

一种快速自适应噪声谱估计方法

焦人杰, 侯丽敏

(上海大学通信与信息工程学院, 上海 200072)

摘要: 提出了一种快速自适应的噪声谱估计方法。该方法在 Doblínger 和 Cohen 的噪声谱估计方法基础上, 提出了改进算法, 修正了带噪语音子带中语音存在的判断门限和噪声谱更新公式, 估计的噪声谱既能快速适应背景噪声变化, 又能保证准确性。实验结果表明, 结合文中噪声谱估计方法构成的语音增强系统能有效抑制带噪语音中平稳和非平稳噪声, 对于噪声突变的情况也适用。

关键词: 噪声谱估计; 最小值追踪算法; 语音增强

中图分类号: TN912.35

文献标识码: A

文章编号: 1000-3630(2007)-04-0735-06

Fast adaptive estimation of noise spectrum

JIAO Ren-jie, HOU Li-min

(Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

Abstract: A fast adaptive method for noise spectrum estimation is proposed in this paper. We improve the algorithm of noise spectrum estimation proposed by Doblínger and Cohen. This is achieved by modifying the threshold for judging presence of speech in sub-bands and the formula of noise spectrum update. The estimated noise spectrum can quickly adapt with background noise and keep accuracy at the same time. Results show that, by using this method, a speech enhancement system has good performance both in stationary and non-stationary noise suppression, and also under sudden change of noise.

Key words: noise spectrum estimation; minimum tracking algorithm; speech enhancement

1 引 言

语音容易受到噪声的污染, 致使语音质量下降, 语音系统性能恶化。语音增强作为一种预处理技术, 能有效解决噪声污染, 改善语音系统性能。

噪声谱估计是语音增强系统中一个重要组成部分。传统的噪声谱估计方法先对带噪语音进行语音活性检测, 区分语音段和非语音段, 然后根据非语音段的功率谱来估计噪声谱。该方法简单, 在背景噪声平稳的情况下是一种有效的噪声谱估计方法, 但它在语音段时不能更新噪声谱, 因此不适用非平稳噪声, 且在信噪比低于 6dB 情况下, 语音活性检测本身的性能也会急剧下降。近些年, 人们提出了一些不

需要语音活性检测的噪声谱估计方法。Doblínger^[1]通过比较前帧带噪语音子带平滑功率谱最小值和后帧带噪语音子带平滑功率谱之间大小, 对带噪语音子带平滑功率谱最小值每帧进行追踪和更新, 并将其作为估计的噪声谱, 该噪声谱估计方法计算效率高, 能快速适应背景噪声的变化, 但它直接把带噪语音子带平滑功率谱最小值作为估计的噪声谱, 使得估计的噪声谱准确度不高, 同时它会将带噪语音中语音电平的突然上升当作噪声电平突变的情况来处理, 造成噪声谱过估计。Martin^[2]提出了统计最小的噪声谱估计方法, 每帧带噪语音子带平滑功率谱的最小值是在一定时间的长窗内进行追踪并更新, Martin 在文献[2]的基础上, 根据带噪语音平滑功率谱和噪声谱均方误差最小的条件, 计算随时频变化的平滑参数对带噪语音功率谱进行最优平滑, 获得噪声谱估计, 提高了精确度^[3]。Cohen^[4]提出最小值控制递归平均法——MCRA 估计噪声谱, 该方法根据带噪

收稿日期: 2006-08-08; 修回日期: 2006-11-26

作者简介: 焦人杰(1982-), 男, 硕士研究生, 研究方向: 语音信号数字处理和语音增强。

通信作者: 侯丽敏, E-mail: lmhou@staff.shu.edu.cn

语音子带平滑功率谱与其最小值之间比值,计算带噪语音子带中语音出现概率,调节时频平滑参数,加权带噪语音功率谱来估计噪声谱。后两种噪声谱估计方法保证了噪声谱估计的准确性,但在追踪带噪语音子带平滑功率谱最小值时都采用了固定时间的窗,因此在噪声突变的情况下,估计的噪声谱存在延时。

本文在 Doblinger 和 Cohen 的噪声谱估计方法基础上,提出了改进算法,使得带噪语音子带平滑功率谱的最小值能快速适应背景噪声变化,又保证了噪声谱估计的准确性。文中提出了带噪语音子带中语音存在概率的细化改进判断,修正了噪声谱更新公式,而且进一步利用带噪语音信噪比调节增益因子以改善语音增强效果,与 Doblinger 和 MCRA 两种方法比较,语音增强性能提高。实验结果表明,结合本文的噪声谱估计方法构成的语音增强系统能有效抑制带噪语音中平稳和非平稳噪声,由于采用 Doblinger 的方法更新带噪语音子带平滑功率谱最小值,因此对于噪声突变的情况也能快速跟踪,证明了该噪声谱估计方法的有效性。

2 噪声谱估计

本文提出的噪声谱估计方法流程如图 1 所示:

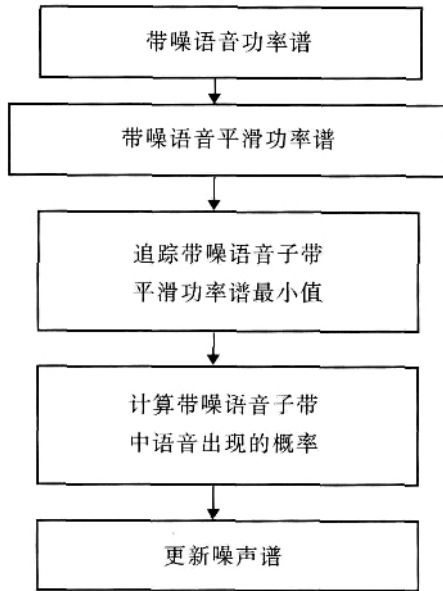


图 1 噪声谱估计流程

Fig.1 Flow diagram of proposed noise estimation algorithm

2.1 追踪带噪语音子带平滑功率谱最小值

设加性噪声 $d(n)$ 与纯净语音 $y(n)$ 不相关,则带噪语音 $x(n)$ 表示为 $x(n)=y(n)+d(n)$, 它们各自短时功率谱为 $|D(\lambda, k)|^2$ 、 $|Y(\lambda, k)|^2$ 和 $|X(\lambda, k)|^2$, 则有

$$|X(\lambda, k)|^2 = |Y(\lambda, k)|^2 + |D(\lambda, k)|^2 \quad (1)$$

其中 λ 为帧号, k 为子带序号。

用一阶递归公式计算带噪语音平滑功率谱 $P(\lambda, k)$, 则有

$$P(\lambda, k) = a_{py} P(\lambda - 1, k) + (1 - a_{py}) |X(\lambda, k)|^2 \quad (2)$$

其中 a_{py} 为平滑因子, 一般取值 0.7 至 0.9, 本文中取值 0.7。因为直接功率谱计算方法计算的短时功率谱包络方差大, 为了减小短时功率谱包络方差对追踪带噪语音子带平滑功率谱最小值的影响, $|X(\lambda, k)|^2$ 用多窗口功率谱计算方法^[5]计算。

本文采用 Doblinger 的带噪语音子带平滑功率谱最小值追踪算法, 对每帧带噪语音子带平滑功率谱最小值 $P_{\min}(\lambda, k)$ 进行追踪并更新, 具体如下:

如果 $P_{\min}(\lambda - 1, k) < P(\lambda, k)$

则 $P_{\min}(\lambda, k) = \gamma P_{\min}(\lambda - 1, k) +$

$$\frac{1 - \gamma}{1 - \beta} [P(\lambda, k) - \beta P(\lambda - 1, k)]$$

否则 $P_{\min}(\lambda - 1, k) = P(\lambda, k)$ (3)

其中 β 和 γ 为经验常数, 本文中 $a_{py}=0.7, \beta=0.96, \gamma=0.998$ 。

2.2 计算带噪语音子带中语音出现的概率

本文计算带噪语音子带中语音出现概率的方法与 Cohen 的 MCRA 方法类似。在文献[4]中, Cohen 根据带噪语音子带平滑功率谱与其最小值的比值来判断带噪语音子带中是否存在语音, 并计算语音出现的概率, 当比值大于门限判定存在语音, 比值小于门限判定不存在语音, 其中门限值为与频率无关的经验常数 $\delta=5$ 。

语音信号的频谱能量主要集中在第一共振峰周围, 并随频率的升高逐渐减小, 因此本文中门限值的设置应与频率有关。在低频段, 门限设置应较低; 高频段时, 门限应较高。设带噪语音子带平滑功率谱与其最小值的比值为

$$S(\lambda, k) = P(\lambda, k) / P_{\min}(\lambda, k) \quad (4)$$

门限值设置为:

$$\delta(k) = \begin{cases} 1.3 & k \in LF \\ 0.0578125k - 0.63 & k \in MF \\ 5 & k \in HF \end{cases} \quad (5)$$

其中 LF 为 0 至 1kHz, MF 为 1kHz 至 3kHz, HF 为 3kHz 至语音抽样频率的一半。

判断带噪语音子带中语音存在的函数 $I(\lambda, k)$ 为

$$I(\lambda, k) = \begin{cases} 1 & S(\lambda, k) > \delta(k) \\ 0 & S(\lambda, k) \leq \delta(k) \end{cases} \quad (6)$$

然后用 $I(\lambda, k)$ 计算带噪语音子带中语音出现概率

的函数 $p(\lambda, k)$

$$p(\lambda, k) = a_{pi} p(\lambda - 1, k) + (1 - a_{pi}) I(\lambda, k) \quad (7)$$

其中 $a_{pi} = 0.2$ 。

2.3 更新噪声谱

根据带噪语音子带中语音出现的概率更新噪声谱。本文对噪声谱更新公式进行了改进。更新的噪声谱包含两部分: 带噪语音子带平滑功率谱最小值和带噪语音的功率谱。更新公式如下:

$$|\hat{D}(\lambda, k)|^2 = p(\lambda, k) |\hat{D}(\lambda - 1, k)|^2 + [1 - p(\lambda, k)] [\eta P_{\min}(\lambda, k) + (1 - \eta) |X(\lambda - 1, k)|^2] \quad (8)$$

本文中 $\eta = 0.8$ 。根据公式(8), 可以看出若带噪语音子带中语音出现的概率小, 则更新的噪声谱中当前帧占的比例大; 若带噪语音子带中语音出现的概率大, 则更新的噪声谱中当前帧占的比例小。

3 语音增强系统

本文采用基于语音信噪比的语音增强系统, 其中语音增强系统如图 2 所示:

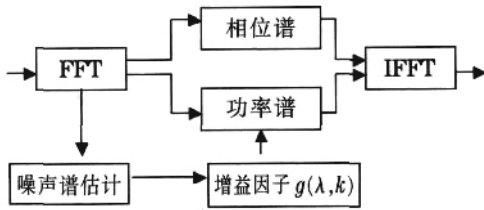


图 2 语音增强系统

Fig.2 Speech enhancement system

根据公式(1), 语音短时功率谱 $|Y(\lambda, k)|^2$ 可以表示为:

$$|Y(\lambda, k)|^2 = \frac{R(\lambda, k)}{R(\lambda, k) + 1} |X(\lambda, k)|^2 \quad (9)$$

其中 $R(\lambda, k) = \frac{|Y(\lambda, k)|^2}{|D(\lambda, k)|^2}$ 。设增益因子 $g(\lambda, k) =$

$\frac{R(\lambda, k)}{R(\lambda, k) + 1}$, 则公式(9)简化为:

$$|Y(\lambda, k)|^2 = g(\lambda, k) |X(\lambda, k)|^2 \quad (10)$$

一般希望语音信噪比高的带噪语音帧经过增强处理后语音失真小, 语音信噪比低的带噪语音帧经过增强处理后残留噪声小, 因此语音信噪比高时增益因子接近 1, 语音信噪比低时增益因子接近 0。本文把增益因子 $g(\lambda, k)$ 设为语音信噪比的函数, 即

$$g(\lambda, k) = \frac{R(\lambda, k)}{R(\lambda, k) + u_k} \quad (11)$$

其中 u_k 为语音信噪比的函数。设语音信噪比 $SNR_{dB} =$

$$10 \lg \frac{\sum_{k \in N} |Y(\lambda, k)|^2}{\sum_{k \in N} |D(\lambda, k)|^2}, \quad \text{式中 } N \text{ 为带频语音的子带数,}$$

根据文献[6]

$$\mu_k = \begin{cases} 1 & SNR_{dB} \geq 20 \\ 8.2 - \frac{9SNR_{dB}}{25} & 20 > SNR_{dB} > -5 \\ 10 & SNR_{dB} \leq -5 \end{cases} \quad (12)$$

4 实验仿真

对本文提出的噪声谱估计方法进行测试, 并与文献[1]和文献[4]比较。实验仿真中, 纯净语音为“北风与太阳”; 语音抽样频率为 8kHz, 帧长 32ms, 帧间重叠 50% 带噪语音功率谱的子带数 $N=256$, 信噪比为全局信噪比, 计算公式如下:

$$SNR_{in} = 10 \lg \frac{\sum y^2(n)}{\sum d^2(n)}$$

$$SNR_{out} = 10 \lg \frac{\sum y^2(n)}{\sum [y(n) - \hat{y}(n)]^2} \quad (13)$$

其中 $y(n)$ 为纯净语音, $\hat{y}(n)$ 为增强语音, $d(n)$ 为真实噪声。

分段相关估计误差^[7](Segmental Relative Estimation Error) 是衡量估计噪声谱与真实噪声谱之间误差的指标, 计算公式为:

$$SegErr = \frac{1}{L} \sum_{\lambda=0}^L \frac{\sum_{k=1}^N [|D(\lambda, k)|^2 - |\hat{D}(\lambda, k)|^2]^2}{\sum_{k=1}^N [|D(\lambda, k)|^2]^2} \quad (14)$$

公式(14)中 L 为语音的总帧数, $|D(\lambda, k)|^2$ 为真实噪声谱, $|\hat{D}(\lambda, k)|^2$ 为估计噪声谱。

表 1 为三种噪声谱估计方法在不同信噪比噪声环境下, 估计的噪声谱与真实噪声谱之间的分段相关估计误差, 其中背景噪声选取 white 噪声、babble 噪声和 factory 噪声。

表 2 为三种噪声谱估计方法在噪声突变环境下, 估计的噪声谱与真实噪声谱之间的分段相关估计误差。

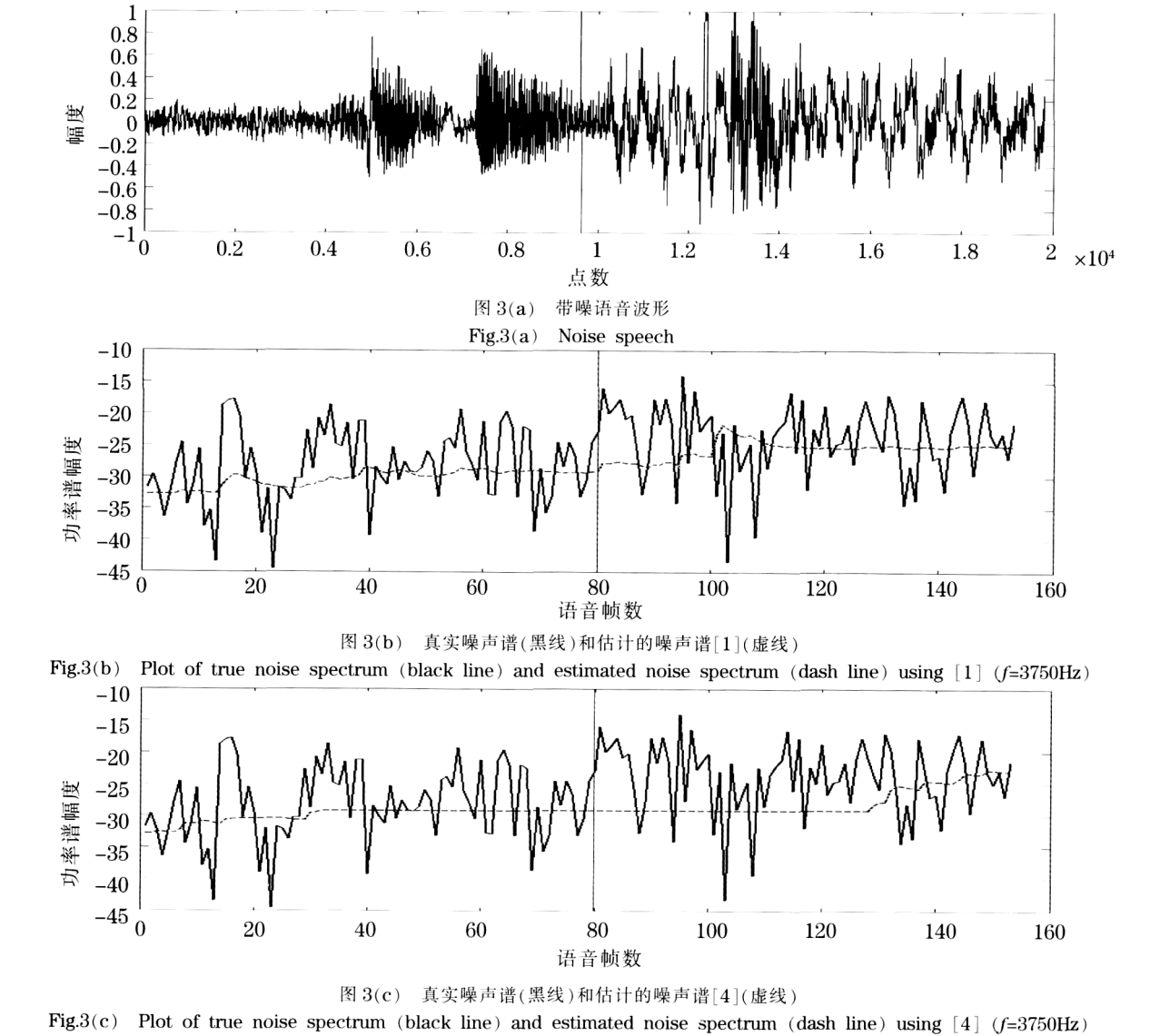
根据表 1 和表 2 看出: 本文噪声谱估计方法估

表 1 不同信噪比噪声环境下估计噪声谱与真实噪声谱之间的分段相关估计误差
Table 1 SegErr between estimated noise spectrum and true noise spectrum under different SNR

	white噪声			babble噪声			factory噪声		
	- 5dB	0dB	5dB	- 5dB	0dB	5dB	- 5dB	0dB	5dB
文献[1]	0.649	0.648	0.668	1.206	4.605	6.003	0.601	2.106	4.603
文献[4]	0.606	0.608	0.618	0.847	1.054	2.015	0.950	1.310	2.045
本文方法	0.420	0.417	0.416	0.641	0.677	1.144	0.537	0.599	1.071

表 2 噪声突变环境下估计噪声谱与真实噪声谱之间的分段相关估计误差
Table 2 SegErr between estimated noise spectrum and true noise spectrum under noise changed quickly condition

	white噪声从 5dB 变化至 0dB	babble噪声从 5dB 变化至 0dB	factory噪声从 5dB 变化至 0dB	0dB babble 噪声变化 至 0dB factory 噪声	5dB babble 噪声变化 至 0dB factory 噪声
文献[1]	0.726	6.965	6.489	5.731	6.841
文献[4]	0.655	1.344	1.101	1.023	1.332
本文方法	0.469	0.730	0.635	0.704	0.734



计的噪声谱与真实噪声谱之间的分段相关误差小于文献[1]和文献[4],在非平稳噪声环境下及噪声突变环境下尤为明显。从分段相关估计误差上看,本文噪声谱估计方法优于文献[1]和文献[4]。

图 3 为噪声突变环境下噪声谱估计的实例,其中带噪语音前 1.3s 为 5dB 的 babble 噪声,后 0.8s 为 0dB 的 factory 噪声。图 3(a)为带噪语音波形,图 3(b)为采用文献[1]估计的噪声谱,图 3(c)为采用文献

[4] 估计的噪声谱, 图 3(d) 为采用本文方法估计的噪声谱。

观察图 3 可以得出, 本文噪声谱估计方法估计的噪声谱能及时追踪噪声突变(图中标注竖线处), 不产生明显噪声谱延时, 且精确度优于文献[1]和文献[4]估计的噪声谱。

结合三种噪声谱估计方法对带噪语音进行语音增强处理, 语音增强系统如图 2 所示。表 3 为不同

信噪比 white 噪声、babble 噪声和 factory 噪声环境下, 带噪语音增强前后的输入输出信噪比。

根据表 3 可以看出: 结合本文噪声谱估计方法的带噪语音增强信噪比改进量优于文献[1]和文献[4]。为了直观进行比较, 图 4 为信噪比 5dB white 噪声情况下, 分别结合三种噪声谱方法处理后的增强语音波形图。

观察图 4 得出: 结合本文噪声谱估计方法处理

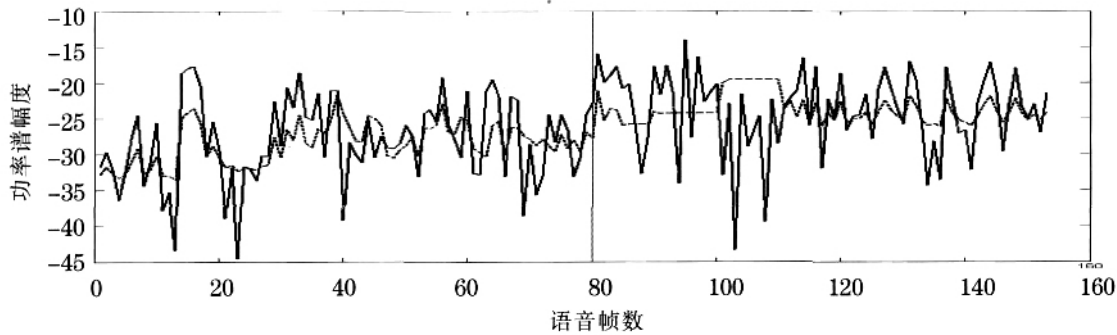


图 3(d) 真实噪声谱(黑线)和采用本文方法估计的噪声谱(虚线)
Fig.3(d) Plot of true noise spectrum (black line) and estimated noise spectrum (dash line) using method proposed here ($f=3750\text{Hz}$)

表 3 不同信噪比情况下增强语音输出信噪比
Table 3 output SNR under different input SNR

	white噪声			babble噪声			factory噪声		
	5dB	0dB	- 5dB	5dB	0dB	- 5dB	5dB	0dB	- 5dB
文献[1]	4.89	2.32	0.91	4.53	1.88	0.42	3.45	1.56	0.54
文献[4]	5.12	2.12	0.97	5.07	2.05	0.54	4.55	1.89	0.67
本文方法	5.81	3.10	1.22	6.13	3.56	1.54	6.12	3.22	1.43

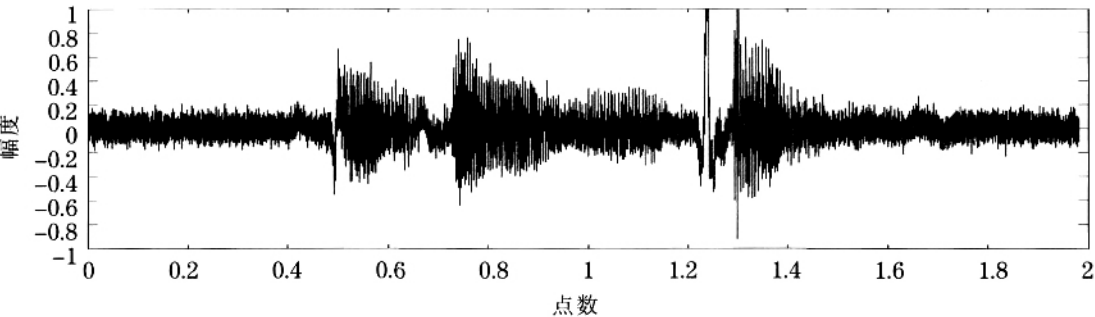


图 4(a) 带噪语音(5dB white 噪声)
Fig.4(a) Noise speech with 5dB white noise

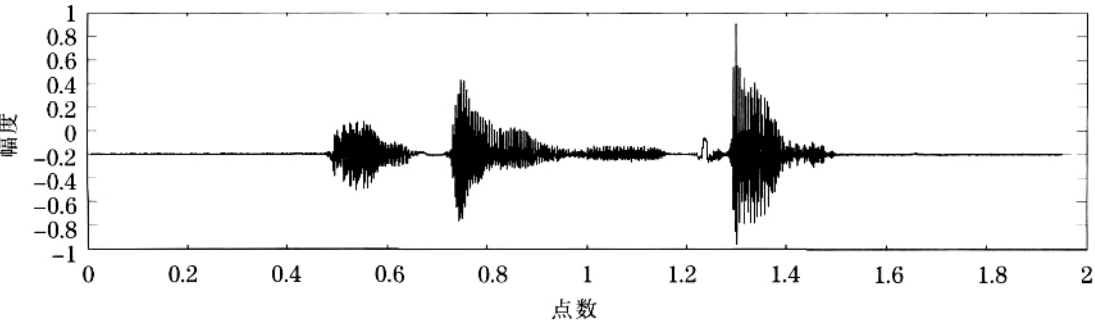


图 4(b) 增强语音结合本文噪声谱方法
Fig4.(b) Enhanced speech using method proposed here

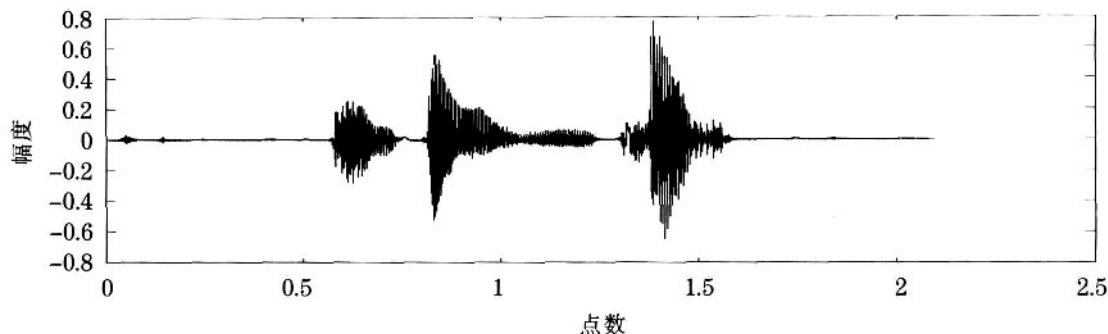


图 4(c) 增强语音结合文献[4]噪声谱方法

Fig.4(c) Enhanced speech using [4]

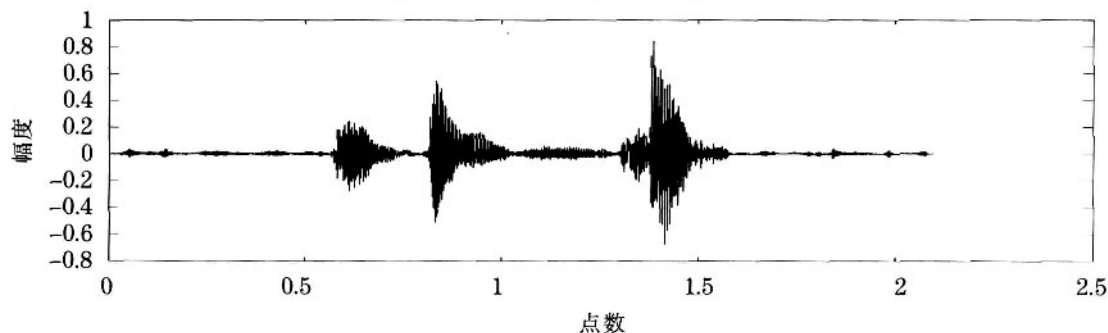


图 4(d) 增强语音结合文献[1]噪声谱方法

Fig.4(d) Enhanced speech using [1]

的增强语音没有残余噪声,结合文献[1]和文献[4]中噪声谱估计方法的增强语音有残余噪声,同时由于增强算法原因,增强语音较纯净语音波形幅度略有减小。

5 结 论

由于噪声复杂多变,语音增强是一项具有挑战性的课题。实验表明,本文提出的噪声谱估计方法能快速追踪背景噪声变化且准确性高,结合该方法构成的语音增强系统在保持较好语音质量的同时,能够有效抑制带噪语音中平稳和非平稳噪声,且对于噪声突变的情况也能适用,有一定通用性。

参 考 文 献

[1] Dobliger G. Computationally efficient speech enhancement

- by spectral minima tracking in sub-bands[A]. in EUROSP-EECH 95[C]. Madrid, Spain, 1995, 9(18-21): 1513-1516.
- [2] Martin R. Spectral subtraction based on minimum statistics[A]. Seventh European Signal Processing Conference [C]. 1994, 9: 1182-1185.
- [3] Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics[J]. Speech and Audio Processing, 2001, 9(7): 504-512.
- [4] Cohen I, Berdug B. Noise estimation by minima controlled recursive averaging for robust speech enhancement[J]. Signal Processing, 2002, 9(1):12-15.
- [5] Walden A T, Percival D B, McCoy E J. Spectrum estimation by wavelet thresholding of multitaper estimators[J]. Signal Processing, 1998, 46(12): 3153-3165.
- [6] HU Y, Loizou P. Speech enhancement based on wavelet thresholding and multitaper spectrum[J]. Speech and Audio Processing, 2004, 12(1): 59-67.
- [7] Cohen I. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging[J]. Speech and Audio Processing, 2003, 11(5): 466-475