

# 一种基于分带谱熵的语音激活检测算法

张敏

(湖南邮政公司, 湖南 长沙 410012)

**摘要:** 为了提高语音激活检测在汽车内部噪声环境下的检测性能, 提出了一种基于分带谱熵的语音激活检测算法。将实验仿真结果与 ITU 标准 G.729B 中的检测性能进行了分析比较, 结果表明, 该算法在汽车内部噪声环境下具有较高的准确率和稳定性, 且算法的复杂度较低, 具有一定的实用价值。

**关键词:** 语音激活检测; 分带谱熵; 汽车内部噪声

中图分类号: TN912.3

文献标识码: A

文章编号: 1674-7720(2010)20-0043-03

## A voice activity detection algorithm based on the band-partitioning spectral entropy

ZHANG Min

(Hunan Post Company, Changsha 410012, China)

**Abstract:** In order to improve the performance of voice activity detection under car interior noise environments, a new voice activity detection algorithm based on the band-partitioning spectral entropy was presented. The performance of the algorithm was compared with the voice activity detection algorithm in standard G.729 annex B of ITU in experimental simulation. Experimental results show that the algorithm has high classification accuracy and stability; moreover, this algorithm has low complexity and applied value.

**Key words:** voice activity detection; band-partitioning spectral entropy; car interior noise

语音激活检测 VAD (Voice Activity Detection) 指采用一定的信号处理技术来检测信号中是否包含语音信号, 并且从背景噪声中找出每一段语音的开始和终止, 以明确找出语音信号存在的范围。目前该技术被广泛应用于语音编码、语音增强、语音合成、回声抵消以及语音通信等领域。

语音激活检测的研究历史很长, 从最早的基于短时能量、过零率的判断, 到基于语音模型和统计知识的各种复杂算法, 语音激活检测算法的原理和实现方法都在不断地更新。近年来, 一些学者提出了基于支持向量机<sup>[1]</sup>、小波理论<sup>[2]</sup>、神经网络<sup>[3]</sup>和循环累积量<sup>[4]</sup>等理论的检测方法, 对语音检测方法的研究给出了新的思路。如何在保证算法鲁棒性的同时降低算法的复杂度是一个极具挑战性的问题, 也是目前语音激活检测中的热点问题。本文设计了一种基于分带谱熵的语音激活检测算法, 该方法采用分带谱熵作为判决参数, 达到了较高的判别准确率及较低的算法复杂度。

### 1 特征提取

特征提取的框图如图 1 所示。首先对加噪语音进行分帧和加窗处理, 然后进行时频变换, 由参考文献[5]知, 与常用的 FFT 相比, 用离散余弦变换谱熵来区分噪声与语音具有更大的隔离度, 因此本文选用离散余弦变换(DCT)。离散余弦变换的定义式为:

$$C_x(0) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \quad (1)$$

$$C_x(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k\pi}{2N} \quad (2)$$

其中,  $k=1, 2, \dots, N-1$ 。

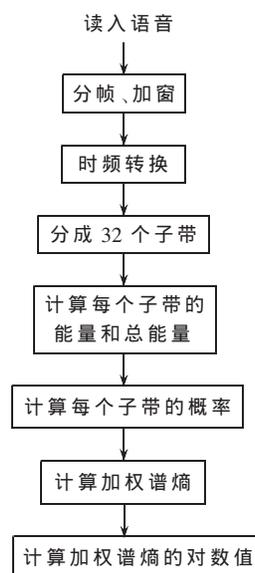


图 1 特征提取框图

## 图形、图像与多媒体

Image Processing and Multimedia Technology

在 DCT 变换之后,对 DCT 系数进行分带,分成 32 个子带,然后分别计算它的能量,其计算公式为:

$$E_l = \sum_{i=8 \times l+0}^{8 \times l+7} C_x^2(i) \quad (3)$$

式中,  $l=0,1,\dots,31$ 。

总的能量  $E_{\text{total}} = \sum_{l=0}^{31} E_l$ , 因此每个子带的概率为  $p_l =$

$$\frac{E_l}{E_{\text{total}}}$$

为了提高概率密度函数分辨语音信号和非语音信号的能力,提出了一些经验性的约束。首先,由于大部分语音信号都在 250 Hz~6 000 Hz 频带内,因此有:

$$E_l=0, l=0 \text{ 或 } l \geq 24 \quad (4)$$

其中,  $E_l$  为第  $l+1$  个子带的能量。

其次,为了消除某些能量集中在特定频率的噪声对方法的影响,需设定概率的上限,即:

$$p_l=0, \text{ if } p_l > 0.9 \quad (5)$$

因此,相应的每一帧语音谱熵的定义为:

$$H_n = - \sum_{l=0}^{31} p_l \log_{10} p_l \quad (6)$$

式中,  $H_n$  是第  $n$  帧语音信号的谱熵。

为进一步提高语音激活检测的准确率,本文采用短时平均能量进行加权的方法,即:

$$H_n' = -E_n \sum_{l=0}^{31} p_l \log_{10} p_l \quad (7)$$

其中,  $E_n = \sum_{k=0}^{N-1} |C_x^2(k)|$  表示第  $n$  帧语音信号的短时平均能量,  $H_n'$  是第  $n$  帧语音信号的加权谱熵。同时,为了减少能量的影响,对加权谱熵取对数,图 2 所示为语音和其加权谱熵的图。

从图 2(c)和图 2(d)可以看出,加权谱熵对帧能量更加敏感,而通过对其取对数,在一定程度上降低了能量的影响。正因为如此,本文采用加权谱熵的对数值作为特征量,提高了算法的鲁棒性。

## 2 语音激活检测算法

(1) 对输入的语音信号  $s(n)$  进行分帧处理,帧长 32 ms (256 个样点),得到每帧信号  $s_i(n)$ ,  $i$  表示第  $i$  帧,并对  $s_i(n)$  进行加窗处理,窗函数采用汉明窗。

(2) 自适应阈值的确定。可合理假设待检测语音的前几帧是非语音信号,本文中假设前 10 帧为非语音信号,计算其加权谱熵的对数值,然后求它们的最大值,门限值就是前 10 帧信号的加权谱熵的对数值中的最大值。

(3) 计算每一帧信号的加权谱熵的对数值,然后把它与门限值进行比较。如果大于门限值,则判为语音帧,否则判为噪声帧或静音帧。

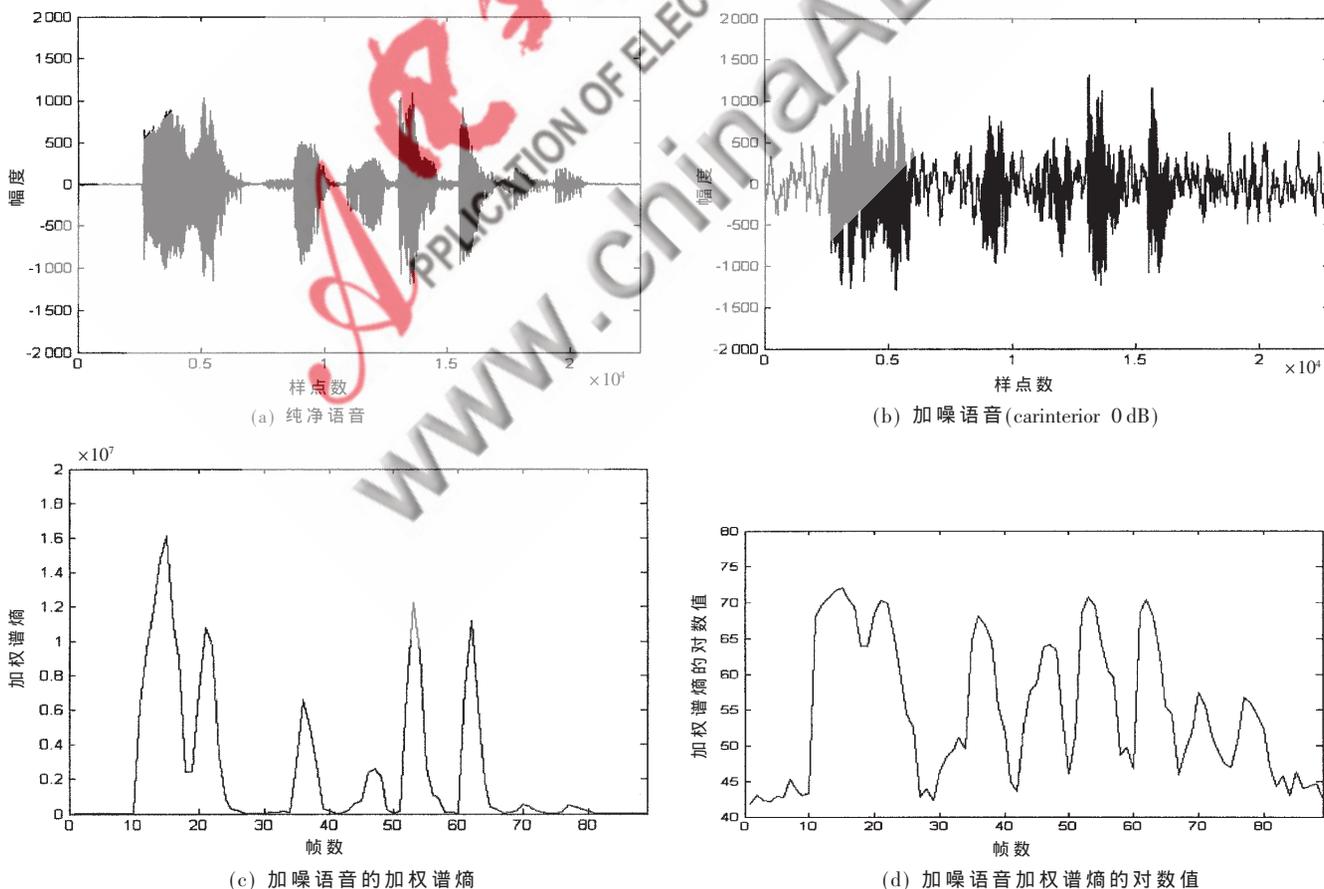


图 2 语音与其加权谱熵

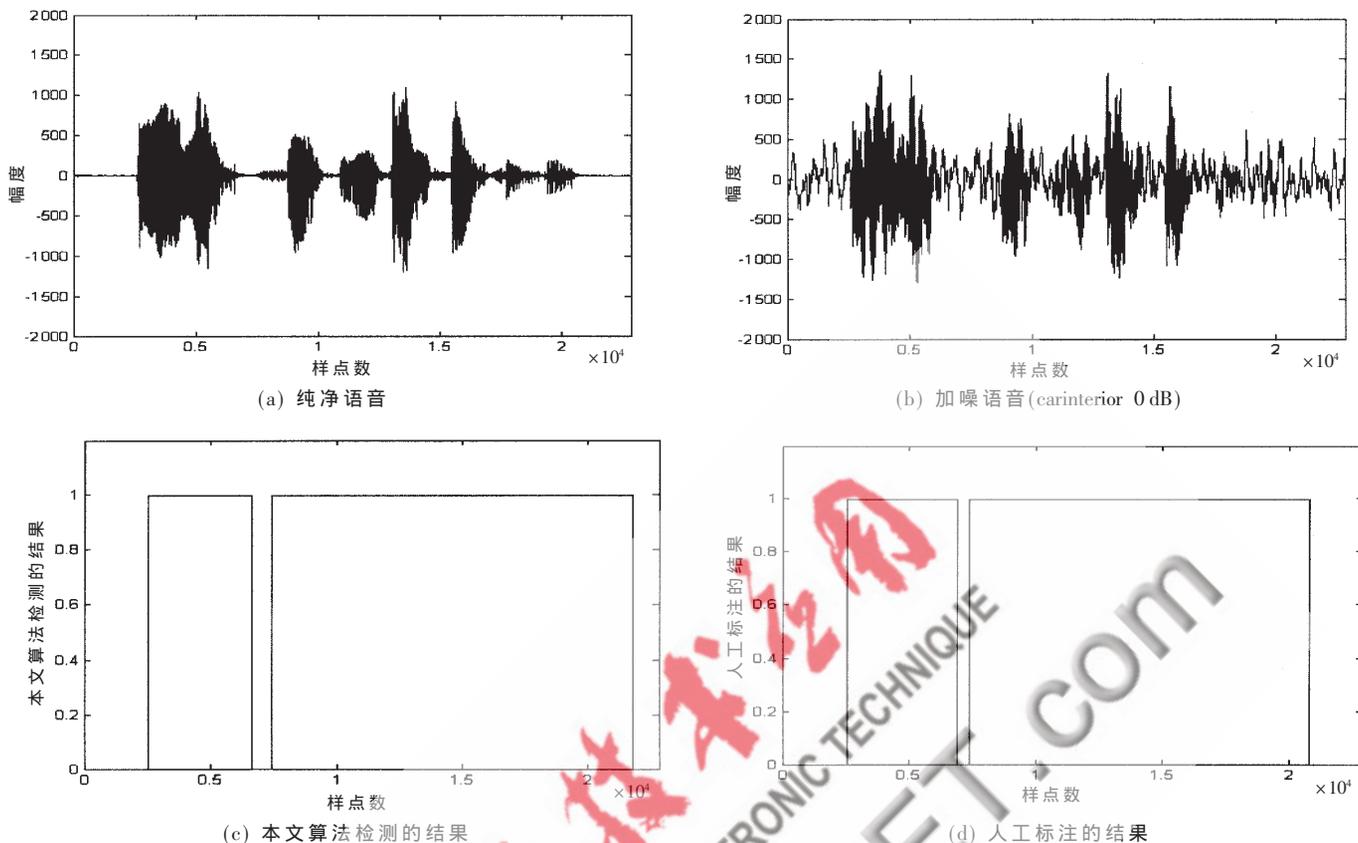


图3 语音波形和判决结果

(4)对判决结果进行平滑处理。为了有效地避免激活状态的频繁切换,本文对判决结果进行平滑后处理。由于无论是非语音信号还是语音信号,一般都会持续一段时间,因此为了剔除判决时的少数坏点,在判决结束后加入了平滑后处理。即:

```

if vadi-3=vadi
    vadi-3=vadi-2=vadi-1=vadi
else
    { vadi-3=vadi-3
      vadi-2=vadi-2
      vadi-1=vadi-1
      vadi=vadi }
    
```

其中,vad<sub>i</sub>表示第*i*帧的检测结果,且当*i*≤10时,vad<sub>i</sub>=0。

### 3 仿真与实验结果

本文的算法用 VC、Matlab 进行仿真。图 3 所示为原始语音、加噪语音(噪声为汽车内部噪声,信噪比为 0dB)、人工标注的结果和使用本文算法检测的结果。图 3(a)是一段纯净语音,其采样率为 8 000 Hz。图 3(b)是加噪语音,所加噪声为汽车内部噪声,信噪比为 0 dB。图 3(c)为本文算法检测的结果。图 3(d)为人工标注的结果。由图 3(c)可以看出,本文的方法对低信噪比条件下的语音依然具有较好的分类效果。

为进一步说明本文算法对加噪语音进行检测的鲁

棒性,对本文算法和 G.729B 中的 VAD 算法进行对比。对 1 000 帧中文信号进行实验,测试环境分别选取了-5 dB、0 dB、5 dB、10 dB、20 dB、30 dB 等 6 个不同的信噪比,充分体现了算法在不同信噪比下的性能。图 4 给出了不同算法在相同背景噪声类型(都为汽车内部噪声)、不同信噪比情况下的检测准确率。

从图 4 可以看出,本文算法的性能基本上不随信噪比的变化而变化。同时可以看出当信噪比逐渐降低时,

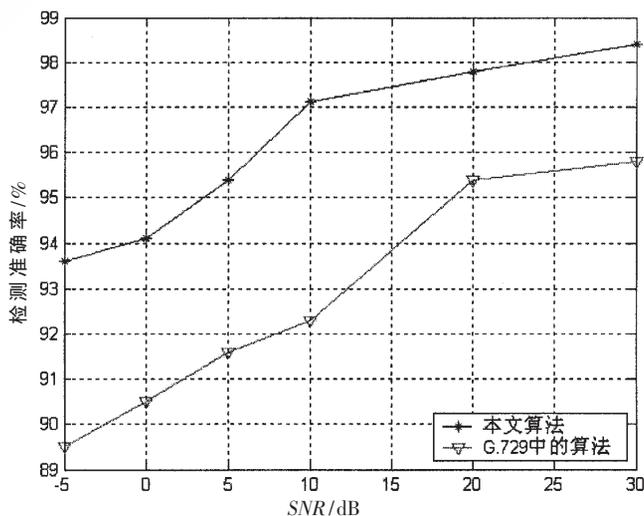


图4 两种算法在同噪声背景不同信噪比环境下的比较

G.729B 算法性能有了明显的下降。

本文中提出了一种能够准确进行 VAD 判决的算法。该算法提取了加权谱熵的对数值作为特征,然后通过自适应阈值,实现语音和静音(或噪声)的准确检测。实验结果表明,该语音激活检测算法在汽车内部噪声条件具有较高的准确率和稳定性,是一种有效、简单的语音激活检测算法。

参考文献

[1] 齐峰岩,鲍长春.一种基于支持向量机的含噪语音的清/浊/静音分类的新方法[J].电子学报,2006,34(4):605-611.

[2] AGHAJANI K H, MANZURI M T, KARAMI M, et al. A robust voice activity detection based on wavelet transform. 2008 Second International Conference on Electrical Engin-

earing (ICEE). Lahore, Pakistan,2008.

[3] 柳燕,鲍长春.基于竞争网络的语音激活算法研究[J].信号处理,2006,22(1):57-60.

[4] 窦慧晶,李如玮,鲍长春.一种基于循环累积量的语音激活检测算法[C].黄山:第九届全国人机语音通讯学术会议(NCMMSC2007),2007.

[5] 汪涛,胡剑凌.基于低信噪比条件下的 VAD 算法研究[J].电子工程师,2005,31(2):42-45.

(收稿日期:2010-07-12)

作者简介:

张敏,女,1979年生,助理工程师,主要研究方向:语音通信。

