

低信噪比环境下改进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:端点检测在语音识别中具有非常重要的作用,其准确性将直接影响语音识别系统的正确率。为了提高低信噪比环境下语音端点检测的正确率,提出了一种基于多窗谱估计的改进谱减法 and 能量谱熵的端点检测算法。该算法首先利用多窗谱估计改进谱减法对含噪语音进行去噪以提高语音信号信噪比,接着对去噪后的语音信号使用新的能量谱熵算法进行端点检测。仿真实验结果表明,同常见端点检测算法相比较,该算法在低信噪比环境下具有较好的端点检测正确率且有一定的鲁棒性,证明了该算法的有效性。

关键词:多窗谱估计;改进谱减法;谱熵;语音增强;端点检测

中图分类号:TN912.35

文献标识码:A

文章编号:1673-629X(2016)03-0071-04

doi:10.3969/j.issn.1673-629X.2016.03.017

Improved Speech Endpoint Detection under Low SNR Environment

DONG Hu

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

Abstract: Endpoint detection has a very important role in speech recognition, its accuracy will directly affect the accuracy of speech recognition system. In order to improve the accuracy of speech endpoint detection under low SNR environment, an endpoint detection algorithm based on spectral subtraction of multitaper spectrum estimation and spectral entropy is proposed. Firstly, it uses improved spectral subtraction of multitaper spectrum estimation to denoise speech signal in order to improve signal to noise ratio, and then it utilizes energy-entropy-ratio algorithm to make endpoint detection for speech signal denoised. Simulation experiment results show that compared with common endpoint detection algorithm, this algorithm has good endpoint detection accuracy and certain robustness in low SNR environment. It proves the effectiveness of the proposed algorithm.

Key words: multitaper spectrum estimation; improved spectral subtraction; spectral entropy; speech enhancement; endpoint detection

0 引 言

语音端点检测是信号处理的一个重要方面^[1]。特征参数的提取是一些常见的端点检测方法所普遍重视的方面,却往往忽略了之前的语音增强工作,这将会对端点检测的准确度产生影响^[2-5]。相对于短时能量特征,文献[6]提出了对数能量特征。它较好地解决了小幅度的辅音和静音混淆问题,更不会出现短时能量中噪声段特征值过大的情况,能较好地区别语音、噪声和静音。然而,在非平稳含噪环境下,能量较难区分语音与背景噪声,而谱熵可克服能量这一缺点^[7-8]。但是,在嘈杂噪声和音乐噪声之下,谱熵变得不稳定,而能量却可抑制该不足之处。根据能量的加性性质,语

音与噪声的能量之和大于噪声能量。

基于上述考虑,文中提出了多窗谱估计^[9-10]改进谱减和能量谱熵相结合的语音端点检测算法。首先,通过多窗谱估计改进谱减法对含噪语音信号进行谱减增强;接着,使用新的能量谱熵算法进行端点检测。仿真结果表明,在低信噪比环境下,文中提出的新算法与文献[11]提出的改进语音端点检测法相比,有更好的端点检测效果及鲁棒性。

1 多窗谱估计

针对同一数据序列,周期图法仅用单个数据窗,而多窗谱用多个正交的数据窗依次求直接谱,接着求平

收稿日期:2015-06-22

修回日期:2015-09-24

网络出版时间:2016-02-18

基金项目:湖南省科技计划项目(2012FJ3010);湖南省教育科研项目(12C0952);长沙师范学院科研基金项目(XXZD201218);长沙师范学院大学生研究性学习和创新性实验计划项目(DXYC201510)

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

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160218.1636.064.html>

均以得到谱估计,从而能得到较小的估计方差^[12]。多窗谱与周期图相比是一种更准确的谱估计方法。

多窗谱定义为:

$$S^{mt}(w) = \frac{1}{L} \sum_{k=0}^{L-1} S_k^{mt}(w) \quad (1)$$

式中: L 为数据窗个数; S^{mt} 为第 k 个数据窗的谱。

$$S_k^{mt}(w) = \left| \sum_{n=0}^{N-1} a_k(n) x(n) e^{-jnw} \right|^2 \quad (2)$$

式中: $x(n)$ 为数据序列; N 为序列长度; $a_k(n)$ 为第 k 个数据窗,满足多个数据窗互相正交。

$$\begin{cases} \sum a_k(n) a_j(n) = 0, k \neq j \\ \sum a_k(n) a_j(n) = 1, k = j \end{cases} \quad (3)$$

数据窗也叫 Slepian 窗。Riedal 和 Siderenko 提出下面的正弦窗:

$$a_k(m) = \sqrt{\frac{2}{N+1}} \sin\left(\frac{\pi k(m+1)}{N+1}\right), m = 0, 1, \dots, N-1 \quad (4)$$

正弦窗产生的局部偏差比 Slepian 窗更小,但却具有大致一样的频谱集中度。

利用 Matlab 软件对多窗谱估计函数和周期图跟真实谱的差异进行对比。其中,周期图使用 Hamming 窗,真实谱采用 AR 模型来估计。对比实验采用两个正弦信号叠加:

$$x(n) = \cos(2\pi \times 50n) + 3\cos(2\pi \times 200n) \quad (5)$$

周期图与多窗谱估计性能对比见图 1。

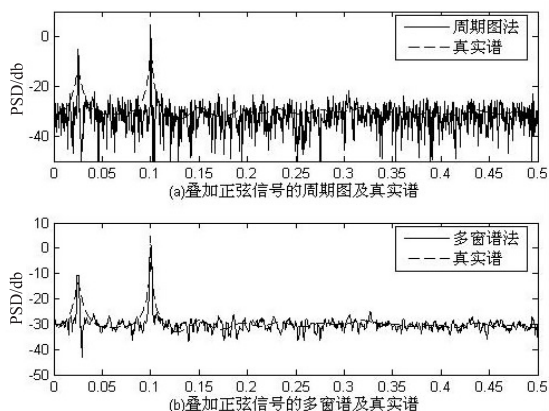


图 1 多窗谱与周期图法谱估计性能对比

从图 1 可知,周期图法的方差性能要劣于多窗谱法。所以,采用多窗谱进行功率谱估计来实现语音增强,能更好地降低音乐噪声。

2 多窗谱估计的改进谱减法

利用 Matlab 的信号处理工具箱中的多窗谱功率谱估算函数 `pmtm`, 可估算多窗谱的功率谱密度,进而获得谱减法中的增益因子,实现谱减语音增强运算^[13]。具体步骤如下:

(1) 令 $x(n)$ 为带噪语音, $x_i(n)$ 为加窗分帧后信号,邻帧之间有重叠。

(2) 对分帧后的信号进行 FFT, 分别求其幅度谱 $|X_i(k)|$ 和相位谱 $\theta_i(k)$, 在相邻帧之间做平滑处理, 计算平均幅度谱 $|\bar{X}_i(k)|$:

$$|\bar{X}_i(k)| = \frac{1}{2M+1} \sum_{j=-M}^M |X_{i+j}(k)| \quad (6)$$

以 i 帧为中心前后各取 M 帧,共有 $2M+1$ 帧进行平均。实际中取 M 为 1,即在 3 帧中进行平均。

(3) 对分帧后的信号 $x_i(m)$ 进行多窗谱估计,可得多窗谱功率谱密度 $P(k, i)$ (i 表示第 i 帧, k 表示第 k 条谱线):

$$P(k, i) = \text{PMTM}[x_i(m)] \quad (7)$$

式中, `PMTM` 表示进行多窗谱功率谱密度估计。

(4) 对多窗谱功率谱密度估计值也进行相邻帧间平滑处理,计算平滑功率谱密度 $P_y(k, i)$:

$$P_y(k, i) = \frac{1}{2M+1} \sum_{j=-M}^M P(k, i+j) \quad (8)$$

以 i 帧为中心前后各取 M 帧,共有 $2M+1$ 帧进行平均。实际中取 M 为 1,即在 3 帧中进行平均。

(5) 已知前导无话段(噪声)占有 NIS 帧,可以计算出噪声的平均功率谱密度值 $P_n(k)$:

$$P_n(k) = \frac{1}{NIS} \sum_{i=1}^{NIS} P_y(k, i) \quad (9)$$

(6) 利用谱减关系。

$$\begin{cases} g(k, i) = (P_y(k, i) - \alpha P_n(k)) / P_y(k, i), P_y(k, i) - \alpha P_n(k) \geq 0 \\ \beta P_n(k) / P_y(k, i), P_y(k, i) - \alpha P_n(k) < 0 \end{cases} \quad (10)$$

式中: α 为过减因子; β 为增益补偿因子。

选择恰当的 α 值能有效消除音乐噪声,但 α 值过大则会引起语音失真。

(7) 由平均幅度谱 $|\bar{X}_i(k)|$ 与增益因子 $g(k, i)$, 可得谱减后的幅度谱:

$$|\hat{X}_i(k)| = g(k, i) \times |\bar{X}_i(k)| \quad (11)$$

将幅度谱 $|\hat{X}_i(k)|$ 与相位谱 $\theta_i(k)$ 进行 IFFT, 将频域 $|\hat{X}_i(k)|$ 还原至时域,即得增强后的信号 $\hat{x}_i(m)$:

$$\hat{x}_i(m) = \text{IDFT}[|\hat{X}_i(k)| \exp[j\theta_i(k)]] \quad (12)$$

多窗谱估计的改进谱减算法见图 2。

3 对增强后语音信号进行端点检测

3.1 对数能量

对第 i 帧语音信号 $s(n)$, 文献[14]提出一种对数能量特征 $\text{LE}(i)$, 计算表达式如下:

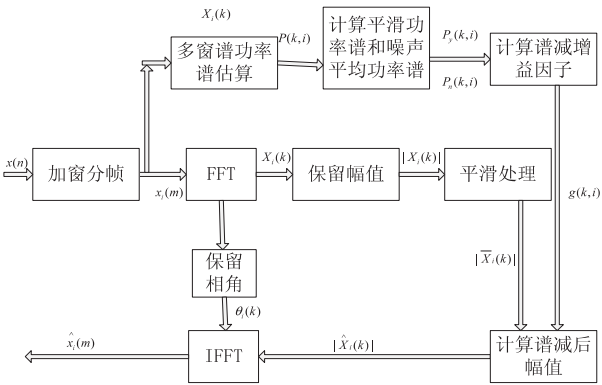


图2 多窗谱估计的改进谱减算法示意图

$$LE(i) = \lg(E(i) + a) - \lg a$$
 (13)

$$E(i) = \sum_{n=1}^N s^2(n)$$
 (14)

式中： $E(i)$ 为第 i 帧信号的短时线性能量； a 为某常数。

文献[14]实验表明,当 $a = 5 \times 10^5$ 时,取得了较好的端点检测性能。

3.2 谱 熵

令 $x(n)$ 为含噪的时域信号,加窗分帧之后得到第 i 帧信号 $x_i(m)$,通过 FFT 变换,得到第 k 条谱线频率分量 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^[15]:

$$p_i(k) = \frac{Y_i(k)}{\sum_{l=0}^{N/2} Y(l)_i}$$
 (15)

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

语音帧的谱熵定义为^[15]:

$$H(i) = - \sum_{k=0}^{N/2} p_i(k) \log p_i(k)$$
 (16)

3.3 能量谱熵特征计算

首先,将含噪语音信号通过多窗谱估计的改进谱减法进行降噪处理;接着,对降噪后的信号进行分帧与加窗,再分别计算对数能量与谱熵;最后,将对数能量与谱熵作除法运算,其结果记为 $B(i)$ 。具体计算表达式定义如下:

$$B(i) = LE(i) / H(i)$$
 (17)

最终能量谱熵特征计算公式定义如下:

$$EH(i) = \sqrt{[1 + |B(i)|]}$$
 (18)

低信噪比环境下可利用谱熵进行端点检测,而语音信号能量大小的变化则可通过对数能量反映。在无噪声情况下,噪音帧和语音帧的对数能量区别非常明显,因此将两者相除,形成了新的特征参数——能量谱熵。结合两种算法自身一定的抗噪能力,将二者结合起来,在低信噪比环境下的检测能力也有了一定程度的提高,鲁棒性得到增强。

4 仿真实验结果分析

实验语音样本使用 TIMIT 语音库中 450 条连续语音,16 kHz 采样频率,16 bit 量化,单声道 wav 格式。加噪数据使用 NOISEX 92 标准噪声库,分别添加 white、pink、volvo 噪声至纯净语音样本中,分别形成 -5dB、0 dB、5dB 和 10dB 的含噪语音信号 1 500 条。

含噪语音信号经多窗谱估计改进谱减法增强之后,接着采用能量谱熵法进行端点检测。为了检验文中提出的端点检测算法的效果,分别在 white、pink 及 volvo 噪声环境下,将其和文献[11]提出的端点检测法作比较,具体见图 3~5。

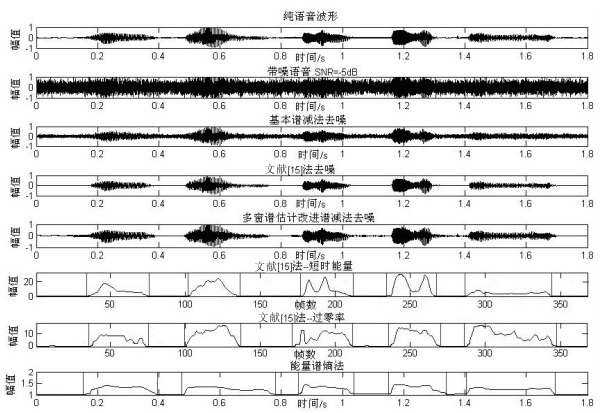


图3 SNR=-5 dB 时 white 环境下端点检测效果比较

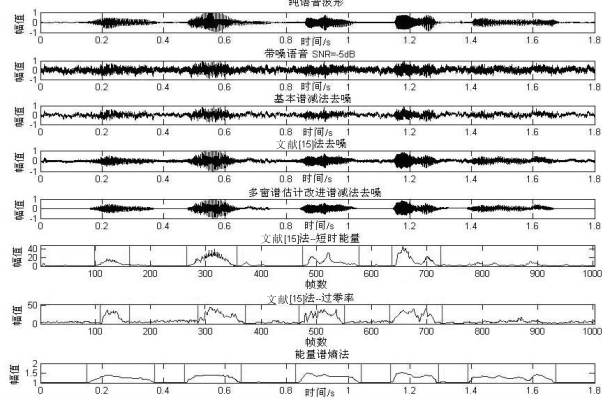


图4 SNR=-5 dB 时 pink 环境下端点检测效果比较

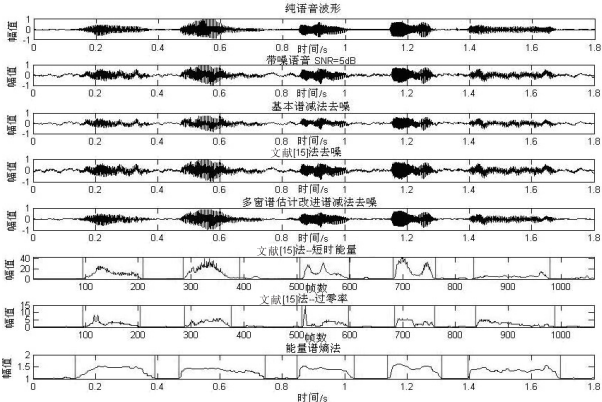


图5 SNR=5 dB 时 volvo 环境下端点检测效果比较

从图3、图5及表1可知,文中提出的能量谱熵算

法和文献[11]提出的端点检测算法均可在低信噪比环境下检测出语音的起止端点,但文献[11]提出的端点检测算法的端点检测正确率却低于文中提出的能量谱熵端点检测法。

表 1 端点检测正确率比较

端点检测算法	white 噪声				pink 噪声				volvo 噪声			
	-5 dB	0 dB	5 dB	10 dB	-5 dB	0 dB	5 dB	10 dB	-5 dB	0 dB	5 dB	10 dB
文献[14]算法	72.3	85.4	90.4	96.1	68.9	78.2	83.4	89.6	失效	76.5	83.0	88.5
文中算法	84.5	92.3	95.1	98.8	80.3	85.1	90.8	94.5	79.7	84.8	88.9	92.8

从图 4 可知,文献[11]提出的端点检测算法能检测出语音端点位置但存在漏检,而文中提出的能量谱熵算法却不存在。说明能量谱熵算法有更好的端点检测特性,在低信噪比环境下有更高的端点检测正确率。端点检测正确率定义如下:

错误帧数= 语音误判为噪声的帧数 + 噪声误判为语音的帧数

(19)

正确率 = (总帧数 - 错误帧数) / 总帧数

(20)

5 结束语

文中提出了一种低信噪比环境下的语音端点检测算法。该算法通过对含噪语音信号先用多窗谱估计改进谱减法进行增强,提高了语音信号的信噪比,接着采用能量谱熵算法进行端点检测。仿真实验结果表明,该算法在低信噪比环境下能较好地实现语音端点检测,具有良好的抗噪性能,鲁棒性较强,适于实际端点检测应用,从而证明该算法是一种较有效的低信噪比端点检测算法。但是,由于实际背景噪声变化性大,如何进一步改善算法,使其适应更复杂的低信噪比环境将是今后研究工作的重点。

参考文献:

[1] 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.

[2] Rabiner L R,Sambur M R. An algorithm for determining the endpoints of isolated utterances [J]. Bell System Technical Journal,1975,54(2):297-315.

[3] 武 薇,范影乐,庞 全. 基于广义维数距离的语音端点检测方法[J]. 电子与信息学报,2007,29(2):465-468.

[4] 闫润强,朱贻盛. 基于信号递归度分析的语音端点检测方法[J]. 通信学报,2007,28(1):35-39.

[5] Jia Chuan,Xu Bo. An improved entropy-based endpoint detection algorithm [C]//Proceedings of ISCSLP. Taiwan, China:[s. n.],2002:96-99.

[6] 肖述才,王作英. 端点检测中的一种新的对数能量特征[J]. 电声技术,2004(6):37-41.

[7] Huang L S,Yang C H. A novel approach to robust speech endpoint detection in car environments[C]//Proc of IEEE international conference on acoustics,speech and signal processing proceedings. Istanbul,Turkey:IEEE,2000:1751-1754.

[8] 李荣荣,胡昌奎,余 娟. 基于谱熵的语音端点检测算法改进研究[J]. 武汉理工大学学报,2013,35(7):134-139.

[9] 武鹏鹏,赵 刚,邹 明. 基于多窗谱估计的改进谱减法[J]. 现代电子技术,2008,31(12):150-152.

[10] 彭 军,王 忠,刘兴涛,等. 基于多窗谱相关加权语音增强[J]. 计算机仿真,2011,28(3):142-145.

[11] 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.

[12] Thomson D J. Spectrum estimation and harmonic analysis[J]. Proc of IEEE,1982,70(9):1055-1096.

[13] Hu Yi,Loizou P C. Incorporating a psycho acoustical model in frequency domain speech enhancement[J]. IEEE Signal Processing Letters,2004,11(2):270-273.

[14] 王 民,孙 广,沈利荣,等. 基于对数能量倒谱特征的端点检测算法[J]. 计算机工程与应用,2014,50(16):198-201.

[15] Zhao H,Zhao L X,Zhao K,et al. Voice activity detection based on distance entropy in noisy environment[C]//Proc of 5th international joint conference on INC,IMS,and IDC. Seoul,Korea:IEEE Computer Society,2009:1364-1367.