一种基于改进谱熵的语音端点检测方法

李 艳 成凌飞 张培玲
(河南理工大学电气工程与自动化学院 焦作 454000)

摘 要 针对常规谱熵端点检测法在非平稳噪声环境下检测效果差的缺陷,提出了一种基于子带谱熵幅度积参数的语音端点检测方法。该方法利用非平稳信号处理技术将语音信号的时域分析和频域分析相结合,在常规谱熵的基础上计算出子带谱熵,再结合时域中的短时平均幅度进行端点检测。仿真结果表明,与常规谱熵端点检测算法和短时平均幅度算法相比,该方法在各种噪声环境下的检测效果都比较好,鲁棒性增强,其有效性得到验证。

关键词 端点检测,谱熵,短时平均幅度,鲁棒性
中图法分类号 TN912.34 **文献标识码** A

Speech Endpoint Detection Based on Improved Spectral Entropy

LI Yan CHENG Ling-fei ZHANG Pei-ling
(School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract In view of the problem that conventional spectral entropy speech endpoint detection algorithm's detection effect is poor under the non-stationary noise, a new feature parameter—sub-band amplitude spectrum entropy was proposed. The new parameter detection of speech endpoint uses non-stationary signal processing technology to combine the signal of time domain and frequency domain characteristics. Firstly, the conventional spectral entropy speech endpoint detection algorithm is improved and the multi-band spectral entropy is calculated, then the endpoint is detected with the combination of short time average magnitude. The simulation results show that this method has better robustness and precision than conventional spectral entropy algorithm and average magnitude algorithm, which proves the effectiveness of the proposed method.

Keywords Endpoint detection, Spectral entropy, Short-time average magnitude, Robustness

1 引言

语音端点检测的目的是在一段信号流中快速区分出语音段和非语音段,并准确地标识出语音段的起始点和结束点[1,2]。准确的端点检测与定位能够提炼出真正的语音数据,从而减少系统存储量、运算量以及处理时间,直接影响到语音信号的特征提取和系统识别率,因此其在语音识别问题中有着举足轻重的地位[3-5]。

目前常见的语音端点检测方法可分为两大类:1)基于语音信号时域特征的算法,例如短时能量、短时平均幅度、短时平均过零率和短时相关分析等[6]。这类方法原理简单、运算量小,在高信噪比环境下可以获得较好的检测效果,但对复杂背景噪声环境下的端点检测误判率升高[7]。2)基于语音信号频域特征的算法,如倒谱特征法[8]、谱熵法[9]和频带方差[10]等。这些算法在纯净语音环境下可以取得良好的检测效果,但在非平稳噪声环境下检测效果骤降,尤其是当噪声与语音信号频域分布类似时,这些算法甚至不能正常工作[11]。

谱熵在嘈杂噪声和音乐噪声环境下非常不稳定[12],而短时平均幅度却能够克服这一缺点,因为幅度有一个很好的加性性质,即语音加噪声的幅度要大于噪声的幅度。因此,本文提出了一种新的基于子带谱熵幅度积参数的语音端点检测方法。该方法通过对频域中常规谱熵算法进行改进,先计算出子带谱熵,再结合信号时域中的短时平均幅度算法得到子带谱熵幅度积,该特征参数将两者所适用的环境进行了融合。仿真实验表明,与常规谱熵端点检测法和短时平均幅度法相比,利用本文提出的新参数检测语音端点可以提高复杂背景噪声下的检测精度。

2 改进谱熵

2.1 短时平均幅度

设语音信号为 $x(m)$,短时平均幅度[13]的定义为

$$M_n=\sum_{m=-\infty}^{\infty}|x(m)|\omega(n-m) \tag{1}$$

其中,$\omega(n)$是窗函数,一般使用 hamming 窗。

与语音信号的短时能量相比,短时平均幅度函数没有平方运算,因此其动态范围(最大值和最小值之比)要比短时能

本文受国家自然科学基金(51244003),河南省高等学校矿山信息化重点学科开放实验室开放基金(KZ2012-01)资助。
李 艳(1991—),女,硕士生,主要研究方向为语音识别;成凌飞(1971—),男,博士,教授,主要研究方向为井下无线通信理论;张培玲(1977—),女,博士,副教授,主要研究方向为语音信号处理、通信信号处理、电力线通信等,E-mail:plzhang@hpu.edu.cn(通信作者)。

量小，对高电平信号不是很敏感，同时在应用于端点检测时也不容易丢失幅值较小的音节和清音。

2.2 基本谱熵

为了解决语音信号能量小时易被噪声所掩盖而难于用能量特征参数进行语音端点检测的问题，Shen[14]等人提出了基于信息熵的端点检测方法，谱熵的基本原理如下。

定义 1 对带噪语音信号 $s(n)$ 进行预加重、加窗、分帧、帧移后计算每一帧信号的短时自相关函数并求其 FFT 变换，得其某频率点的功率谱幅度 $X(k,m)$，如式(2)所示：

$$X(k,m)=\sum_{n=1}^{N} r(n)\mathrm{e}^{-j\frac{2\pi kn}{N}},1\leqslant k\leqslant N,n=0,1,\cdots,L \tag{2}$$

其中，$r(n)$ 表示每一帧信号的短时自相关函数，L 为窗长，N 为 FFT 变换长度，$X(k,m)$ 表示第 m 帧第 k 频率点的功率谱幅度，对实际信号来说，$X(k,m)$ 是关于 $N/2+1$ 对称的，所以功率谱能量的计算如下式：

$$X_{power}(k,m)=X(k,m),1\leqslant k\leqslant N/2 \tag{3}$$

由此计算每一个频率分量的功率谱能量占整个这一帧的功率谱能量的概率：

$$p(i,m)=\frac{X_{power}(i,m)}{\sum_{m=1}^{q}\sum_{k=1}^{N/2}X_{power}(k,m)},1\leqslant i\leqslant N/2 \tag{4}$$

由于语音信号的能量主要集中在 250～4500Hz，为了增强区分语音和非语音段的能力，对式(4)引入约束条件：

$X(k,m)=0$, if $k<250$Hz or $k>4500$Hz

考虑上述约束条件后，在每帧上相应的功率谱熵为：

$$H(m)=\sum_{i=1}^{N/2}p(i,m)\cdot\log[1/p(i,m)] \tag{5}$$

对于频带宽度受限的信号而言，频率在 250～4500Hz 范围内，语音信号的随机事件比较多，平均信息量大，熵值也大，而背景噪声在此频率范围受限的情况下，熵值较小。本文中所有的仿真结果都是在频率范围受限(250～4500Hz 之间)的情况下得到的。

2.3 子带谱熵

从式(5)中可看出功率谱熵只依赖每个功率谱能量的变化，不依赖总的谱能量。相应地，谱熵参数对噪声的变化是鲁棒的，但是每一个功率谱点上的幅度易受噪声的干扰，在更低的信噪比下该参数的性能会有所下降。文献[15]对这一问题提出了解决方法，即子带谱熵的概念，其思想是将一帧语音信号分成若干个子带，再对每一子带求谱熵，这样就消除了每一谱点的幅值受噪声影响的问题。所以本文对带噪的语音信号也进行了多带分析，这样就部分地克服了在噪声环境中功率谱幅度的敏感性。假设 q 为帧数，N_b 为每一帧的子带数，本文中 $N_b=N/8$，$E_b(l,m)$ 表示第 m 子带的子带能量，定义如下：

$$E_b(l,m)=\sum_{k=1+(l-1)\cdot 4}^{1+(l-1)\cdot 4+3}X_{power}(k,m),1\leqslant l\leqslant N_b \tag{6}$$

定义子带功率谱概率为：

$$p_b(l,m)=\frac{E_b(l,m)}{\sum_{m=1}^{q}\sum_{k=1}^{N_b}E_b(k,m)},1\leqslant l\leqslant N_b \tag{7}$$

由此定义子带功率谱熵：

$$H_b(m)=\sum_{l=1}^{N_b}p_b(l,m)\cdot\log[1/p_b(l,m)] \tag{8}$$

2.4 改进谱熵-子带谱熵幅度积特征

针对常规谱熵端点检测算法在非平稳噪声环境下检测效果差的缺陷，以及短时平均幅度端点检测算法在复杂背景噪声环境下端点检测误判率升高的问题，本文提出了一种改进子带谱熵的语音端点检测方法。该方法在常规谱熵的基础上计算出子带谱熵，再结合信号时域中的短时平均幅度进行端点检测，也即利用子带谱熵幅度积参数进行语音端点检测。具体实现步骤如下：

①对输入的语音信号进行预加重、加窗、分帧、帧移等处理；

②对进行上述处理后的语音信号分别在时域计算其短时平均幅度，在频域计算其子带谱熵；

③为了融合语音信号的时域和频域特征，并充分利用基于短时平均幅度端点检测方法与基于子带谱熵的端点检测方法各自的优点，将进行平移调整后的语音短时平均幅度和子带谱熵相乘，即得到每一帧的子带谱熵幅度积特征参数以对语音进行端点检测。

因此，本文所提出的新特征参数——子带谱熵幅度积的计算表达式为：

$$H_b(m)M_n=(H_b(m)-Ave_{H_b})\cdot(M_n-Ave_{M_n}) \tag{9}$$

其中，Ave_{H_b} 和 Ave_{M_n} 分别为前 10 帧的子带谱熵和短时平均幅度特征的平均值。

图 1 给出了计算子带谱熵幅度积特征的算法框架，在对输入的语音信号进行加窗分帧处理时，本文使用 hamming 窗，设置帧长为 256，帧移为 128。

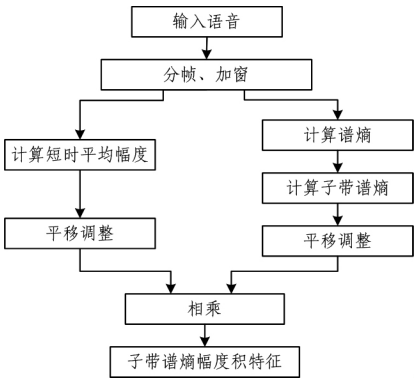

图 1 计算子带谱熵幅度积的算法框图

3 实验仿真

为了验证本文所提算法的可行性以及优越性，利用 MATLAB 软件做了相应的仿真实验。仿真过程中所使用的语音样本是使用计算机声卡在安静环境下录制的采样频率为 8kHz、16bit 量化的.wav 声音文件，所录语音样本的内容是“三”，噪声是采自 NOISEX-92 噪音库的典型噪声，分别为白噪声(White)、粉红噪声噪声(pink)、背景说话噪声(Babble)、工厂噪声(Factory)和高速行驶汽车噪声(Volvo)。利用所录语音和采集的不同噪声合成多段信噪比为 -5dB 的含噪语音，然后分别利用文献[13]中所提出的短时平均幅度法、文献[14]中提出的常规谱熵检测法和本文提出的子带谱熵幅度积对所录制的纯净语音信号和加入了各种噪声的带噪语音进行端点检测。实验结果如图 2－图 7 所示。

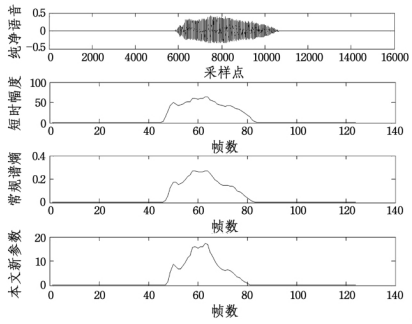

图 2　纯语音时 3 种方法的检测结果

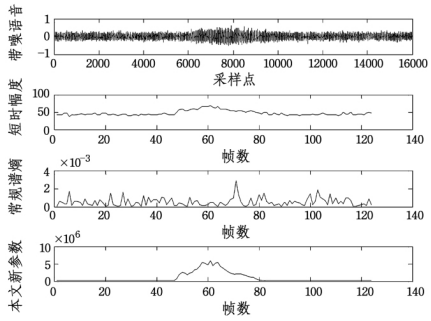

图 3　混有－5dB 白噪声时 3 种方法的检测结果

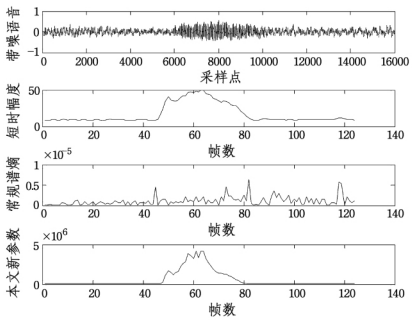

图 4　混有－5dB 工厂噪声时 3 种方法的检测结果

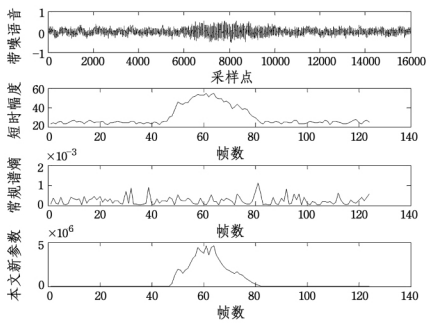

图 5　混有－5dB 粉红噪声时 3 种方法的检测结果

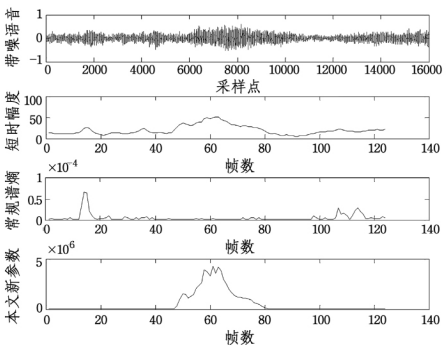

图 6　混有－5dB 背景说话噪声时 3 种方法的检测结果

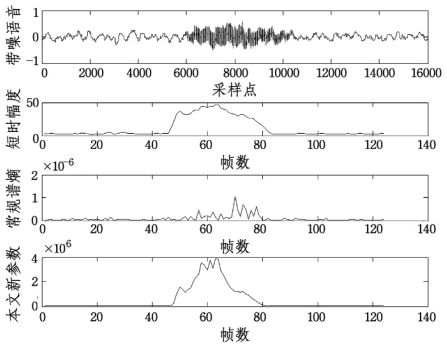

图 7　混有－5dB 高速行驶汽车噪声时 3 种方法的检测结果

从以上仿真结果可以看出，在无背景噪声环境下，3 种方法的检测效果都很好；当混入－5dB 的不同噪声时，常规谱熵法基本失效，因为该算法在求得每一帧语音信号各频率段的功率谱和该帧总的功率谱后，需要计算所对应的各个频率段的概率，而非平稳噪声的频谱概率分布是不均匀的，这就是导致非平稳噪声环境下对应谱熵的检测结果比较差的主要原因。短时平均幅度法由于只应用了语音信号时域中的某一特征，没有考虑其他的相关特征，也不能应对各种噪声环境。而本文提出的检测方法利用非平稳信号处理技术将语音信号的时域分析和频域分析相结合，检测的鲁棒性明显增强，在不同背景噪声环境下的检测效果都比较理想。

结束语　本文结合语音的短时平均幅度，通过对常规谱熵端点检测算法进行改进，提出了一个新的特征参数——子带谱熵幅度积。仿真实验结果表明，利用这个新参数进行语音端点检测相对于短时平均幅度法和谱熵法能更有效地从背景噪声中检测出含噪语音信号的起止端点，提高了检测率，且稳定性更强。该方法原理简单，易于实现，在保证运算量没有显著增大的前提下能提高端点检测的性能。

参 考 文 献

[1] Ouzounov A. Noisy speech endpoint detection using robust feature [M] // Biometric Authentication. New York: Spring International Publishing, 2014: 105-117

[2] Ouzounov A. Telephone speech endpoint detection using mean-delta feature [J]. Cybernetics and Information Technologies, 2014, 14(2): 127-139

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

[4] 韦国刚，周萍，杨青. 一种简单的噪声鲁棒性语音端点检测方法[J]. 测控技术，2015，34(2)：31-34

[5] 朱恒军，于泓博，王发智. 小波分析和支持向量机相融合的语音端点检测算法[J]. 计算机科学，2012，39(6)：244-246，265

[6] 蔡莲红，黄德智，蔡锐. 现代语音技术基础与应用[M]. 北京：清华大学出版社，2003：26-29

[7] 赵新燕，王炼红，彭林哲. 基于自适应倒谱距离的强噪声语音端点检测[J]. 计算机科学，2015，42(9)：83-85

[8] 胡光锐，韦晓东. 基于倒谱特征的带噪语音端点检测[J]. 电子学报，2000，28(10)：95-97

[9] Zhao H, Zhao L X, Zhao K, et al. Voice activity detection based

on distance entropy in noisy environment[C]// 5th International Joint Conference on INC, IM S, and IDC. Seoul, Korea: IEEE Computer Society, 2009: 1364-1367

[10] 朴春俊,马静霞,徐鹏. 带噪语音端点检测方法研究[J]. 计算机应用,2006,26(11):2685-2686

[11] 张宇波,邢立钊. 基于小波分析与 PSO-ELM 的语音端点检测算法研究[J]. 中北大学学报(自然科学版),2016,37(1):33-37

[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] 柳春. 一种改进的基于短时平均幅度的语音端点检测算法研究[J]. 西北民族大学学报(自然科学版),2009,30(73):56-59

[14] Shen J L, Hung J W, Lee L S. Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments [C]// ICSLP(S0160-5840). Sydney, Australia, 1998: 232-235

[15] Wu Bing-fei, Wang K C. Robust Endpoint Detection Algorithm Based on the Adaptive Band-Partitioning Spectral Entropy in Adverse Environments[J]. IEEE Transactions on Speech and Audio Processing (S1063-6676), 2005, 13(5): 762-775

(上接第 209 页)

实验结果如表 1—表 3 所列,对于 Pepper 图片,图 1 给出了所提算法在不同噪声下的去噪效果。从实验结果可以看出,算法相对其他算法有优势,尤其在大噪声的情况下。

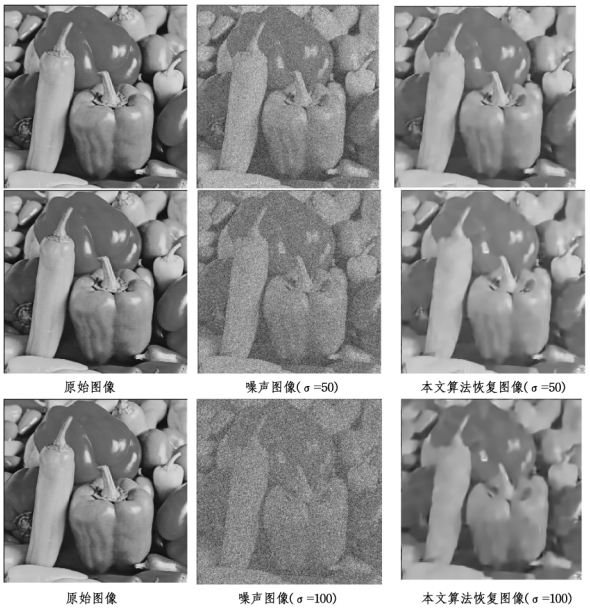

图 1 所提算法在不同噪声下的去噪效果

结束语 本文对传统的基于稀疏表示的图像去噪模型进行改进,将非局部自相似性和稀疏系数的权重融入到稀疏表示模型中,提出了一种新的图像去噪模型,并给出该模型的优化算法。仿真结果表明,所提模型在很多情况下较其他一些先进的模型具有优势。

参 考 文 献

[1] Tomasi C, Manduchi R. Bilateral filtering for gray and color images[C]// ICCV. 1998: 839-846

[2] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(7): 629-639

[3] Osher S, Burger M, Goldfarb D, et al. An iterative regularization method for total variation-based image restoration[J]. Multiscale Modeling & Simulation, 2005, 4(2): 460-489

[4] Starck J L, Candes E J, Donoho D L. The curvelet transform for image denoising[J]. IEEE Transactions on Image Processing, 2002, 11(6): 670-684

[5] Dong W, Zhang L, Shi G, et al. Nonlocally centralized sparse representation for image restoration[J]. IEEE Transactions on Image Processing, 2013, 22(4): 1620-1630

[6] Ji H, Liu C, Shen Z, et al. Robust video denoising using low rank matrix completion[C]// CVPR. 2010: 1791-1798

[7] Donoho D L. De-noising by soft-thresholding[J]. IEEE Trans. Inf. Theor., 1995, 41(3): 613-627

[8] Chang S G, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression[J]. IEEE Transactions on Image Processing, 2000, 9(9): 1532-1546

[9] Portilla J, Strela V, Wainwright M J, et al. Image denoising using scale mixtures of Gaussians in thewavelet domain[J]. IEEE Transactions on Image Processing, 2003, 12(11): 1338-1351

[10] Weiss Y, Freeman W T. What makes a good model of natural images? [C]// CVPR. 2007: 1-8

[11] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. Physica D: Nonlinear Phenomena, 1992, 60(1): 259-268

[12] Osher S, Burger M, Goldfarb D, et al. An iterative regularization method for total variation-based image restoration[J]. Multiscale Modeling & Simulation, 2005, 4(2): 460-489

[13] Schmidt U, Roth S. Shrinkage fields for effective image restoration[C]// CVPR. 2014: 2774-2781

[14] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]// CVPR. 2005: 60-65

[15] Peyre G, Bougleux S, Cohen L D. Nonlocal regularization of inverse problems[J]. Inverse Problems and Imaging, 2011, 5(2): 511-530

[16] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on Image Processing, 2007, 16(8): 2080-2095

[17] Mairal J, Bach F, Ponce J, et al. Nonlocal sparse models for image restoration[C]// ICCV. 2009: 2272-2279

[18] Wang S, Zhang L, Liang Y. Nonlocal spectral prior model for low-level vision[C]// ACCV. 2013: 231-244

[19] Dong Wei-sheng, Li Xin, Zhang Lei, et al. Sparsity-based image denoising via dictionary learning and structure clustering[C]// CVPR. 2011: 457-464

[20] Gu S, Zhang L, Zuo W, et al. Weighted nuclear norm minimization with application to image denoising[C]// CVPR. 2014: 2862-2869

[21] Xu J, Zhang L, Zuo W, et al. Patch Group Based Nonlocal Self-Similarity Prior Learning for Image Denoising[C]// ICCV. 2015