

一种基于谱熵的语音端点检测方法

文/李战明 尚丰

摘要

在低信噪比环境下,为了提高语音端点检测的效果,提出了一种适应于低信噪比环境的语音端点检测方法。基于子带谱熵法,引入正参数对基本的谱熵法进行算法改进,得到改进后的子带谱熵法,通过增加预判环节选择合适的正参数,加大语音信号与噪声信号的区分度,进一步改善在低信噪比环境下算法的效果,得到新的语音端点检测算法。仿真实验表明,新的算法不仅快速高效,具有较强鲁棒性,而且适合在低信噪比环境中较准确的检测出语音端点。

【关键词】语音端点检测 子带谱熵 参数可变 鲁棒性

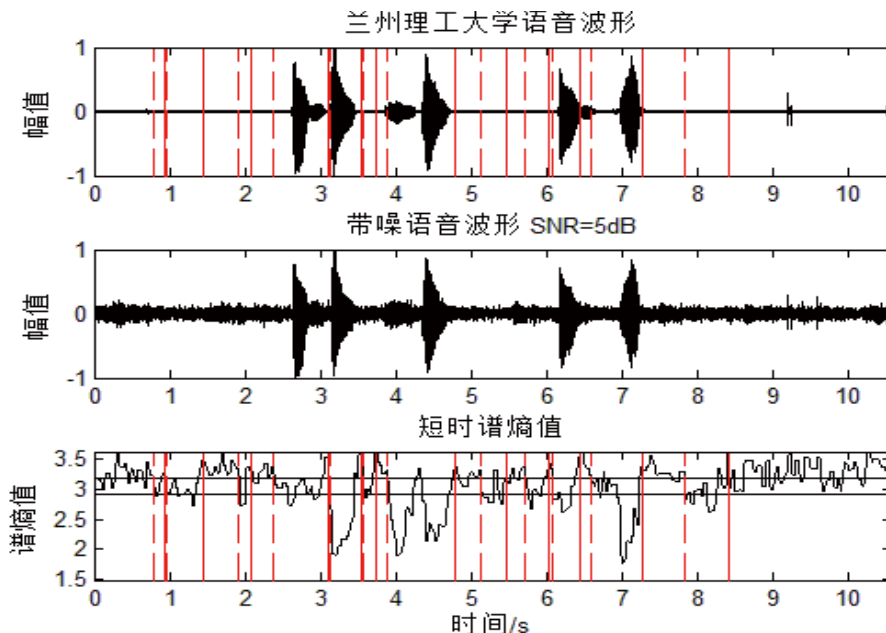


图 1: 未改进谱熵法

1 引言

随着语音识别、语音增强等技术的飞速发展,端点检测越来越重要。端点检测的目的是在复杂环境下识别出语音信号与语音、噪声混合信号,确定语音信号的开始与结束端点,以便进一步的语音信号处理。在助听器、手机等硬件环境下,由于实时性与准确性要求较高,计算量即不能过大以免硬件不能满足要求,又要求语音信号处理快速、准确。准确的端点检测可以大幅度减少计算量,提高系统的语音处理效率。现阶段运用的端点检测算法有短时能量与短时过零率双门限法、自相关函数法、方差法、谱距离法、能零比与能熵比等。在噪声环境下这些方法的效果并不是很理想,语音信号的能量并不大,所以单纯用能量参数并不能很好的将语音信号检测出来。Shen 提出了一种基于信息熵的端点检测方法,信息熵只与语音信号的统计特性有关,与单纯的能量幅值大小无关,所以这种方法能够较好的将语音信号与噪声信号检测出来,并且计算量不大,适用于助听器、手机等硬件环境。但是对于周期性能量较为集中的噪声而言,误判率较高。为了消除每帧信号 FFT 变换后的谱线幅值受噪声影响,把每条谱线的谱熵值改为子带的谱熵,从而改善在周期性噪声环境中端点检测的效果。为提高在低信噪比环境下端点检测的效果,在谱熵计算中引入一个正常量 K 。经过实验证明这种方法确实比一般方法正确率要高,但是随着信噪比不断降低,端点检测误判率又会不断提高,单纯的运用加入正常量的子带谱熵法还是不能适用于噪声较严重的场合。

针对上述情况,本文提出了一种自适应正常量可变的谱熵端点检测改进算法。这种方法基于子带谱熵法,引入一个可变的正数 K ,计算信号谱熵。基于最大离散谱熵定理,通过正数 K 的调节,增大有语音信号帧与无语音信号帧谱熵值的差距,增强语音与噪声的区分度,使算法在低信噪比环境下也能有效的进行。

2 端点检测算法

2.1 谱熵法

熵值来源于热力统计学,是描述紊乱程度的物理量。1948 年香农把熵的概念引入信号处理中,作为随机事件不确定性的度量。一个系统越是有序熵值越低,越是混乱熵值越高。Shen 第一次将谱熵作为端点检测的特征参数。设含噪语音 $x(n)$ 经过加窗、去直流分量、分帧处理后得到第 i 帧信号 $x_i(n)$, FFT 后得到频率 f_k 对应的谱线为 $X_i(k)$,对应的能量谱为 $Y_i(k)=|x_i(k)|^2$,则每个频率分量的归一化谱概率密度是:

$$p_i(k)=\frac{Y_i(k)}{\sum_{m=0}^{N/2}Y_i(l)} \quad (2-1-1)$$

式中, $p_i(k)$ 是第 i 帧频率分量 f_k 对应的概率密度, N 是 FFT 的长度。

短时谱熵定义为:

$$H(i)=-\sum_{k=0}^{N/2}p_i(k)\lg p_i(k) \quad (2-1-2)$$

若熵函数有

$$H(p_1,p_2,\dots,p_q)=H(1/q,1/q,\dots,1/q)=\lg q \quad (2-1-3)$$

即等概率分布时,熵函数取最大值,表明等概率分布时信号不确定性最大,称为最大离散熵原理。

由于语音信号具有共振峰结构,归一化谱概率密度函数分布不均匀,所以语音信号的谱熵值一般都低于噪声的谱熵,在统计学概率上与噪声信号有所区别,与能量大小无关,所以谱熵法在噪声环境中的效果优于其他算法。但是对于周期性噪声,未改进的谱熵法效果一般。如图 1 所示,语音信号是青年男音:“兰州理工大学”,背景噪声为工厂噪声,信噪比为 5DB,误判率较高。

2.2 改进后的子带谱熵法

为了消除每帧信号 FFT 后每条谱线受噪声影响,可把每条谱线的谱熵法改为子带的谱熵。子带谱熵法的思想是把每一帧信号分为若干个子带,然后求每条子带的谱熵,这样就消除了单独谱线受噪声的干扰。如前面所述语音信号 $X_i(n)$, FFT 后得到:

$$X_i(k)=\sum_{l=0}^{N-1}X_i(l)e^{-j2\pi kln/N} \quad (2-2-1)$$

谱概率密度函数、谱熵分别为式 2-1-1 与 2-1-2。假设每个子带由 m 条谱线组成,共有 N_b 个子带,则第 i 帧中的第 m 个子带能量为:

$$E_b(m,i)=\sum_{k=(m-1)*n}^{(m*n)-1}Y_i(k) \quad 1 \leq m \leq N_b \quad (2-2-2)$$

相应的, 子带能量概率 $p_b(m,i)$ 和子带谱熵 $H_b(i)$ 分别为:

$$p_b(m,i) = \frac{E_b(m,i)}{\sum_{k=0}^{N_b} E_b(k,i)} \quad 1 \leq m \leq N_b \quad (2-2-3)$$

$$H_b(i) = -\sum_{m=0}^{N_b} p_b(m,i) \lg p_b(m,i) \quad (2-2-4)$$

引入子带谱熵法后可以有效的解决单独谱线受噪声影响的问题, 考虑到语音信号频率在 250~3500HZ 频带内, 则剔除有效频带之外的谱线。根据语音特性可知, 由于共振峰结构某个频带能量有可能比较集中, 但是不会过于集中, 能量过于集中的往往是含有爆破音的噪声, 也将之剔除。即若 $p_i(k) > 0.9$, 则 $p_i(k) = 0$ 。Chuan Jia, Bo Xu 等人提出了在谱上的计算中引入一个正常量 K 在式 2-2-3 中, 得到:

$$p_b'(m,i) = \frac{E_b(m,i) + K}{\sum_{k=1}^{N_b} (E_b(k,i) + K)} \quad (2-2-5)$$

$$\text{与 } H_b'(i) = -\sum_{m=0}^{N_b} p_b'(m,i) \lg p_b'(m,i) \quad (2-2-6)$$

多次仿真实验发现算法效果优于谱熵法, 如图 2 所示, 由于子带谱熵的引入, 由于谱线引入的误判大幅度降低。继续降低信噪比, 改进后的子带谱熵法误判率也会随之升高, 如图 3 所示, 有误判的出现。为使高噪声环境下端点检测算法可以有好的效果, 现在提出一种参数可变的子带谱熵法。

2.3 参数可变的子带谱熵算法

在谱熵的计算中引入一个新的正参数 K , 得到式 2-2-5 与式 2-2-6, 则:

$$\begin{aligned} \Delta p_b(m,i) &= p_b'(m,i) - p_b(m,i) \\ &= \frac{[E_b(m,i) + K] \cdot \sum_{k=1}^{N_b} E_b(k,i) - E_b(m,i) \cdot [\sum_{k=1}^{N_b} (E_b(k,i) + K)]}{[\sum_{k=1}^{N_b} (E_b(k,i) + K)] \cdot \sum_{k=1}^{N_b} E_b(k,i)} \\ &= \frac{\sum_{k=1}^{N_b} E_b(k,i) \cdot E_b(m,i) + K \sum_{k=1}^{N_b} E_b(k,i) - E_b(m,i) \cdot [\sum_{k=1}^{N_b} E_b(k,i) + N_b \cdot K]}{(\sum_{k=1}^{N_b} E_b(k,i) + N_b \cdot K) \cdot \sum_{k=1}^{N_b} E_b(k,i)} \\ &= \frac{K[1 - p_b(m,i)N_b]}{\sum_{k=1}^{N_b} E_b(k,i) + N_b \cdot K} = \frac{1 - p_b(m,i)N_b}{\frac{\sum_{k=1}^{N_b} E_b(k,i)}{K} + N_b} \end{aligned}$$

式中, N_b 是子带数为常量, $\sum_{k=1}^{N_b} E_b(k,i)$ 是一帧中的总能量也为常数。 $p_b(m,i)$ 是第 m 个子带能量的概率。引入一个正常量 K 后, 当 $1 - p_b(m,i)N_b < 0$ 时, 即 $p_b(m,i) > 1/N_b$ 时, 子带能量较大, 新子带能量概率小于原子带能量概率; 当 $1 - p_b(m,i)N_b > 0$ 时, 即 $p_b(m,i) < 1/N_b$ 时, 子带能量较小, 新子带能量概率大于原子带能量概率; 当 $1 - p_b(m,i)N_b$ 趋近于 0 时, 新子带能量概率基本不变, 这就说明引入正常量 K 后会将子带能量平均化, 从而使谱熵值

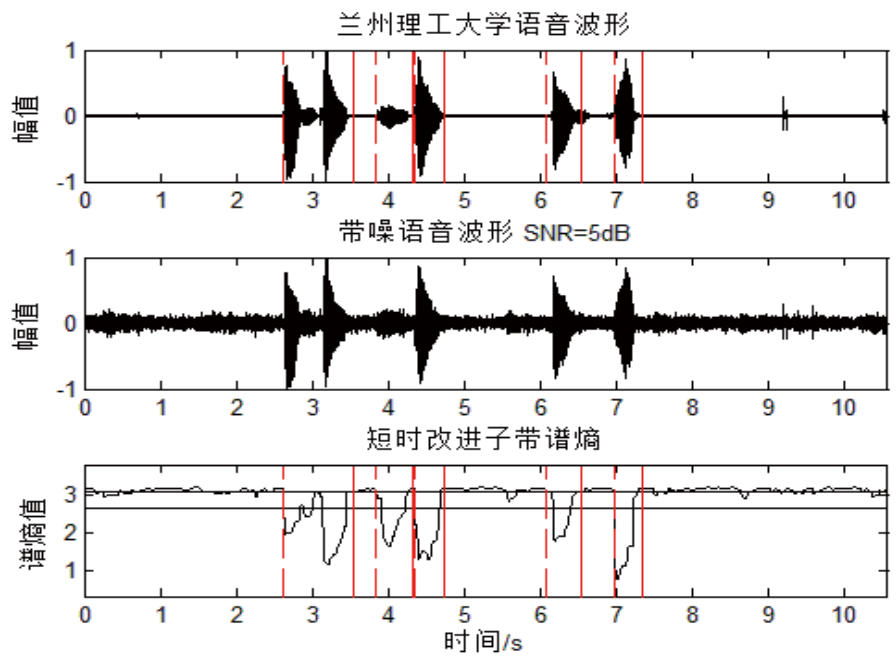


图 2: 改进后子带谱熵法

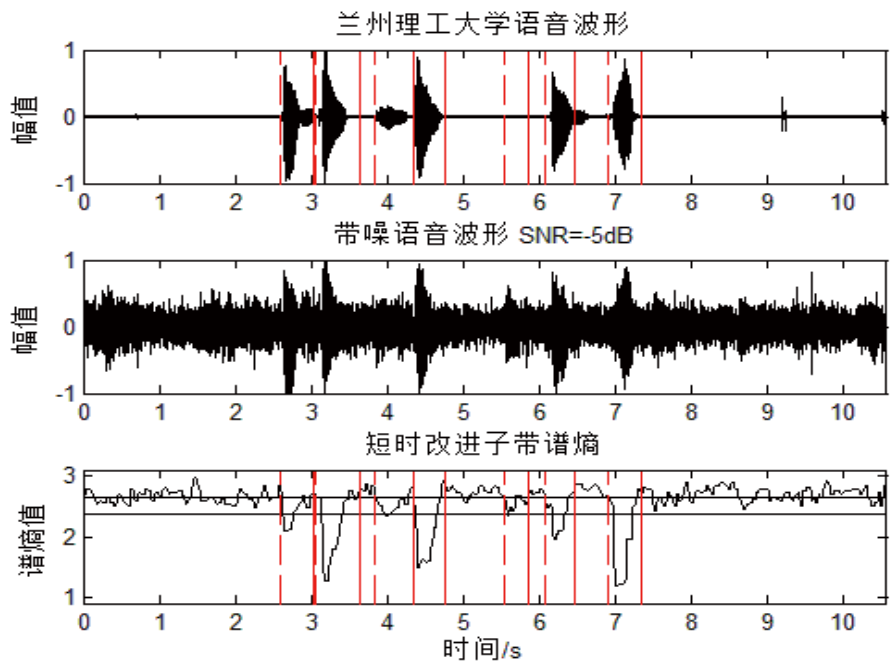


图 3: 低信噪比时改进后的子带谱熵法

增大。而语音信号的总能量要大于噪声信号的总能量, 所以语音信号中 $\Delta p_b(m,i)$ 的变化小于噪声信号中 $\Delta p_b(m,i)$ 的变化, 故加入正常量 K 后谱熵值增大, 其中语音帧谱熵值增量比噪声谱熵值增量要小。单独取出 $\Delta p_b(m,i) = \frac{1 - p_b(m,i)N_b}{\sum_{k=1}^{N_b} E_b(k,i) + N_b \cdot K}$, 对 K 进行求导, 得:

$$\frac{d[\Delta p_b(m,i)]}{dK} = (1 - pN) \cdot \left(\sum_{k=1}^{N_b} E_b(k,i) + KN_b \right)^{-2}$$

$$= \frac{1 - p_b(m,i)N_b}{\left(\sum_{k=1}^{N_b} E_b(k,i) + KN_b \right)^2} \quad (2-3-2)$$

当 $p_b(m,i) < 1/N_b$ 时, 即子带能量较小时, $\frac{d[\Delta p_b(m,i)]}{dK} > 0$, 又 $\Delta p_b(m,i) > 0$, 故 $|\Delta p_b(m,i)|$ 随着 K 的增大而增大; 当 $p_b(m,i) > 1/N_b$ 时, $\frac{d[\Delta p_b(m,i)]}{dK} < 0$, 又 $\Delta p_b(m,i) < 0$, 则 $|\Delta p_b(m,i)|$ 也随着 K 的增大而增大; 这就说明随着 K 的

增大, 谱熵值会随着增大。每段信号开始都有一段无语音段 $X_N(n)$, 可以求出无语音段谱熵值得平均值以确定一个阈值。在之后的端点检测中, 如果某帧谱熵值较大, 则增大参数 K 的值; 如果某帧谱熵值较小, 则减小参数 K 的值。这样可以使语音段与噪声段谱熵值差距更大, 从而提高端点检测的准确性。

2.4 算法流程

设前导无话段帧长为 NIS , 先求出这段信号中的谱熵 $H_N(i)$, 由于噪声是无序、杂乱的, 所以 $H_N(i)$ 的谱熵比较大。将 $H_N(i) \in [1, NIS]$ 求出, 并进行平滑处理, 再求出均值 $\overline{H_N}$ 。用 $\overline{H_N}$ 找出一个阈值 V_b 。在之后的端点检测中先进行比较, 若 $H_b(i-1) \geq V_b$, 则选取较大的 K 值计算 $H_b(i)$; 若 $H_b(i-1) < V_b$, 则选取较小的 K 值计算 $H_b(i)$, 使语音段与噪声段谱熵值差距加大, 改善在低信噪比时端点检测的效果。算法具体步骤如下:

- Step1: 计算前导无话段信号的谱熵 $H_N(i)$, 并依上所述计算阈值 V_b ;
- Step2: 通过计算前导无话段的平均谱熵值 $\overline{H_N}$, 求出判定阈值 S_1 与 S_2 ;
- Step3: 计算 $H_b(i-1)$, 若 $H_b(i-1) \geq V_b$, 则选取较大的 K 值, 若 $H_b(i-1) < V_b$, 则选取较小的 K 值;
- Step4: 计算 $H_b(i)$, 当 $H_b(i)$ 小于 S_2 时肯定是语音, 若不小于则继续 step2 直到检测到语音开始点;
- Step5: 检测到语音开始点后比较 $H_b(i)$ 与 S_1 , 若大于 S_1 则为结束点, 若不大于则继续 step5
- Step6: 重复 step4、step5 直到语音段结束。

2.5 实验仿真结果

改进算法后在 MATLAB 中进行仿真, 新算法的效果经实验验证要好于改进前的算法, 如图 4 所示, 在低信噪比为 -5DB 时, 算法也有比较好的效果。为了说明一般性, 录制男女各 10 人, 样率为 8000HZ、16bit 量化的纯净语音信号, 分别叠加白噪声、工厂噪声对比改进前与改进后算法的效果见表 1 与表 2, 实验证明新算法的误判率较低, 端点检测效果较好。

3 小结

本文提出了一种可以根据环境噪声选择参数的子带谱熵端点检测算法, 即改进了可引入常数的谱熵法, 得到一种抗噪能力较强的子带谱熵法, 并通过仿真实验证明与之前的算法相比在低信噪比环境中检测效果更好。

参考文献

[1] 杨行峻, 迟惠生. 语音信号数字处理 [M]. 北京: 电子工业出版社, 1995.
[2] 卢艳玲. 一种基于多特征的带噪语音信号端点检测与音节分割算法 [J]. 电声技术, 2005 (07): 60-62.

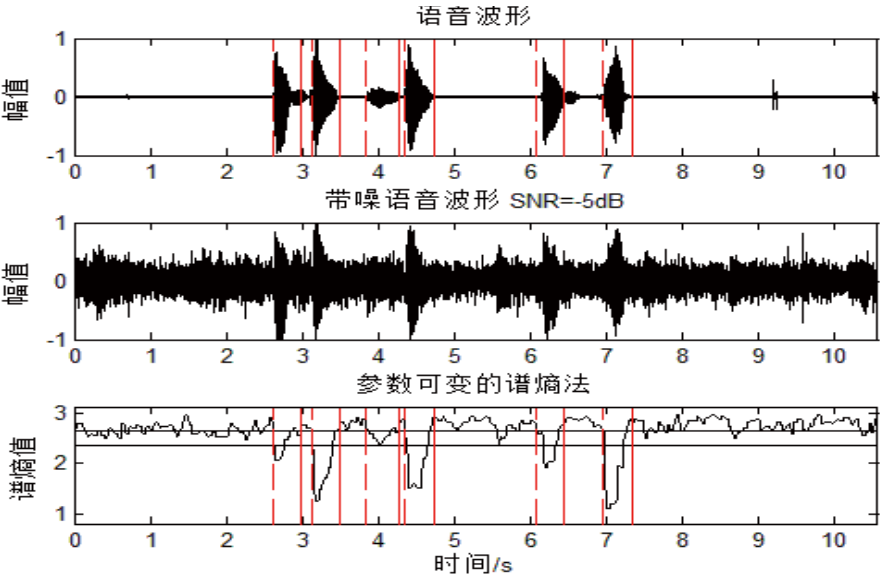


图 4: 参数可变的谱熵法

表 1: 工厂噪声环境下语音端点检测结果 (误检率 / %)

算法	5DB	0DB	-5DB
未改进党的子带谱熵法	21.4	30.3	41.5
引入参数的子带谱熵法	18.2	21.7	38.6
参数可变的子带谱熵法	13.3	19.6	27.5

表 2: 白噪声环境下语音端点检测结果 (误检率 / %)

算法	5DB	0DB	-5DB
未改进党的子带谱熵法	14.3	17.6	25.8
引入参数的子带谱熵法	8.8	10.9	15.7
参数可变的子带谱熵法	9.3	11.1	13.4

[3] 孙炯宁, 傅德胜, 徐永华. 基于熵和能量的语音端点检测算法 [J]. 计算机工程与设计, 2005, 26 (12): 3429-3431.
[4] 张徽强. 带噪语音信号的端点检测和声韵分离 [D]. 国防科学技术大学, 2005.
[5] Shen J, Hung J, Lee L. Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments [C] // Proceeding of International Conference on Spoken Language Processing. Sydney: [s, n], 1998: 232-238.
[6] 王琳, 李成荣. 一种基于自适应谱熵的端点检测改进方法 [J]. 计算机仿真, 2010, 27 (12): 373-375.
[7] Chuan Jia, Bo Xu. An Improved Entropy-Based Endpoint Detection Algorithm [C]. ISCSLP, 2002, 285-288.

作者简介
李战明 (1962-), 教授、博导, 兰州理工大学电气工程与信息工程学院院长。主要研究方向为复杂系统的建模与控制, 神经模糊系统与软计算, 智能信息处理与模式识别, 计算机控制系统的理论与工程等。
尚丰 (1990-), 男, 甘肃甘南州合作市人。硕士学位。研究方向为信号与信息处理。

作者单位
兰州理工大学 甘肃省兰州市 730050