· 计算机工程与应用 ·

改进的基于听觉掩蔽特性的语音增强

马义德, 邱秀清, 陈昱莅, 刘映杰, 朱敬锋

(兰州大学信息科学与工程学院 兰州 730000)

【摘要】该文利用人耳感知的掩蔽特性,研究了一种改进的基于人耳听觉掩蔽特性的语音增强方法。该方法改进了对含噪语音中噪声的估计,即在估计噪声时,先在帧上判断语音的有无,然后在频率点上判断语音的有无,最后再从感知的角度,利用估计出的语音信号计算掩蔽阈值,以此在时域和频域上合理地调整增强系数。仿真结果表明,该方法不仅使信噪比有较大的提高,而且减少了语音失真,同时也很好地抑制了背景噪声和残余音乐噪声。

关键 词 听觉掩蔽;噪声估计;减谱算法;语音增强中图分类号 TP391 文献标识码 A

An Improved Method for Speech Enhancement Based on Human Auditory Masking Properties

 $MA\ Yi\text{--de},\ QIU\ Xiu\text{--qing},\ CHEN\ Yu\text{--li},\ LIU\ Ying\text{--jie},\ and\ ZHU\ Jing\text{--feng}$

(School of Information Science & Engineering, Lanzhou University Lanzhou 730000)

Abstract This paper suggests an improved speech enhancement method based on human auditory masking properties. In this method, an improved noise estimation algorithm is used, then, the estimated results can be used to compute the masking threshold of the speech, after that, the coefficients of time and frequency are adjusted according to perception. The results show that the improved method leads to better signal to noise ratio, significant reduction of background noise, and unnatural structure of the residual noise.

Key words auditory masking; noise estimate; spectral subtraction algorithm; speech enhancement

语音识别是下一代操作系统的核心技术,而语音增强是语音识别系统普及化与实用化的主要瓶颈问题。目前,在平稳的声学环境以及信噪比较大的情况下可以得到较好的增强效果,但在非平稳环境下,低信噪比的增强仍然是一项具有挑战性的工作。

处理宽带噪声最通用的方法是减谱法^[1],但会产生较大的残余"音乐噪声"。众所周知,人类听觉系统具有掩蔽效应,噪声信号的存在之所以影响到人耳对纯语音信号的接收,从本质上讲是因为噪声信号对纯语音信号产生了掩蔽,语音增强就是要减少甚至消除噪声信号对纯语音信号的掩蔽。文献[2-4]提出了模拟人耳的听觉系统来进行语音增强的方法。在众多基于掩蔽效应的增强方法中,对噪声的估计大都使用了文献[5-6]提出的"MS"方法。但经试验发现这种噪声估计方法并不是最好的,一些学者也提出了一些其他算法^[7],收到了一定的效果,

但仍不理想。而对噪声的估计不当,则会使得对后续的掩蔽参数调整出现偏差,从而影响增强效果;若对噪声估计合理,则可以得到很好的增强效果。为此本文引入了一种可以快速跟踪外界环境变化的噪声估计方法来估计噪声,该方法先在帧上判断语音的有无,而后在频率点上判断语音的有无,然后利用估计出的噪声同时结合掩蔽效应自适应设定增强系数。仿真结果表明这样的结合在提高信噪比、抑制背景噪声、减少语音失真等方面取得了较好的效果。

1 算法描述

假设含噪语音信号y(n)表示为: y(n) = s(n) + d(n), 其中s(n)为纯净语音信号; d(n)为加性噪声。由于增强是按帧进行的,所以可把上述模型写成帧的形式:

收稿日期: 2006-09-18; 修回日期: 2007-06-20

基金项目: 国家自然科学基金(60572011); 985特色项目(LZ985-231-58262)

作者简介: 马义德(1963-), 男, 教授, 博士生导师, 主要从事计算机应用系统、生物医学图像处理、DSP与信号处理等方面的研究.

$$y(m,n) = s(m,n) + d(m,n)$$

 $m=1,2,\dots,N; n=0,1,\dots,N-1$

式中 m 为帧号; N 为帧长, 对其进行傅里叶变换得 Y(m,k) = S(m,k) + D(m,k)。引用一种常用的语音增强减谱法模型^[3]:

$$|\hat{S}(m,k)| = G(m,k) |Y(m,k)|$$
 (1)

且当 $SNR_{post(m,k)}$ < $(1/\alpha(m,k)+\beta(m,k))$ 时有:

$$G(m,k) = \sqrt{1 - \alpha(m,k) \text{SNR}_{post(m,k)}}$$
 (2)

否则

$$G(m,k) = \sqrt{\beta(m,k) \text{SNR}_{post(m,k)}}$$
 (3)

式中 $SNR_{post(m,k)} = \hat{D}(m,k)|^2/|Y(m,k)|^2$; $\alpha(m,k)$ 和 $\beta(m,k)$ 分别为时间和频率的函数, $\alpha(m,k)$ 用于控制增强语音的残留音乐噪声,其值根据人耳的听觉掩蔽特性确定;而系数 $\beta(m,k)$ 用于控制增强语音的信噪比。其值根据最小均方误差准则确定。具体计算过程参看文献[2],本文直接引用结果:

$$\alpha(m,k) = \frac{T(m,k)}{\Gamma(\lambda/2+1)\left| (\xi(m,k) - \rho(m,k) + 1)E\left|D_{m,k}^{2}\right| \right|}$$
(4)

$$\beta(m,k) = 1 + \left(\frac{\alpha(m,k) - 1}{\alpha(m,k)}\right) \xi^{\lambda/2}(m,k)$$
 (5)

式中 $\rho(m,k) = E[Y_{m,k}^2]/E[D_{m,k}^2]$,称为语音信号的后验信噪比; $\xi(m,k) = E[X_{m,k}^2]/E[D_{m,k}^2]$,称为语音信号的先验信噪比。估计出噪声后, $\rho(m,k)$ 可直接计算,由于不能求得纯净语音的功率谱, $\xi(m,k)$ 的估计用下式计算:

$$\xi(m,k) = \eta \ \xi(m-1,k) +$$

$$1 - \eta \frac{\max((Y_{m,k}^2 - E[D_{m,k}^2]), 0)}{E[D_{m,k}^2]}$$
(7)

1.1 噪声的估计^[8-9]

先在三个频率子带(0~1000 Hz、1000~3000 Hz、3000 Hz以上)上计算带噪语音与估计噪声的信噪比,当这三个子带上的信噪比都小于某个门限 σ 时,该帧就被认为是非语音帧,噪声被更新为: $D(m,k)=\varepsilon D(m-1,k)+(1-\varepsilon)|Y(m,k)|^2$;否则被认为是语音帧。再按照频率点来判断语音是否存在,具体方法为:先找出当前带噪语音的最小能量 $p_{\min}(m,k)$,如果 $p_{\min}(m-1,k)< p(m,k)$,那么 $p_{\min}(m,k)=p_{\min}(m-1,k)+((1-\gamma)/(1-\beta))(p(m,k)-\beta p(m-1,k))$,否则 $p_{\min}(m,k)=p(m,k)$;然后计算带噪语音能量与当前最小能量的比值 $S_r(m,k)=p(m,k)/p_{\min}(m,k)$,将该比值与一个依赖于频率的

门限 $\delta(k)$ 相比,若大于该门限则认为在这个频率点上包含了语音,此时平滑参数 $\alpha_s(m,k) = \alpha_1$;否则没有包含语音, $\alpha_s(m,k) = \alpha_2$ (其中的 α_1 、 α_2 均为经验值)。至此噪声利用下式确定为:

$$D(m,k) = \alpha_s(m,k)D(m-1,k) + (1 - \alpha_s(m,k))|Y(m,k)|^2$$
(8)

1.2 掩蔽阈值的计算[2]

本文首先将语音信号进行临界频带分析或Bark分析。利用FFT求得语音信号功率谱E(m,k),将语音信号的频率域划分为不同的临界频带,语音在各个临界频带的能量 $B_i = \sum_{k=h}^{bh_i} E(m,k)$,其中 $bh_i \setminus bl_i$ 分别

表示临界频带i的上限和下限, $i=1,2,\cdots,i_{max}$ 是临界带号。考虑到各个频带间相互掩蔽的影响,定义如下的传播函数:

 $SF_{ij}=15.81+7.5(\Delta+0.474)-17.5\sqrt{1+(\Delta+0.474)}^{2}$ (9) 式中 $\Delta=i-j$, $i,j=1,2,\cdots,i_{\max}$ 且 $|\Delta| \leqslant i_{\max}$ 。 则考虑频带间影响后的能量为:

$$C_i = \sum_{i=1}^{l_{\text{max}}} B_i \times SF_{ij}$$

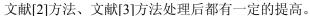
由于噪声和音调的掩蔽特性不同,因此首先判断各个频带是噪声还是音调,可以根据谱平坦度(SFM= μ_g / μ_a)来判断,其中 μ_g 、 μ_a 分别表示各个频带功率谱的几何和算术平均,且SFM \in [0,1],为0表示具有纯音调特性,纯音调的掩蔽阈值偏移量为(14.5+i)dB,为1表示具有白噪声特性,白噪声的掩蔽阈值偏移量为5.5 dB,根据SFM的定义式,定义音调系数 ϕ = min(SFM $_{dB}$ /-60,1),则相对的掩蔽阈值偏移量为: O_i = ϕ (14.5+i)+5.5(1- ϕ)dB,此时掩蔽阈值T(m,i)= $10^{\lg(C_I-(O_I/10))}$ 。在各个临界频带上,语音信号具有相同的掩蔽特性,将T(m,i)扩展到各个频谱上,记为T'(m,i)。则最终的掩蔽阈值T(m,i)= max (T'(m,i), $T_a(m,i)$),其中 $T_a(m,i)$ 为绝对听阈,定义为: 3.6 $f^{-0.8}$ -7 exp(f -3.3) 2 + 0.007 f^4 。

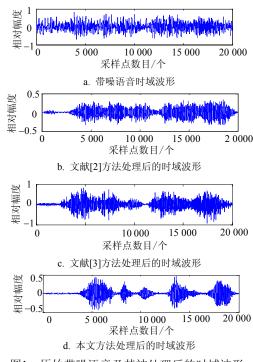
2 结果及性能评价

2.1 实验一

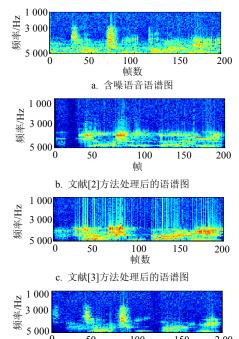
实验一所用语音是在马路边随机录取的带噪语音。本文将录制的带噪语音、文献[2]方法处理结果、文献[3]方法处理结果、本文方法处理结果如图1和图2所示。

通过图1、图2不难发现,对含噪语音在使用本 文方法处理后,从时域结果图以及语谱图上比采用





原始带噪语音及其被处理后的时域波形



d. 本文方法处理后的语谱图 原始带噪语音及其被处理后的语谱图

100

帧数

150

2.00

50

2.2 实验二

实验二所用语音是在安静的环境下录取的纯净 语音,背景噪声为加入的高斯白噪声。对输入信噪 比分别为: -5 dB、0 dB、5 dB的带噪语音信号分别 采用文献[2]方法、文献[3]方法以及本文方法进行语 音增强,并对输出信噪比进行比较。

增强语音的信噪比对比 表1

	信噪比/dB				
输入的带噪语音	-5	0	5		
文献[2]方法处理结果	5.526 4	9.321 4	11.354 6		
文献[3]方法处理结果	5.035 4	8.867 9	10.983 2		
本文方法处理结果	9.151 2	10.754 8	14.354 1		

从表1的试验结果可看到,本文方法在提高信噪 比上比文献[2]方法、文献[3]方法都有了一定的提 高;尤其是输入为低信噪比语音时,本文方法对信 噪比的改善效果更加明显。

实验三 2.3

考虑语音信号的分段失真, 时域形式的增强语 音信号失真度D定义为[10]:

$$D = \frac{1}{L} \sum_{m=1}^{L} \left(\sum_{n=1}^{N} [s(m,n) - s'(m,n)]^{2} / \sum_{n=1}^{N} [s(m,n)]^{2} \right)$$

式中 s(n, i)为纯净语音; s'(n, i)为增强后语音; N为帧长; L为总帧数。在安静的环境下录取的纯净语 音中加入四种不同类型的噪声(leopard、destroyer engine、factory2、f16),得如表2所示对比结果(噪声 来源于噪声库NOISEX-92)。

不同噪声环境下语音失真度(D)比较

	Leopard	factory2	destroyer engine	f16
文献[2]方法处理结果	0.428 6	0.402 3	0.359 7	0.202 4
文献[3]方法处理结果	0.463 5	0.412 3	0.371 6	0.210 3
本文方法处理结果	0.263 5	0.232 8	0.203 2	0.165 4

从表2的数据可看到,在各个噪声环境下,本文 的方法对语音造成的失真比文献[2]、文献[3]中的字 方法都小。

为了确认客观性评估,本文进行了主观听觉测 试,通过将增强语音进行非正式主观试听,可以感 觉到音乐噪声以及背景噪声有所减少, 同时语音的 失真也有所下降, 所得的结果更加容易让人接受。

结 论 3

本文针对目前众多基于掩蔽效应的语音增强方 法中对带噪语音噪声的估计不足,提出了一种改进 的噪声估计方法,同时借助人耳感知的掩蔽特性, 依据噪声的掩蔽门限,在时域和Bark频域上动态调 整增强系数,以达到实现语音增强的目的。最后将 使用本文方法处理结果与使用文献[2]和文献[3]方 法处理结果进行比较, 表明该算法在提高信噪比、

(下转第304页)

关的结论。这些结果初步显示了深入开展音乐情绪 感知的脑电研究的可行性。音乐作为一种有效的情 绪诱发源,不论在情绪感知的心理和生理研究方面, 还是在实际应用方面,都具有很好的应用前景。尤 其是近年来临床上有关音乐疗法以及脑神经科学上 有关脑机接口等方面发展的兴起,使得对音乐、情 绪、脑电之间关系的研究显得更有价值。若能根据 人对不同音乐的情绪反应在脑电节律上的变化规律 来编曲,并用于临床上的音乐治疗,其应用价值是 不言而喻的。

参考文献

- [1] KOELSCH S. Investigating emotion with music[J]. Ann N Y Acad Sci, 2005, 1060: 1-7.
- [2] SCHMIDT L A, TRAINOR L J. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions[J]. Cogn Emot, 2001, 15(4): 487-500.

- [3] DAVIDSON R J, EKMAN P, SARON C D, et al. Approach-withdrawal and cerebralasymmetry: emotional expression and brain physiology I[J]. J Pers Social Psychol, 1990, 58(2): 330-341.
- [4] 孟昭兰. 人类情绪[M]. 上海: 上海人民出版社, 1989.
- [5] KHALFA S, ISABELLE P, JEAN-PIERRE B, et al. Event-related skin conductance responses to musical emotions in humans[J]. Neuroscience Letters, 2002, 328: 145-149.
- [6] EKMAN P, FRIESEN W V, ANCOLI S. Facial signs of emotional experience[J]. Journal of Personality and Social Psychology, 1980, 39: 1125-1134.
- [7] 尧德中. 脑功能探测的电学理论与方法[M]. 北京: 科学出版社, 2003.
- [8] BAUMGARTNER T, ESSLEN M, JANCKE L. From emotion perception to emotion experience: emotions evoked by pictures and classical music[J]. International Journal of Psychophysiology, 2006, 60: 34-43.

编辑税红

.....

(上接第257页)

抑制背景噪声、减少语音失真、提高增强语音的感知质量等方面得到了改进。试验结果表明,阈值的选择具有广泛的适应性,对不同的噪声均有较好的增强效果。

参考文献

- [1] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Transaction on Acoustic Speech and Signal Processing, 1979, 27(2): 113-120.
- [2] 刘海滨, 吴镇扬, 赵 力, 等. 平稳环境下基于人耳听觉 掩蔽特性的语音增强[J]. 信号处理, 2003, 19(4): 303-307.
- [3] 江小平, 姚天任, 傅 华. 基于最小统计量和掩蔽效应的单通道语音增强[J]. 通信学报, 2003, 24(6): 23-31.
- [4] 陈国明. 基于人耳掩蔽效应的语音增强算法研究[D]. 南京: 东南大学, 2005.
- [5] MARTIN R. Spectral subtraction based on minimum statistics[C]//In Proceeding of 7th Euro Signal Processing Conference. Edinburgh: [s. n.], 1994: 1182-1185.

- [6] MARTIN R. Noise power spectral density estimation based on optimal smoothing and minimum statistics[J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(5): 504-512.
- [7] ISRAEL C. Noise estimation by minima controlled recursive averaging for robust speech enhancement[J]. IEEE Signal Processing Letters, 2002, 9(1): 12-15.
- [8] SUNDARRAJAN R. Noise estimation algorithms for highly non-stationary environments[D]. Dallas: University of Texas, 2004.
- [9] SUNDARRAJAN R, PHILIPOS C L, YI Hu. A noise estimation algorithm with rapid adaptation for highly non-stationary environments[C]//In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP'04). Montreal: ICASSP APOS, 2004, 1(1): 305-308.
- [10] 卜凡亮, 王为民, 戴启军, 等. 基于噪声被掩蔽概率的 优化语音增强方法[J]. 电子与信息学报, 2005, 27(5): 753-756.

编辑张俊