第35卷 第18期 计 算 机 工 程 2009年9月
Vol.35 No.18 Computer Engineering September 2009
·人工智能及识别技术· 文章编号：1000—3428(2009)18—0164—03 文献标识码：A 中图分类号：TP391

基于谱熵噪声估计的改进减谱法

李振静[1]，王国胤[1,2]，杨 勇[1,2]，罗 飞[1]
(1. 重庆邮电大学计算机科学与技术研究所，重庆 400065；2. 西南交通大学信息科学与技术学院，成都 610031)

摘 要：为优化基本减谱法在低信噪比下的性能，必须对噪声进行更为准确的估计。提出一种基于谱熵噪声估计的改进减谱法，通过谱熵值估计每一帧语音的噪声，利用减谱法减去噪声以达到降噪的目的。实验结果表明，该方法与传统消噪方法相比，在不同噪声环境下均能取得较好的去噪效果。

关键词：谱熵；减谱法；语音识别

Improved Spectral Subtraction Method Based on Spectral Entropy Noise Estimation

LI Zhen-jing[1], WANG Guo-yin[1,2], YANG Yong[1,2], LUO Fei[1]
(1. Institute of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065;
2. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031)

【Abstract】 In order to improve the performance of traditional spectral subtraction in low SNR environment, more accurate noise estimation is needed. This paper proposes a novel spectral subtraction based on noise estimation of spectral entropy. It estimates the noise of each frame according to its spectral entropy, and subtractes it to enhance the frame. Experimental results show that in comparison with other classical methods, the proposed method can obtain better performance in different noisy environment.

【Key words】 spectral entropy; spectral subtraction method; speech recognition

1 概述

目前语音识别在安静环境下已能达到较高的识别率，但在噪声环境下识别效果仍不理想。典型的语音消噪方法有减谱法、维纳滤波法(Weiner)、基于最小均方误差法(MMSE)等。减谱法由于其计算复杂度低、实时性强、易于实现，因此得到广泛运用。

减谱法中的主要问题是如何准确地估计噪声，现有方法多是选取语音段的寂静帧作为噪声帧并直接减去。这首先就要求噪声帧的选取要准确，这在低信噪比环境下十分困难，还要假定语音中确实有足够的寂静帧的存在。再者，由于噪声是随机变化的，在全部帧都减去相同的噪声值也不准确。

谱熵被用来描述语音信号中频谱分布的平均不确定性。文献[1]将谱熵的概念运用到语音的端点检测中，通过实验发现，谱熵不仅可以区分语音帧和噪声帧，而且在不同类型的噪声环境下也有效，最后得到了较好的检测结果。目前谱熵大多运用于语音端点检测领域以及特征提取领域。

本文提出一种基于谱熵噪声估计的改进减谱法，该方法可以更为准确地估计噪声帧。

2 基本减谱法

设

$$y(n)=s(n)+d(n) \tag{1}$$

其中，$y(n)$是带噪语音；$s(n)$是纯净语音；$d(n)$是噪声。

对式(1)的两边进行傅里叶变换(FFT)，得到频谱：

$$Y(k)=S(k)+D(k) \tag{2}$$

忽略噪声和纯净语音之间的相位差别，可以得到：

$$|Y(k)|=|S(k)|+|D(k)| \tag{3}$$

利用人耳对语音相位的不敏感性，通过式(3)直接从带噪语音谱中减去噪声谱，用带噪语音的相位来近似语音的相位，就得到了消噪后的语音：

$$\hat{S}(k)=[|Y(k)|-|D(k)|]\mathrm{e}^{\mathrm{j}\theta y(k)} \tag{4}$$

基本减谱法是将噪声的统计均值估计为当前噪声帧，但由于噪声是随机变化的，因此在减谱时容易产生残留噪声，这就是音乐噪声。为抑制基本减谱法中音乐噪声问题，出现了许多改进的减谱法。有学者提出在减谱前进行语音活性检测[2]，但在低信噪比环境下表现不理想；也有学者将听觉掩蔽模型应用于减谱法[3]，但模型中参数须人为确定，而且系统复杂度增加，会引入新的频谱失真。

3 谱熵噪声估计

3.1 谱熵的定义及性质

一般来说，语音信号是短时(0～32 ms)平稳的。将一段语音分成若干个短时帧，然后对其进行快速傅里叶变换即得到短时频谱，然后对所有频率分量的频谱进行归一化，即可得到其概率密度函数(PDF)：

基金项目：重庆市杰出青年科学基金资助项目(2008BA2041)；重庆市自然科学基金资助项目(2007BB2445)；重庆市计算机网络与通信重点实验室开放基金资助项目

作者简介：李振静(1979－)，男，硕士研究生，主研方向：语音识别，情感计算；王国胤，教授、博士生导师；杨 勇，副教授、博士研究生；罗 飞，硕士研究生

收稿日期：2009-01-04 **E-mail：**lzj_cqupt@sina.com

$$p_i = s(f_i)/\sum_{k=1}^{N} s(f_k),\ i = 1,\ 2,\ \cdots,\ N \tag{5}$$

其中，$s(f_i)$ 就是每一个频率分量 f_i 的频谱值；p_i 是对应的概率密度值；N 是 FFT 的频率点数。

则每一帧的谱熵可定义如下：

$$H = -\sum_{k=1}^{N} p_k \mathrm{lb}(p_k) \tag{6}$$

谱熵是熵的一种形式，熵的一些有关性质如下：

(1)对称性

在 $p_k = (p_1, p_2, \cdots, p_N)$ 中，各分量的次序任意改变，熵值不变。

(2)极值性

对于离散随机变量集合 p_k，当集合中的事件等概率发生时，熵达到最大值。例如在式(6)中，$H \le \mathrm{lb}(N)$。

(3)渐化性

$$H(p_1, p_2, \cdots, p_k) = H(p_1 + p_2, p_3, \cdots, p_k) + (p_1 + p_2) H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$$

其中，$0 \le p_k \le 1$，$k = 1,\ 2, \cdots,\ N$，$\sum_{k=1}^{N} p_k = 1$，$p_1 + p_2 > 0$。

3.2 谱熵噪声估计

3.2.1 谱熵与频谱的关系及变化规律

根据熵的性质和谱熵的定义，推出谱熵与频谱的关系及变化规律：由于纯净语音是短时平稳的(0～32ms)，频率范围较小，即频谱 $s(f_i)$ 集中在若干频率点上，则其概率分布 p_i 值也相对集中在若干频率点上，在此可以表示为 $p_{i1} = (p_1, p_2, \cdots, p_s, 0, 0, \cdots, 0)$ $i = 1 \cdots N, s << N$；而噪声的频谱信息比较丰富，频谱 $s(f_i)$ 分散在大多频率点上，那么其概率分布 p_i 值也相较分散，在此可以表示为 $p_{i2} = (p_1, p_2, \cdots, p_n, 0, 0, \cdots, 0)$ $i = 1,\ 2, \cdots,\ N,\ n \approx N$；对于 $H(p_{i1})$，$H(p_{i2})$，由于 $s << n$，因根据谱熵的渐化性(性质 3)以及幅度无关性可以得到 $H(p_{i1}) < H(p_{i2})$，即噪声的谱熵值总大于纯净语音的谱熵值。

不失一般性，由熵的对称性可知，谱熵与频谱分量的次序无关，又由于纯净语音频率范围较小，因此纯净语音的谱熵较为相似；从谱熵的定义可以看出，谱熵值的大小仅与语音信号中频谱的分布情况有关，与频谱值的大小没有直接的关系。例如：若 $p_k = (p_1,\ p_2,\ p_3,\ p_4)$，则对谱熵值 H 有 $H(1,\ 1,\ 1,\ 1) = H(10,\ 10,\ 10,\ 10) > H(100,\ 20,\ 20,\ 20)$。

综上所述，谱熵与信号频谱分布情况密切相关，可以用于区分噪声帧和语音帧；谱熵与频谱的幅度不直接相关，在区分噪声帧和语音帧时有一定的鲁棒性。

3.2.2 利用谱熵进行噪声帧的选择

由上述分析可知，谱熵能区分噪声帧和语音帧。本文首先选取谱熵值最大的一帧作为原始噪声帧。图 1～图 5 为一些语音信号谱熵值的对比，本文 1 帧取 512 个样点，由谱熵的极值性可知，其谱熵值总小于等于 $\mathrm{lb}512 = 9$。

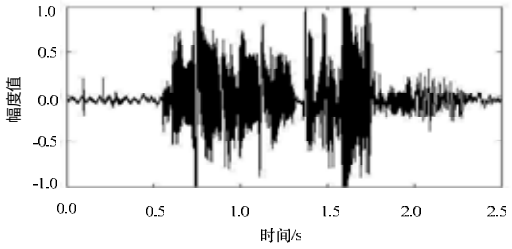


图 1　纯净语音信号 S_0

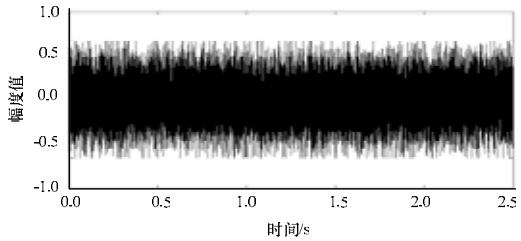


图 2　纯高斯白噪声 N_0

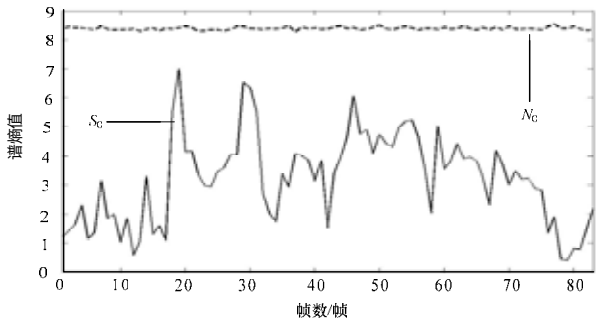


图 3　S_0 及 N_0 的谱熵值

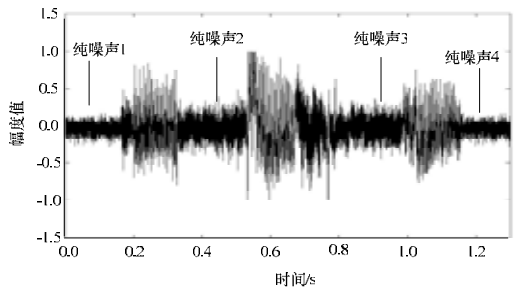


图 4　含噪语音 S_1

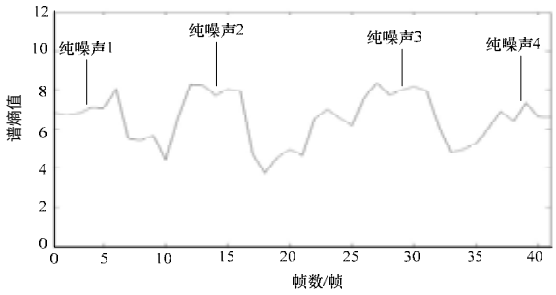


图 5　S_1 的谱熵值

从图 1 和图 3 可以看出，纯净语音信号的谱熵都较小，主要在 4.0 上下波动；而噪声的谱熵都较大，在图 3 中维持在 8.0 以上；同样从图 4 和图 5 可以看出，在含噪语音中，语音帧的谱熵值也是一般小于噪声帧的谱熵值。因此，可将一段语音中谱熵最大的一帧作为原始噪声帧。

3.2.3 利用谱熵进行噪声帧的更新

利用减谱法进行去噪时，若噪声估计值不准确，减谱后会产生音乐噪声。为了减少音乐噪声，本文根据各帧谱熵的变化来动态的更新噪声值。

从图 4、图 5 中可以看出，噪声幅度在各个语音段的值会是随机变化的(如图 4 中的纯噪声点 1～噪声点 5)，但其谱熵值变化不大(如图 5 所示)；在所有语音段都减去相同的噪声显然是不准确的，如在图 4 中纯噪声点 2 及其附近各帧处

仍减去纯噪声点 1 时的噪声就会造成残余噪声较大，因此，在纯噪声 2 处及后续帧应该将噪声值更新为新的噪声值。

通过及时跟踪并更新噪声值，可使每一帧都减去与前面最邻近的噪声值，由于语音信号是短时平稳的，因此这样进行减谱较为准确。

4 基于谱熵噪声估计的减谱法

本文方法分为 3 个部分：(1)原始噪声估计，选取语音中谱熵值最大的一帧作为噪声帧；(2)噪声更新，当新的一帧到来时，将其谱熵值与前一噪声帧的谱熵值进行比较，若两者比值在一个阈值内(根据实验取为 0.95)，就认为该语音帧是新的噪声帧，此时将噪声谱更新为此帧谱值；(3)减谱，在每一帧分别减去该帧的噪声帧估计值就完成消噪。算法流程如下：

输入 含噪语音

输出 去噪后的语音

Step1 对每一语音帧进行 FFT 变换，得到语音频谱 S_i，其中，i=1, 2, …, N

Step2 计算每一帧的谱熵值 $e(S_i)$，将谱熵值最大的一帧 m 作为纯噪声帧即 $Noise=S_m$

Step3 若新的一帧的谱熵值与前一纯噪声帧的比值大于阈值 γ (取为 0.95)，即 $e(S_i)/e(S_m)>\gamma$, i=1, 2, …, N

此时便更新噪声估计 $Noise=S_i$

Step4 利用前面已得到的语音谱 S_i 以及噪声帧估计 Noise 进行减谱即 $\hat{S}(k)=[|Y(k)|-Noise]e^{j\theta y(k)}$，即得到消噪后的增强语音。

5 实验与分析

5.1 实验 1

实验 1 与一种改进减谱法进行对比。首先根据金山词霸词库中 one 到 ten 的 10 个英文单词的读音，采集的语音样本都是在安静的实验室环境下录制的 16 kHz 采样、16 bit 量化的语音。噪声类型为随机产生的高斯白噪声，以不同的信噪比(-10 dB～10 dB)加入到原始样本中，然后分别应用各种方法进行去噪，结果表 1 所示。

表 1 英文单词 10 个数字语音信号的信噪比 dB

语音信号	one	two	three	four	five	six	seven	eight	nine	ten
原信噪比	-10.00	-10.00	-5.00	-5.00	0.00	0.00	5.00	5.00	10.00	10.00
基本减谱法去噪	0.15	0.21	2.74	2.63	4.25	4.16	9.52	9.47	17.28	18.86
改进方法去噪	3.04	2.96	4.62	4.54	5.53	5.64	9.68	9.71	19.35	20.81
本文方法去噪	2.11	2.33	3.81	5.38	8.74	8.99	12.52	13.56	19.50	21.40

从表 1 可见，本文方法能更有效地去除噪声，并在对-5 dB 以上的语音去噪时比基本减谱法及改进方法都有更好的效果，在低于-5 dB 时稍差于改进方法，主要是由于在信噪比小于-5 dB 的含噪语音中，噪声谱的大小已大于语音谱的 3 倍以上，其频谱分布也就非常接近于纯噪声的谱熵，仅利用谱熵在区分语音帧和噪声帧时会出现较多误判，另外由于英文单词大多是单音节词，所以本文算法的噪声更新步骤就没有发挥充分的作用。

5.2 实验 2

实验 2 在 2 个较大词汇量的语音库上进行信噪比的对比。语音库 1 是中科院的 CASIA 汉语情感语料库[4]，共 1 200 条语音样本；语音库 2 是自建库，共 1 433 条语音样本。

2 个语音库的样本都是在安静的实验室环境下录制的 16 KHz 采样、16bit 量化的语音。噪声类型同实验 1。实验结果如表 2 和表 3 所示。

表 2 不同方法在语音库 1 上的信噪比比较 dB

去噪方法	语音 *SNR*				
	-10	-5	0	5	10
MMSE	-1.62	0.86	3.05	4.70	5.73
Weiner	1.07	1.73	2.40	3.56	5.45
基本减谱法	-4.80	-0.83	2.88	5.93	8.02
本文方法	0.53	2.78	5.60	8.51	11.41

表 3 不同方法在语音库 2 上的信噪比比较 dB

去噪方法	语音 *SNR*				
	-10	-5	0	5	10
MMSE	0.29	3.52	6.53	8.60	9.63
Weiner	1.44	2.32	3.20	5.10	9.25
基本减谱法	-5.69	-0.97	3.75	8.23	12.28
本文方法	0.95	3.87	7.14	11.11	14.89

可见，对-5 dB 及以上的含噪语音进行消噪，本文方法消噪后的信噪比是最大的，而在小于-5 dB 以下时则没有 Weiner 滤波法效果好，主要原因如下：在信噪比小于-5 dB 的含噪语音中，噪声谱的大小已大于语音谱的 3 倍以上，其频谱分布也就非常接近于纯噪声的谱熵，所以本文方法在区分语音帧和噪声帧时会出现较多误判；而 Weiner 滤波法在低信噪比下对含噪信号进行更强地抑制(相当于过度减谱)，这样去噪后的语音幅度较小、听觉舒适度较好，而且体现在信噪比上是信噪比提升较大。但由于对含噪语音的过度抑制会消去有用的信号，这样在滤波以后就引入了新的频谱失真，会引起语音识别率的下降。

6 结束语

本文提出一种将谱熵用于噪声估计的改进减谱法，该方法是谱熵与减谱法的一种结合。实验结果表明，该方法在不同噪声环境下均有较好的效果，运用在语音情感识别中对识别率也有所提高。

本文只考虑了在高斯白噪声情况下的应用，下一步的工作是分析在不同噪声情况下的应用；另外还可考虑谱熵结合其他的特征分析方法进行消噪，以进一步提高消噪效果。

参考文献

[1] Shen Jialin, Jeihweih H, Linshan L. Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments[C]//Proc. of ICSLP'98. Sydney, Australia: ASSTA, 1998.

[2] Kitaoka N, Yamamoto K, Kusmizu T, et al. Development of VAD Evaluation Framework CENSREC-1-C and Investigation of Relationship between VAD and Speech Recognition Performance[C]//Proc. of ASRU'07. Kyoto, Japan: IEEE Signal Processing Society, 2007.

[3] 马义德, 邱秀清, 陈昱莅, 等. 改进的基于听觉掩蔽特性的语音增强[J]. 电子科技大学学报, 2008, 37(2): 255-257.

[4] 中国科学院自动化所. CASIA 汉语情感语料库[Z]. http://www.chineseldc.org. 2005.

编辑 金胡考