基于临界带特征矢量距离的端点检测算法

武文娟 顾宏斌 潘秀林
(南京航空航天大学民航学院 南京 210016)

摘 要 端点检测是语音数字信号处理中一个重要的环节。在前人研究的基础上提出了一种新的基于临界带特征矢量距离的端点检测方法,由计算得到的每帧各临界带中的功率谱之和作为特征矢量,并且通过计算各帧之间的矢量距离得到其距离轨迹,以此设定门限进行语音端点的检测。对比实验表明,相对于基于谱熵的算法及基于倒谱距离的算法,本方法具有更好的鲁棒性和较高的正确率。

关键词 端点检测,临界带,特征矢量距离
中图法分类号 TN912.34 **文献标识码** A

Voice Activity Detection Method Based on Selected Sub-bands Vector Distance

WU Wen-juan GU Hong-bin PAN Xiu-lin
(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract Voice activity detected (VAD) is a very important step for speech signals processing, a new method of the VAD was proposed based on the selected sub-bands vector distance. It calculates the summation of power spectrum in very frame as characteristic vector. For detecting the voice activity, the method calculates the characteristic vector distances between frames to acquire the distance track. Theoretical analysis and experimentation show that compared with entropy-based algorithm and cepstrum-based algorithm this method has better robustness and higher correctness.

Keywords Voice activity detected, Selected sub-bands, Characteristic vector distance

语音作为人类信息交流的最自然、最有效、最灵活而又最广泛的使用途径,理应成为未来人机交互的主要方式。语音的端点检测作为语音通信、语音合成、语音识别和语音增强中的一个重要环节,直接影响到后续工作的质量。正确确定语音段端点,不仅可以减少计算量,而且可以提高上述系统的正确率。端点检测是指从包含语音的一段信号中确定出语音的起始点及结束点。有效的端点检测不仅可以减少系统的处理时间,提高系统的处理实时性,而且能排除噪声的干扰,使后续的识别性能大大提高。所以,语音端点检测至今仍是一个值得深入研究的课题。

目前,研究者们提出了许多语音端点检测的算法,例如基于语音短时能量、短时平均幅度和过零率、基于谱熵、基于倒谱距离等。其中很多方法在强噪声环境下无法检测到准确的端点,尤其是在非恒定噪声环境下[1]。为了提高端点检测的正确率,选择合适的声学特征也至关重要。本文尝试性地提出了基于临界带矢量特征的端点检测方法。试验表明,在低信噪比情况下,该方法仍有较好的端点检测能力。

1 常用端点检测方法

由于短时能量与过零率检测方法比较简单,因此最常被人们采用。但其并非是实时化的算法,而且它需要大量存储空间存储语音原始数据,使得其端点检测稳定性大大下降[2]。

基于倒谱距离的端点检测方法是以倒谱系数作为参数的,测量方法步骤类似于基于能量的端点检测,但将倒谱距离代替短时能量来作为门限。信号的复倒谱定义为信号能量密度谱函数 $S(w)$ 的对数的傅立叶级数,倒谱系数为

$$c_0=\frac{1}{2\pi}\int_{-\pi}^{+\pi}\log S(w)\mathrm{d}w \tag{1}$$

对于一对谱密度函数 $S(w)$ 与 $S'(w)$,利用 Parseval 定理,用谱的倒谱距离表示对数谱的均方距离,而对数谱的均方距离表示两个信号谱的差别[3],故可用来作为一个判决参数。但是,在低信噪比的情况下,这种方法的检测的正确率大大下降。

基于谱熵的端点检测方法是引入了信息论中的熵函数[4]。先求信号的短时功率谱,在此基础上定义和计算概率密度函数,进而得出信息熵,然后通过熵值大小来区分语音段和无声段。每帧语音信号的信息熵定义为:

$$H=-\sum_{i=1}^{N/2+1}p_i\log p_i \tag{2}$$

其中,p_i 为每个频率分量的归一化谱概率密度函数,且 $P_i=0$ 时,$P_i\log P_i=0$[5]。

这种方法,在高信噪比时,噪声的谱熵变化比较平缓,容易确定阈值、切分语音;而在低信噪比时,噪声的谱熵起伏变大,而语音信号的谱熵突变不明显,尤其是在噪声与语音的边

到稿日期:2008-03-04 本文受民航总局科技基金项目资助(E9905)资助。
武文娟(1983—),女,硕士,主要研究领域为文语转换、人工智能,E-mail:wwjfengyun@yahoo.com.cn;顾宏斌(1957—),男,教授,博士生导师,主要研究领域为航空器仿真、虚拟现实技术、基于计算机的培训;潘秀林(1982—),男,硕士,主要研究领域为文语转换、机器学习。

缘地带，很难确定一个准确的阈值，将语音与噪声切分开来。

2 基于临界带矢量特征距离的端点检测算法

2.1 临界带特征矢量方法的提出

后来提出的方法，基本上都是基于信号短时谱的，如前面提到的基于谱熵及基于倒谱距离的端点检测，都是利用短时傅立叶变换求取的语音信号的短时谱或倒谱。短时谱是按实际频率分布的，而符合人耳听觉特性的频率分布应该是按临界带频率分布的。所以，如果按实际频率分布的频率作为语音特征，由于它不符合人耳的听觉特征，将会降低语音信号处理系统的性能。

本文所阐述的方法，正是在临界带频谱的基础上提出的，首先要求得临界带特征矢量。临界带特征矢量是指：将一帧信号的功率谱按频率高低分成若干个临界带，对每个临界带中的功率谱求和，即可得到相应的临界带特征矢量。

每一帧信号都对应一个若干维的临界带特征矢量，因此，无论语音帧与噪声帧都对应于不同的临界带特征矢量，我们可以用欧式距离来计算特征矢量的畸变。而且，噪声帧与噪声帧之间的距离要远远小于语音帧与噪声帧之间的距离。正是基于这种思想，我们提出了基于临界带特征矢量距离的端点检测算法。

2.2 临界带特征矢量算法的过程

第一步，求出每一帧的加窗语音 $X_n(m)$：$m=0\sim(N-1)$ 的 DFT 的模平方值 $|X_n(k)|^2$，此即为功率谱。本文中，做 512 点的 DFT 变换，采样频率 f_s 为 8kHz，窗长为 38ms(即 $N=300$)，窗形为汉明窗。$|X_n(k)|^2$ 与原始加窗语音信号的频谱模平方 $|X_n(\exp(j\omega_k))|^2$ 具有下列关系：

$$|X_n(k)|^2=|X_n(\exp(j\omega_k))|^2 \quad 0\sim511 \tag{3}$$

其中，$\omega_k=2\pi f_k$，$f_k=\frac{f_s}{512}k$

第二步，划分临界带。在 $0\sim f_s/2$ 中确定 $\hat{f}_1,\hat{f}_2,\hat{f}_3\cdots$ 若干个临界带频率分割点。确定的方法是将 $i=1,2,3\cdots$ 代入式(4)，即可求出相应的 $\hat{f}_i$(以 Hz 为单位)。

$$i=\frac{26.81\hat{f}_i}{1960+\hat{f}}-0.53 \tag{4}$$

由此可以求出 $\hat{f}_1,\hat{f}_2,\hat{f}_3\cdots$，并且，由 $\hat{f}_1$ 与 $\hat{f}_2$ 构成第一临界带，$\hat{f}_2$ 与 $\hat{f}_3$ 构成第二临界带，以此类推。此处，在 0.1 至 4kHz 范围内需要安排 16 个临界带[6]。

第三步，求临界带特征矢量。将每一个临界带中的功率谱 $|X_n(k)|^2$ 取和，即可得到相应的临界带特征矢量。如果用 $H=[h_1,h_2,\cdots,h_l,\cdots,h_L]$ 表示临界带特征矢量，则每一个分量可通过式(5)求得。从而，每帧都可以得到一个十六维的临界带特征矢量：

$$h_L=\sum_{\hat{f}_l<\hat{f}_k\leqslant\hat{f}_{l+1}}|X_n(k)|^2 \tag{5}$$

第四步，求临界带特征矢量距离。假设前几帧信号是背景噪声(这里取前 4 帧)，将这几帧的临界带特征矢量求均值，即可得到噪声的平均特征矢量值。并且，利用式(5)对每一帧的特征矢量求其与噪声平均特征矢量的均方距离，即可得到特征矢量距离轨迹：

$$d^2_{cep}=\sum(c_i-c_0)^2 \tag{6}$$

其中，c_i 表示当前帧的临界带特征矢量，c_0 表示噪声的平均特征矢量值。

第五步，设定阈值，进行端点检测。设定一个阈值 D，逐帧进行比较：如果第 $i-1$ 帧、第 $i-2$ 帧的特征矢量距离都小于 D，而第 $i+1$ 帧、第 $i+2$ 帧的特征矢量距离都大于 D，我们就认为第 i 帧为语音段的起始位置。同样，如果第 $i-1$ 帧、$i-2$ 帧的特征矢量距离都大于 D，而第 $i+1$ 帧、第 $i+2$ 帧的特征矢量距离都小于 D，我们就认为第 i 帧为语音段的结束位置。

3 实验结果及性能分析

通过实验来说明本文方法的有效性，与文献中的基于倒谱距离的方法和基于频谱谱熵的方法进行比较。实验的共同条件为：采样频率 8kHz，16bit 量化，帧长 38ms，帧间重叠 50%，即每帧有 300 个点，帧移为 150 个点。

图 1 为“亚洲国家”的原始语音波形；图 2 为原始语音波形基础上加入噪声后的波形。下面，就不同信噪比情况，对本文提出的方法与上述两种方法进行对比。

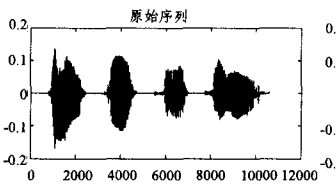


图 1 原始语音波形

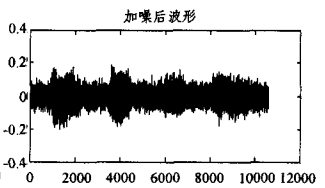


图 2 加噪后波形

(1)同基于倒谱距离检测方法的实验对比

本文提出的基于临界带矢量特征的方法和基于倒谱距离的方法，都用到了矢量之间的距离，而文献中的矢量距离也是计算其欧式距离。这里，我们对两种方法进行了比较。

由图 3 和图 4 可以看出，在高信噪比情况下，两种方法中语音信号与噪声的倒谱距离轨迹及临界带距离轨迹都有明显的起伏。只要设定一定的阈值，即可将语音信号分离出来。笔者也进行了实验，取得了良好的效果。

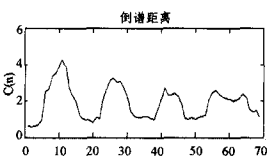


图 3 高信噪比时倒谱距离轨迹

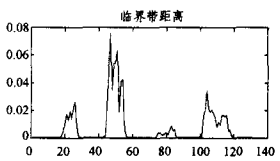


图 4 高信噪比时临界带距离轨迹

但是，在低信噪比情况下，同样以这两种方法进行实验。由图 5 与图 6 可以看出，临界带距离轨迹幅度变化仍然比较明显，可以很准确地切分出语音信号；而倒谱距离轨迹幅度变化差距变小，使端点检测变得困难。此时，通过确定阈值来切分的语音正确率比较低。

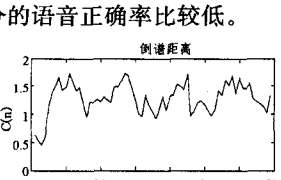


图 5 低信噪比时倒谱距离轨迹

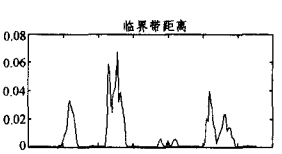


图 6 低信噪比时临界带距离轨迹

(2)同基于谱熵检测方法的实验对比

我们将基于临界带特征矢量的方法与基于谱熵函数的检

(下转第 237 页)

$$m_2(\tau)=\begin{cases}20+\int_0^{\tau}-4\mathrm{d}x & \tau\in[0,5]\\ 0+\int_{5k}^{\tau}-1.8\mathrm{d}x & \tau\in[5k,5k+0.5]\\ 0.9+\int_{5k+0.5}^{\tau}0.2\mathrm{d}x & \tau\in[5k+0.5,5(k+1)]\end{cases}$$

$$m_3(\tau)=\begin{cases}0 & \tau\in[0,5]\\ \int_5^{\tau}0.8\mathrm{d}x & \tau\geqslant 5\end{cases}$$

即在任意时刻 τ 缓冲区 B3 的产品数量$=m_1(\tau)$，缓冲区 B4 的产品数量$=m_3(\tau)$，其中 k 为正整数。

由上述公式可知在任意时刻 τ 均有 $m_1(\tau)+m_2(\tau)=30$，因此缓冲区 B3 的产品数量在任意时刻均小于等于缓冲区的最大容量。

结束语 为了实现区间速率连续 Petri 网的模糊控制，本文建立了区间速率连续 Petri 网的模糊模型，定义了区间速率连续 Petri 网的模糊规则。另外，对区间速率连续 Petri 网的模糊控制进行了讨论，给出了库所标识收敛的定理。ICPNs 中每个变迁的模糊模型都由两条模糊规则构成，由于他们都是线性的，可以适用于控制应用工程。下一步，我们将对区间速率连续 Petri 网的模糊模型在系统控制与优化的应用进行研究。

参 考 文 献

[1] David R, Alla H. Continuous Petri Nets // 8 th European Workshop on Applications and Theory of Petri Nets. Saragosse (E), Juin 1987:275-294

[2] Ball J LE, Alla H, David R. Asymptotic Continuous Petri Nets. J. of Discrete Event Dynamic Systems: Theory and Applications, 1993:235-263

[3] Haounani M, Lefebvre D. Variable Speed Continuous Petri Net. Gringdeld, Switzerland // Proc. of the 17th IASTED Int. Conf. Modeling, Identification And Control. 1998

[4] 叶志宝，赵义军，董焕河. 最大速度变化的连续 Petri 网(VCPN)的动态演变及性质判定. 计算机研究与发展，2002，39(3)：330-334

[5] Gu Tianlong, Dong Rongsheng, Tian Yu-chu. Continuous Petri Nets Augmented with Maximal and Minimal Firing Speeds // IEEE International Conference on Systems, Man and Cybernetics (SMC). 2003

[6] Gu Tianlong, Dong Rongsheng. Novel Continuous Model to Approximate Time Petri Nets: Modeling And Analysis. Journal of Application Mathematic and Computer Science, 2005, 15 (1): 141-150

[7] Liao Weizhi, Gu Tianlong. Optimization And Control Of Production Systems Based On Interval Speed Continuous Petri Nets // IEEE International Conference on Systems, Man and Cybernetics (SMC). Hawaii, USA: 2005: 1212-1217

[8] 廖伟志，古天龙，王汝凉. 区间速率连续 Petri 网模型行为分析研究. 小型微型计算机系统，2006，27(8)：1490-1494

[9] 廖伟志，古天龙. 区间速率连续 Petri 网的有效冲突及其消解. 计算机科学，2006，33(10)：221-224

[10] 廖伟志，文 瑛，王汝凉. 一类区间速率连续 Petri 网的可达稳态分析. 系统仿真学报，2005，17(z1)：44-47

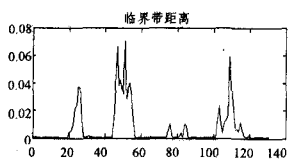
[11] 廖伟志，王汝凉. 区间速率连续 Petri 网可达稳态必要性分析. 计算机工程与应用，2005，41(25)：78-80

[12] 周必水，唐云廷，罗勇. 离散时间 Petri 网的模糊模型. 计算机研究与发展，2003，40(5)：657-660

[13] Hennequin S, Lefebvre D, Elmoudni A. Fuzzy Control of Variable Continuous Petri Nets[C] // Proceeding of the Conference on Decision & Control. Phoenix, Arizona USA: 1999: 1352-1356

(上接第 221 页)

测方法仍然是在低信噪比情况下进行比较。可以说，这两种方法的检测效果不相上下，但是经过前期的算法处理，最后都需要确定一个阈值，以此来确定是语音还是噪声。所以这个阈值的确定很重要，直接影响到切分出来的语音是否完整、是否准确。由图 7 和图 8 可以看出，在语音与噪声的边缘地带，临界带距离轨迹的突变很明显，这使得阈值的确定比较容易，而且切分准确。而谱熵函数由于在噪声部分起伏比较大，所以在语音与噪声的边缘地带的函数突变不是很明显，这就给确定阈值带来了不便。如果确定不准确，极有可能造成漏切或错切。



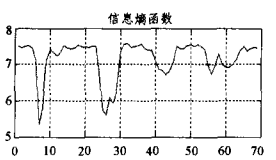


图 7 低信噪比时时临界带距离轨迹　图 8 相同信噪比时谱熵波形

结束语 本文提出了一种新的基于短时谱临界带矢量特征的语音端点检测方法。实验对比表明，该方法在强噪声环境下比基于倒谱距离和基于频谱熵的端点检测方法具有更好的鲁棒性。并且，所需的变换可用高效的 FFT 来完成，计算开销较小，检测精度较高，可广泛用于语音编码与语音识别。

参 考 文 献

[1] LI Ye, WANG Tong, CUI Huijuan, et al. Voice Activity Detection in Non-stationary Noise // IMACS Multi Conference on Computational Engineering in Systems Applications (CESA). Beijing, China, October 2006: 1573-1575

[2] 陈斌，郭大勇，等. 基于 DSP McBSP 的语音实时采集与噪声环境下的端点检测研究[J]. 测控技术，2004(Z1)：212-214

[3] 胡光瑞，韦晓东. 基于倒谱特征的带噪语音端点检测[J]. 电子学报，2001(10)：95-97

[4] Shen J L, Hung J W, Lee L S. Robust Entropy- based End point Detection for Speech Recognition in Noisy Environments [C] // Proceedings of ICSLP- 98. 1998

[5] 陈四根. 基于熵函数的语音端点检测方法[J]. 声学与电子工程，2001：28-30

[6] 赵立. 语音信号处理[M]. 北京：机械工业出版社，2003：42-45