低信噪比环境下几乎失去端点检测能力; 而提出的

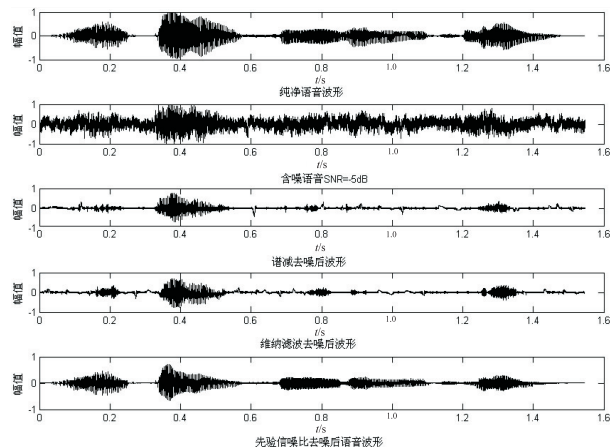


图 2 -5 dB 粉红色噪声下三种增强方法比较

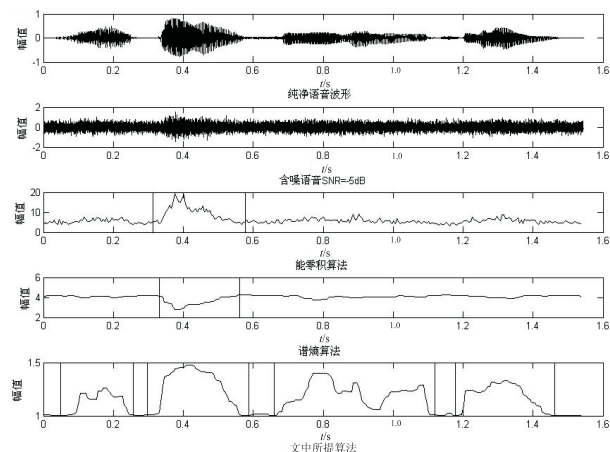


图 3 -5 dB 白噪声下三种端点检测效果比较

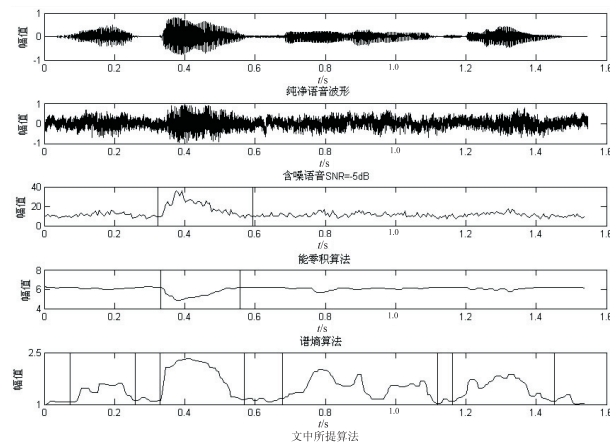


图 4 -5 dB 粉红色噪声下三种端点检测效果比较

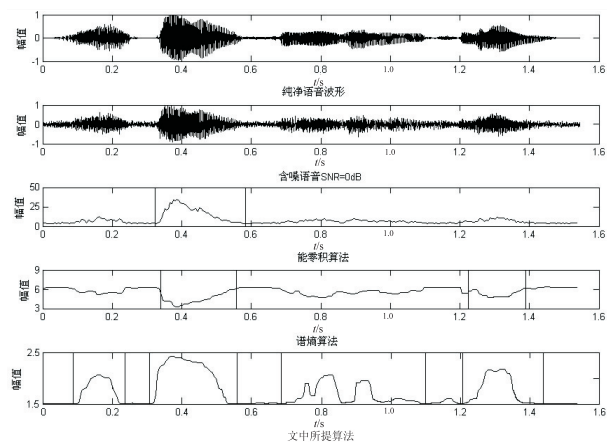


图5 0 dB babble 噪声下三种端点检测效果比较
端点检测算法却拥有较强的抗噪声能力与鲁棒性,在

表1 三种端点检测算法在不同信噪比下的检测正确率

SNR/dB	白噪声			粉红色噪声			babble 噪声		
	能零积/%	谱熵/%	所提算法/%	能零积/%	谱熵/%	所提算法/%	能零积/%	谱熵/%	所提算法/%
10	86.4	92.5	98.6	84.9	91.2	97.3	84.1	90.6	95.6
5	81.2	89.1	94.8	80.1	88.1	93.5	79.2	85.7	92.1
0	73.6	84.7	92.3	71.7	83.2	91.6	70.8	81.5	90.0
-5	61.5	81.3	90.6	59.8	80.6	89.3	61.3	78.5	85.9

4 结束语

低信噪比环境下的语音端点检测是语音处理的难点,为此,提出了端点检测算法。该算法在端点检测前进行了先验信噪比语音增强处理,有效提高了端点检测正确率,在传统的对数能量、过零率和谱熵端点检测算法的基础上,提出了能零熵端点检测算法。该算法不仅计算量小,而且有效利用了能零积和谱熵算法的优势,克服各自的缺陷。仿真结果表明,提出的端点检测算法在低信噪比环境下具有端点检测正确率高、鲁棒性好等优点,是一种行之有效的方法。但鉴于背景噪声的随机性,该算法对于更复杂低信噪比环境的适应性改进将是下一步研究的方向。

参考文献:

[1] 董 胡. 低信噪比环境下改进的语音端点检测算法[J]. 计算机技术与发展,2016,26(3):71-74.

[2] 李阳春,俞一彪. 倒谱本征空间结构化高斯混合模型语音转换方法[J]. 声学学报,2015,40(1):12-19.

[3] Ghosh P K,Tsiartas A,Narayanan S. Robust voice activity detection using long-term signal variability[J]. IEEE Transactions on Audio Speech & Language Processing,2011,19(3):600-613.

[4] Morita S,Unoki M,Lu X,et al. Robust voice activity detection based on concept of modulation transfer function in noisy reverberant environments[J]. Journal of Signal Processing Systems,2016,82(2):163-173.

[5] Zhang X L,Wang D. Boosting contextual information for deep

低信噪比的白噪声、粉红色及 babble 噪声环境下端点检测效果较好。进一步,可通过端点检测的正确率来比三种语音端点检测算法的性能。实验前,手工标记出每个语音样本的端点位置,作为评价端点检测是否正确的标准。端点检测正确率(R)计算公式如下:

$$R = (\text{语音帧总数} - \text{错误帧数}) / \text{语音帧总数} \times 100\%$$

(19)

其中,错误帧数为语音帧误判为噪声帧数以及噪声帧误判为语音帧数之和。

表1 给出了在不同信噪比环境下三种端点检测算法通过多次实验后的结果。由表1可知,在不同低信噪比的白噪声、粉红色及 babble 噪声环境下,所提算法的正确率要高于传统谱熵和能零积算法,鲁棒性好。

neural network based voice activity detection[J]. IEEE/ACM Transactions on Audio Speech & Language Processing,2016,24(2):252-264.

[6] 陈振锋,吴蔚澜,刘 加,等. 基于 Mel 倒谱特征顺序统计滤波的语音端点检测算法[J]. 中国科学院大学学报,2014,31(4):524-529.

[7] Ephraim Y,Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator[J]. IEEE Transactions on Acoustics Speech and Signal Processing,1985,33(2):443-445.

[8] Rangachari S,Loizou P C. A noise estimation algorithm for highly non-stationary environments[J]. Speech Communication,2006,48(2):220-231.

[9] Chen F,Loizou P. Impact of SNR and gain -function over-and underestimation on speech intelligibility[J]. Speech Communication,2012,54:272-281.

[10] 郭利华,马建芬. 具有高可懂度的改进的维纳滤波的语音增强算法[J]. 计算机应用与软件,2014,31(11):155-157.

[11] Plapous C,Marro C,Scalart P. Improved signal-to-noise ratio estimation for speech enhancement[J]. IEEE Transactions on Speech and Language Processing,2006,14(6):2098-2108.

[12] 赵 欢,王纲金,赵丽霞. 一种新的对数能量谱熵语音端点检测方法[J]. 湖南大学学报:自然科学版,2010,37(7):72-77.

[13] 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):77-79.

[14] 韩志艳,王 旭,王 健. 基于短时能零积和鉴别信息的语音端点检测[J]. 东北大学学报:自然科学版,2009,30(12):1690-1693.