

基于自适应倒谱距离的强噪声语音端点检测

赵新燕 王炼红 彭林哲

(湖南大学电气与信息工程学院 长沙 410082)

摘要 在有噪声干扰的情况下,传统的语音端点检测方法的检测准确度明显下降。为了在强背景噪声环境下有效区分出语音信号和非语音信号,针对倒谱距离端点检测方法进行了研究,提出了一种基于自适应倒谱距离的强噪声语音端点检测方法。该方法引入倒谱距离乘数和门限增量系数,针对不同信噪比采用不同的倒谱距离乘数,并采用自适应判决门限的方法进行语音端点检测。MATLAB 仿真实验结果显示,在不同背景噪声和不同信噪比下,该方法对于语音端点检测具有较高的检测正确率,其端点检测效果明显优于传统端点检测方法,适用于强背景噪声下的端点检测。

关键词 倒谱距离,自适应判决门限,强噪声,端点检测

中图分类号 TP301.6 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.9.017

Adaptive Cepstral Distance-based Voice Endpoint Detection of Strong Noise

ZHAO Xin-yan WANG Lian-hong PENG Lin-zhe

(College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract In the case of noise interference, accuracy of speech endpoint detection using the traditional method dramatically declines. In order to effectively distinguish the speech signal and non-voice signal in strong background noise environment, this paper presented a strong noise speech endpoint detection method based on adaptive cepstral distance. The method introduces cepstral distance multiplier and the threshold increment coefficient. Different cepstral distance multipliers are used for different SNR and adaptive decision threshold method is used for voice activity detection. MATLAB simulation results show that, under different background noise and different SNR, the method for voice activity detection has high detection accuracy. Its detection is better than the traditional endpoint detection method, and is suitable for endpoint detection under strong background noise.

Keywords Cepstral distance, Adaptive decision threshold, Strong noise, Endpoint detection

1 引言

语音端点检测的实质是利用语音与噪声对于相同参数的不同特征表现来对它们进行区分。传统的语音端点检测算法主要有诸如短时能量、短时平均过零率等,并采用门限判定法来检测语音,此类方法原理简单、运算方便,在高信噪比环境下可以获得较好的检测效果,但对于复杂背景噪声环境下的端点检测误判率升高^[1-3]。还有一些检测方法虽然检测准确性好、稳健性强,但方法自身复杂度高、运算量大,如基于统计模型和模式分类的方法、基于神经网络模型的方法等,这类检测方法不便于语音的实时处理,因此在实时语音信号系统中很少采用^[4-6]。

然而实际的语音识别系统常用于不同的检测环境,背景噪声复杂多变,信噪比很低,传统的检测方法就变得不再适用。倒谱特征能很好地表示语音的特征^[7],采用倒谱系数作为特征参数很有优势,但在信噪比不断变换的情况下其适

性有所下降。因此,本文提出一种基于自适应倒谱距离的强噪声语音端点检测方法。先对语音信号采用带通滤波,尽量保留低频大部分能量,同时又去除工频 50Hz 噪声带来的影响,再对滤波后的带噪语音信号求倒谱距离,然后针对不同信噪比采用不同的倒谱距离乘数,并采用自适应判决门限的方法进行语音端点检测。用 MATLAB 做仿真实验,结果表明,在不同背景噪声和多变信噪比下,该方法对于语音端点检测都具有较高的检测正确率,相比传统端点检测方法适应性更强,检测正确率更高。

2 倒谱距离

2.1 倒谱距离定义

信号倒谱的一种定义是信号的能量谱密度函数 $S(\omega)$ 的对数值的逆傅里叶变换。对于一对谱密度函数 $S(\omega)$ 与 $S'(\omega)$,依据 Parseval 定理^[8,9],用谱的倒谱距离表示对数谱的均方距离:

到稿日期:2014-09-01 返修日期:2014-10-22 本文受国家自然科学基金(61174140),湖南省自然科学基金(14JJ4026)资助。

赵新燕(1988-),女,硕士,主要研究方向为语音图像传输与处理、嵌入式系统及应用,E-mail: 1033926587@qq.com;王炼红(1971-),女,副教授,主要研究方向为语音图像传输与处理、数据挖掘技术、嵌入式系统及应用、现代通信技术;彭林哲(1990-),男,硕士,主要研究方向为现代通信网络技术、嵌入式系统及应用。

$$d_{cep}^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |\log S(\omega) - \log S'(\omega)| d\omega = \sum_{n=-\infty}^{+\infty} (c_n - c_n')^2 \quad (1)$$

式中, c_n 和 c_n' 分别为谱密度 $S(\omega)$ 与 $S'(\omega)$ 的倒谱系数, 在不同声学单元或者是有声段和无声段之间跳转的时候, 相邻帧之间的倒谱差异较大, 所以采用倒谱距离作为语音端点检测的一个特征参数^[10]。

2.2 传统倒谱距离端点检测步骤

1) 预处理: 预加重 8kHz 的采样信号, 然后做分帧加窗处理。

2) 估计噪声倒谱系数: 将前几帧信号假设为背景噪声, 计算这些帧的倒谱系数, 利用前几帧倒谱系数的平均值作为背景噪声的倒谱系数, 噪声倒谱系数的近似值按如下规则不断更新:

$$c_n' = \alpha c_n' + (1 - \alpha) c_n' \quad (2)$$

式中, α 为更新因子, c_n' 为上一帧噪声的倒谱系数。

同时采用式(1)计算当前帧的后 5 帧倒谱距离的平均值, 将其作为背景噪声倒谱距离的估计值。

3) 计算每帧的倒谱距离 d_{cep} : 计算每帧倒谱系数, 并结合噪声倒谱系数的估计值, 利用式(1)计算倒谱距离。

4) 确定判决门限: 设置两个门限 G_1 和 G_2 :

$$G_i = k_i d_{cep}, i = 1, 2 \quad (3)$$

式中, d_{cep} 为倒谱距离的估计值, k_1, k_2 分别为判决门限的两个乘系数, 且为了满足 $G_1 < G_2$, 要求 $k_1 < k_2$, 这里取 $k_1 = 1.5, k_2 = 2.0$ 。

5) 根据倒谱距离 d_{cep} 绘制倒谱距离轨迹, 采用类似于短时能的检测方法, 利用倒谱距离检测语音段和噪声段。

倒谱对于语音的特征具有很好的表征, 在强噪声环境下倒谱系数常用来作为端点检测的特征量^[11]。传统倒谱距离端点检测方法在恒定信噪比下检测正确率与其他方法相比具有明显的优势, 但对于信噪比不断变化的复杂环境下的端点检测效果不太理想, 因此有必要对传统方法进行改进, 使其适应多变的噪声环境。

3 自适应倒谱距离端点检测算法

3.1 算法改进

在用传统倒谱距离检测语音端点的实验中观察到, 信噪比较小时, 噪声倒谱与带噪语音倒谱相差不大, 此时的倒谱距离小, 语音段与噪声段的倒谱距离虽仍有差别, 但是差别不大、难以区分, 与设定的倒谱距离门限值相比较容易导致误判。同样, 传统的倒谱距离检测方法中门限乘系数恒定, 调整出的门限值难以同时适应低信噪比和高信噪比环境, 所以也需要建立随信噪比 SNR 变化的检测门限来提高语音端点检测的准确性。因此, 本算法提出自适应倒谱距离, 引入倒谱距离乘数; 针对不同信噪比采用不同的倒谱距离乘数; 同时, 在恒定门限乘系数的基础上添加门限增量系数, 高信噪比时采用高门限, 低信噪比时采用低门限, 仍采用双门限判决。

改进的算法的基本思想如下:

1) 逐帧估计信噪比 SNR。采用经典的直接判决法 (D-D)^[12], 逐帧估计短时信噪比 SNR_{prio} , 其表达式如下:

$$SNR_{prio}^k(i) = \eta \frac{|\hat{X}_k(i-1)|^2}{|\hat{D}_k(i-1)|^2} + (1 - \eta) \cdot \max[SNR_{post}^k(i) - 1, 0] \quad (4)$$

式中, $\hat{X}_k(i-1)$ 表示上一帧纯净语音信号谱分量估计值; $\hat{D}_k(i-1)$ 表示上一帧噪声功率谱分量估计值, 可以由语音间歇的静音帧估计得到; $SNR_{prio}^k(i)$ 为先验信噪比; $SNR_{post}^k(i)$ 为后验信噪比; η 为平滑因子, 一般在 0.8~1 之间取值。

2) 由信噪比 SNR 计算倒谱距离乘数 β 。

首先, 自定义自适应倒谱距离为:

$$d'_{cep} = \beta d_{cep} \quad (5)$$

式中, β 为倒谱距离乘数, d_{cep} 为传统方法中的倒谱距离, d'_{cep} 为新的自适应倒谱距离。

为提高高低信噪比情况下倒谱距离的适应性并且使倒谱距离轨迹尽量平滑, 进行了大量实验数据规整建模, 最后得出, 倒谱距离乘数 β 与信噪比 SNR 的关系式为式(6)时, 检测效果最佳。

$$\beta = \begin{cases} \ln(70 - SNR)/4, & -25 \leq SNR \leq 25 \\ 1, & SNR > 25 \\ 1.7, & SNR < -25 \end{cases} \quad (6)$$

3) 由信噪比 SNR 确定判决门限。判决门限在恒定门限乘系数的基础上添加门限增量系数, 重新定义判决门限与信噪比的关系式为:

$$G_i = \begin{cases} k_i d_{cep} + \gamma SNR, & -25 \leq SNR \leq 25 \\ k_i d_{cep}, & \text{others} \end{cases} \quad (7)$$

式中, d_{cep} 为倒谱距离的估计值, k_1, k_2 分别为判决门限的两个乘系数, 且为了满足 $G_1 < G_2$, 要求 $k_1 < k_2$, 这里取 $k_1 = 1.5, k_2 = 2.0$; γ 为门限增量系数, 此处经大量实验后设置为 $\gamma = 0.07$ 。

3.2 算法实现流程

根据上节中提出的改进的算法思想, 采用自适应倒谱距离和可更新倒谱距离门限对带噪语音信号进行端点检测, 以提高低信噪比、复杂噪声环境下检测的准确率。在语音端点检测前先对带噪语音信号进行预处理, 依据语音信号的短时平稳特性对信号进行分帧加窗处理。一般, 语音信号的能量集中在 60~3400Hz 频段^[11], 采用带通滤波器对带噪语音进行预处理, 滤波频率大于 60Hz 即可以在尽量保留低频能量的同时去除工频 50Hz 噪声带来的影响。

改进的端点检测方法的检测流程如图 1 所示。

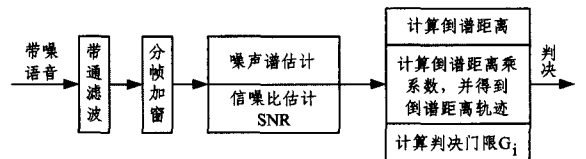


图 1 改进的自适应倒谱距离端点检测流程

假设 $y(n), x(n), d(n)$ 分别为经带通滤波后的带噪语音信号、纯净的语音信号、噪声信号, 则有 $y(n) = x(n) + d(n)$, n 为采样点, 具体步骤如下:

1) 预处理。对 8kHz 采样信号进行带通滤波处理, 通带为 60~3400Hz; 然后分帧加窗, 一般地, 帧长选取 14~30ms 为宜, 本算法采用帧长 25ms, 即 200 个采样点; 帧移 12.5ms, 即 100 个采样点, 则有 $y(i, k) = x(i, k) + d(i, k), k = 1, 2, \dots, 200$, 式中 $y(i, k), x(i, k), d(i, k)$ 分别表示分帧加窗后第 i 帧的第 k 个采样点对应的信号帧。

2) 噪声谱估计。假定前几帧信号是无声段, 用前 5 帧信

号谱的平均值作为噪声谱的估计值。在无声段噪声谱的近似值仍按式(2)规则进行更新,而当检测到语音端点的时候为减小噪声谱估计的误差,需放缓噪声谱的估计;检测到语音段终点时,重新开始更新噪声倒谱,用当前段语音终点的后一帧带噪语音谱重新估计噪声谱,并按照之前的速度更新噪声谱。

3)逐帧估计信噪比 SNR。假定前几帧无声段信号的信噪比为 0dB,用当前帧带噪语音谱减去噪声谱估计值作为当前帧语音信号谱估计值,并按照式(4),采用递推的方法逐帧估计信噪比 SNR_{prio} 。

4)计算每帧的倒谱距离 d_{xp} 。计算每帧倒谱系数,再由每帧信号的倒谱系数和噪声倒谱系数的估计值通过式(1)计算倒谱距离。

5)计算倒谱距离乘数 β 。将步骤第 3)中得到的信噪比估计值 SNR_{prio} 代入关系式(6),得到倒谱距离乘数 β ,将 β 代入关系式(5)计算得到自适应的倒谱距离,并绘制倒谱距离轨迹。

6)计算自适应判决门限。将步骤 3)中得到的信噪比估计值 SNR_{prio} 代入关系式(7),得到自适应判决门限。

7)端点检测。利用自适应倒谱距离和自适应判决门限进行当前帧的端点检测,并不断循环执行步骤 2)–6),在检测到语音起始点和终点时对相应系数进行记录和更新。

4 实验仿真及分析

本文提出的算法采用 MATLAB 软件进行仿真,实验采用的语音信号为“我到北京去”,如图 2 所示,经 8kHz 采样和 16bit 量化处理,wav 格式存储为纯净语音样本。以分别加入不同信噪比的白噪声和粉色噪声作为样本进行测试,与用传统的倒谱距离做端点检测做对比。

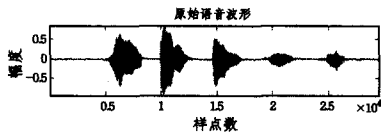


图 2 原始语音信号波形

图 3、图 4 分别为加入 30dB 高斯白噪声和 30dB 粉色噪声的含噪语音,以及分别采用传统倒谱距离和改进的自适应倒谱距离检测方法进行端点检测的结果。

从图 3、图 4 对比可看出,在同样信噪比情况下,改进的算法与传统倒谱距离算法对于语音的端点检测都有很好的检测准确性,但改进算法的处理结果比传统算法的效果稍好。

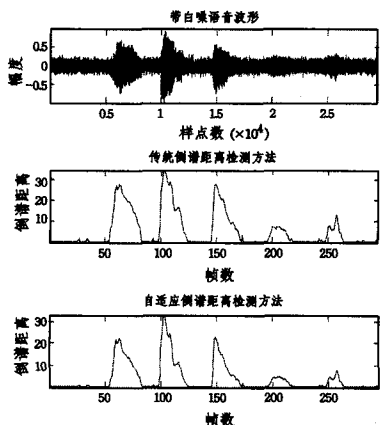


图 3 含 30dB 白噪声的语音

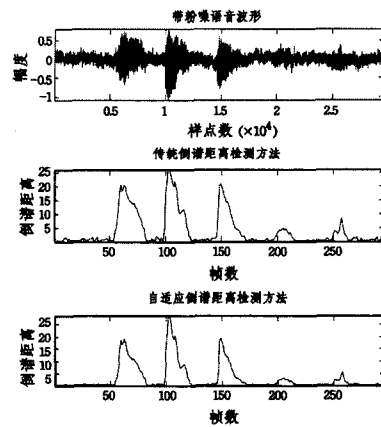


图 4 含 30dB 粉色噪声的语音

图 5 为同一段原始语音信号分别加入不同信噪比 (30dB/5dB/20dB) 的高斯白噪声后利用传统倒谱距离和改进的倒谱距离端点检测方法进行端点检测的效果。

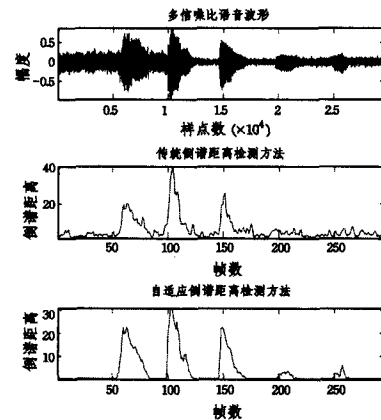


图 5 加入不同信噪比的白噪声

从图 5 中可以看出,对于不同时间段加入不同信噪比噪声的语音信号,传统的倒谱距离端点检测算法在信噪比骤变的情况下因为倒谱距离适应性差而致使检测结果严重偏离,导致误判;而本文提出的自适应倒谱距离端点检测方法则仍然具有较好的检测效果,能更好地适应多变的强噪声环境。

现采用一段公交车上嘈杂环境下的语音信号“下一站,荫马塘”进行实验仿真,并与用传统的倒谱距离方法做端点检测进行对比。信号经 8kHz 采样和 16bit 量化处理,wav 格式存储为纯净语音样本,仿真效果如图 6 所示。

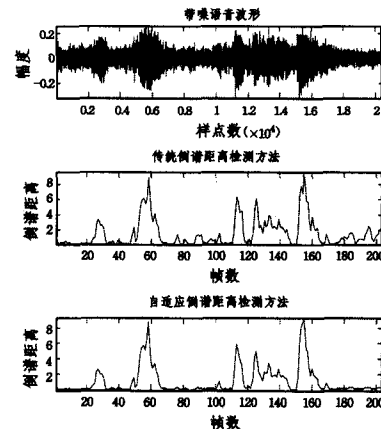


图 6 带噪语音端点检测

Services [C]//Proc. of the 3rd International Conf. on Business Process Management (BPM 2005). LNCS 3649, Berlin, Heidelberg; Springer-Verlag, 2005; 109-119

- [23] 林闯. 随机 Petri 网和系统性能评价(第 2 版)[M]. 北京:清华大学出版社, 2005
Lin Chuang. Stochastic petri net and performance evaluation of system(2nd Edit) [M]. Beijing: TsingHua University Press,

(上接第 85 页)

从图 6 对比可看出,在真实的嘈杂环境中,自适应倒谱距离算法的端点检测效果要优于改进的倒谱距离向量算法。

表 1 给出了两种算法在不同信噪比环境下通过多次实验后的实验结果。

表 1 两种方法下带噪声语音端点检测的检测正确率

SNR	白噪声		粉色噪声	
	Dcep	Dcep1	Dcep	Dcep1
-5dB	82.5%	91.0%	81.7%	90.2%
0dB	85.1%	92.4%	84.6%	91.7%
5dB	90.9%	95.3%	90.4%	94.9%
15dB	94.2%	98.9%	93.5%	98.1%
多信噪比	40.2%	87.4%	38.3%	86.7%

表 1 中, Dcep 表示传统倒谱距离端点检测方法, Dcep1 表示改进的自适应倒谱距离端点检测方法。从表中可以看出,改进的倒谱距离在单信噪比环境下的语音端点检测正确率要高于传统倒谱距离检测算法;在多信噪比环境下的端点检测效果明显高于传统倒谱距离检测算法。总的来说,改进的自适应倒谱距离端点检测算法明显优于传统的倒谱距离检测算法。

结束语 本文在传统倒谱距离端点检测的基础上引入倒谱距离乘数和门限增量系数,提出了自适应的倒谱距离语音端点检测方法,针对不同信噪比采用不同的倒谱距离乘数,并采用自适应判决门限的方法进行语音端点检测。用 MATLAB 软件进行仿真,在纯净语音中添加白噪声和粉噪声以及在同一段语音中添加多信噪比噪声进行实验,同时还采用了大量嘈杂环境下的实际语音信号进行实验仿真。实验结果表明,在不同背景噪声和多变信噪比环境下,与传统倒谱距离检测方法相比较,该方法检测正确率更高且运算量不大,更适用于复杂多变的强背景噪声环境下的语音端点检测。

在信噪比低于 -20dB 时,采用改进的倒谱距离波形用于端点检测,强度较弱的语音信号或清音信号仍然会出现丢失或是误检现象,并且清音检测的平滑度不够好,此问题将是下一步研究的重点。

参 考 文 献

- [1] 李晋,刘甫,王玲,等.改进的语音端点检测技术[J].计算机工程与应用,2009,45(24):133-135,149
Li Jing, Liu Fu, Wang Ling, et al. Improved technology for speech endpoint detection[J]. Computer Engineering and Applications, 2009, 25(24): 133-135, 149
- [2] 吕钊.噪声环境下的语音识别算法研究[D].合肥:安徽大学, 2011
Lv Zhao. Research on speech recognition algorithm in noisy environment[D]. Hefei: Anhui University, 2009
- [3] 张婷,何凌,黄华,等.基于临界频带及能量熵的语音端点检测

2005

- [24] 何炎祥,沈华.一种基于随机 Petri 网的 Web 服务组合性能瓶颈定位策略[J].计算机学报,2013,36(10):1953-1966
He Yan-xiang, Shen Hua. A Stochastic Petri Net-Based Performance Bottleneck Location Strategy for Web Services Composition [J]. Chinese Journal of Computers, 2013, 36(10): 1953-1966
- [J]. 计算机应用, 2013, 33(1): 175-178
Zhang Ting, He Ling, Huang Hua, et al. Speech endpoint detection based on critical band and energy entropy[J]. Journal of Computer Applications, 2013, 33(1): 175-178
- [4] Li Jie, Zhou Ping, Jing Xin-xing, et al. Speech endpoint detection method based on TEO in noisy environment[J]. Procedia Engineering, 2012, 29(1): 2655-2660
- [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
Zhu Heng-jun, Yu Hong-bo, Wang Fa-zhi. Speech Endpoints Detection Algorithm Based on Support Vector Machine and Wavelet Analysis[J]. Computer Science, 2012, 39(6): 244-246, 265
- [6] Wei Xiao-dong, Hu Guang-rui, Ren Xiao-lin. Endpoint detection of noisy speech by the use of cepstrum [J]. Engineering Village, 2014, 34(2): 185-188
- [7] 胡光锐,韦晓东.基于倒谱特征的带噪声语音端点检测[J].电子学报,2000,28(10):95-97
Hu Guang-rui, Wei Xiao-dong. Endpoint Detection of Noisy Speech Based on Cepstrum[J]. Acta Electronica Sinica, 2000, 28(10): 95-97
- [8] 沈红丽,曾毓敏,王鹏.一种改进的基于倒谱特征的带噪声语音端点检测方法[J].通信技术,2009,42(2):156-158
Shen Hong-li, Zeng Yu-min, Wang Peng. An Improved Speech Endpoint Detection Method based on Cepstrum Distance [J]. Communications Technology, 2009, 42(2): 156-158
- [9] 王博,郭英,李宏伟,等.基于倒谱距离的语音端点检测改进算法[J].空军工程大学学报(自然科学版),2006,7(1):59-63
Wang Bo, Guo Ying, Li Hong-wei, et al. An Improved Voice Activity Detection Method Based on Cepstrum Distance [J]. Journal of Air Force Engineering University (Natural Science Edition), 2006, 7(1): 59-63
- [10] 缪涛,侯丽敏.基于倒谱距离窗移最小失真分割的语种辨识[J].上海大学学报(自然科学版),2007,13(2):116-120
Miu Wei, Hou Li-min. Language Identification Based on Minimum Distortion of Cepstrum Distance Segmentation [J]. Journal of Shanghai University (Natural Science Edition), 2007, 13(2): 116-120
- [11] 于迎霞,史家茂.一种改进的基于倒谱特征的带噪声端点检测方法[J].计算机工程,2004,30(19):85-87
Yu Ying-xia, Shi Jia-mao. A Modified Endpoint Detection Method of Noisy Speech Based on Cepstral [J]. Computer Engineering, 2004, 30(19): 85-87
- [12] 蒋建中,张东方,张连海.一种新的强噪声环境下的语音增强算法[J].计算机工程与应用,2013,49(20):222-225
Jiang Jian-zhong, Zhang Dong-fang, Zhang Lian-hai. Speech enhancement algorithm for high noise environment [J]. Computer Engineering and Applications, 2013, 49(20): 222-225