

基于数字助听器声音场景分类的噪声抑制算法

汪家冬¹ 邹采荣¹ 蒋本聪¹ 王青云²

(1. 广州大学机械与电气工程学院, 广州, 510006; 2. 东南大学信息科学与工程学院, 南京, 210096)

摘要: 提出了一种基于声音场景分类的噪声抑制算法。算法使用调制滤波法对纯语音、纯噪声和含噪语音 3 种场景进行分类, 并根据分类结果调整噪声抑制算法参数集, 得到不同的抑制系数。本文方法在助听器测试系统中取得了良好的实验效果, 场景分类正确率在 95% 以上。在不同噪声类型情况下, 经过本文算法处理的输出语音信号取得了良好的信噪比和 MOS 评分的提升。本文算法可以有效地提高数字助听器输出语音质量。

关键词: 声音场景分类; 调制滤波; 噪声抑制; 信噪比

中图分类号: TP391.42 **文献标志码:** A

Noise Reduction Algorithm Based on Acoustic Scene Classification in Digital Hearing Aids

Wang Jiadong¹, Zou Cairong¹, Jiang Bencong¹, Wang Qingyun²

(1. School of Mechanical and Electric Engineering, Guangzhou University, Guangzhou, 510006, China; 2. School of Information Science and Engineering, Southeast University, Nanjing, 210096, China)

Abstract: A new noise reduction algorithm based on acoustic scene classification is proposed. Three acoustic scenes of pure speech, noise, noisy speech are classified by modulation filter. The parameters of noise reduction algorithm are adjusted by the result of scene classification. Different attenuation coefficient is adopted according to the different acoustic scene. Satisfied experimental results are achieved in the digital hearing aid testing system. Better than 95% accuracy is acquired in acoustic scene classification experiment. In the environment of different kinds of noises input, the signal-noise ratio (SNR) and MOS score are increased apparently. The quality of output speech in digital hearing aids is improved effectively.

Key words: acoustic scene classification; modulation filter; noise reduction; SNR

引 言

在数字助听器系统中, 语音、噪声和音乐等不同的场景中常常需要调整不同的信号处理策略和参数, 而系统对声音场景进行自动分类的能力则制约了系统性能^[1]。高性能的数字助听器能够根据当前

声音场景自动切换程序,调整参数,处理声音,提高信噪比,改善用户体验^[2]。

近年来,针对数字助听器应用,很多学者对声音场景分类算法进行了研究^[3]。这些方法各有特色,实验所用的数据库也各有不同。很多学者研究声学特征参数集的选取和分类模型的建立^[4]。合理地选取适合区分声音场景的特征能够提高整个分类系统的性能,降低模型的计算量。在这些研究中,短时能量、线性回归系数、过零率、基音频率、共振峰、熵信息以及倒平共谱信息等都是主要使用的特征。很多学者也对声音场景提出各种分类算法,如人工神经网络^[5]、支持向量机(Support vector machine, SVM)、隐马尔可夫模型(Hidden Markov model, HMM)^[6]以及混合高斯模型(Gaussian mixture model, GMM)等。2015年 Ipek Sen 等对这些分类器进行了比较,提出了新的参数设置方法^[7]。但是,基于特征提取和模式识别的方法导致数字助听器计算量变大,实时性变差,在实际系统中常常由于功耗过大无法应用。而基于调制深度的声音场景分类算法^[8]由于其计算量小、实时性高成为研究热点

本文研究了一种基于调制滤波法的数字助听器声音场景分类算法。利用输入声音的调制深度,区分纯语音、噪音和含噪语音 3 种场景。并根据场景分类结果,调整噪声抑制参数,获得不同的抑制系数。本方法可以保证语音信号顺利通过系统,而噪声则受到最大程度的抑制,从而在有噪声的情况下提高信噪比。本文方法在助听器测试系统中取得了良好的实验效果,场景分类正确率在 95% 以上。在不同的噪声类型情况下,经过本文算法处理的输出语音信号取得了良好的信噪比和 MOS 评分^[9]的提升。本文算法可以有效提高数字助听器输出语音质量。

1 基于调制滤波法的声音场景分类算法

在数字助听器中,声音信号经传声器采集、A/D 模块转换成数字信号。声音场景分类算法在全带内对输入数字信号进行处理,其处理得到的场景分类结果供后续噪声抑制模块使用,用来调节抑制函数参数。基于调制滤波法的数字助听器声音场景分类算法流程如图 1 所示,一般以 1 s 为一个观测窗口,也可以根据实际需要调整观测频度。

图 1 中,每帧信号最大值记为 $v(i), i=1, \dots, N, i$ 为帧号, N 为一个观测窗口的帧数。则观测窗口的有效值为

$$V_{\text{rms}} = \sqrt{\frac{1}{N} \left(\sum_{i=1}^N v^2(i) \right)} \quad (1)$$

对每个观测窗口的包络曲线 $v(i)$ 进行微分,有

$$H(z) = 0.5(1 - z^{-1}) \quad (2)$$

微分之后得到去除直流分量的信号 $d_v(i)$ 。根据噪声和语音的频谱特性,将 $d_v(i)$ 通过调制滤波器组分成 3 个调制频带 0~4 Hz, 4~16 Hz, 16~64 Hz, 得到调制滤波后的信号 $d'_{v1}(i)$, $d'_{v2}(i)$ 和 $d'_{v3}(i)$ 。调制滤波器可以使用六阶 IIR 滤波器实现。调制滤波后的信号需要进行修正,得

$$\begin{cases} d_{v1}(i) = K_1 \cdot d'_{v1}(i) \\ d_{v2}(i) = K_2 \cdot d'_{v2}(i) \\ d_{v3}(i) = K_3 \cdot d'_{v3}(i) \end{cases} \quad (3)$$

式中修正系数 K_1, K_2 和 K_3 可以根据不同系统进行修正。

下一步计算 3 个调制频带的调制深度,具体如下

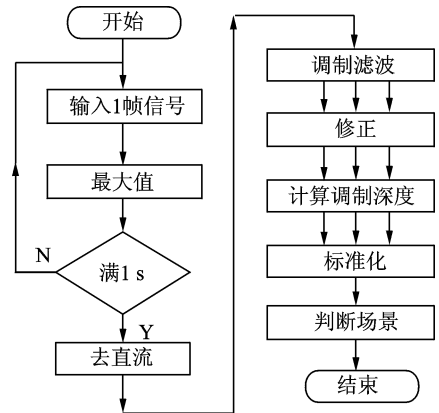


图 1 基于调制滤波法的数字助听器声音场景分类算法流程图

Fig. 1 Algorithm flowchart of acoustic scene classification based on digital hearing aid of modulation filtering method

$$\begin{cases} V_1 = \sqrt{\frac{1}{N} \left(\sum_i^N d'_{v1}(i) \right)} \\ V_2 = \sqrt{\frac{1}{N} \left(\sum_i^N d'_{v2}(i) \right)} \\ V_3 = \sqrt{\frac{1}{N} \left(\sum_i^N d'_{v3}(i) \right)} \end{cases} \quad (4)$$

对3个通道的调制深度进行正则化,具体如下

$$\begin{cases} m_1 = \frac{V_1}{V_{rms}} \\ m_2 = \frac{V_2}{V_{rms}} \\ m_3 = \frac{V_3}{V_{rms}} \end{cases} \quad (5)$$

最后,根据正则化后的调制深度进行声音场景分类,具体如下

$$\begin{cases} m_1 - m_3 \geq TH_{speech} & \text{纯语音} \\ TH_{noise} \leq m_1 - m_3 < TH_{speech} & \text{含噪语音} \\ m_1 - m_3 < TH_{noise} & \text{噪声} \end{cases} \quad (6)$$

式中语音阈值 TH_{speech} 和噪声阈值 TH_{noise} 可以在 $0 \sim 1$ 之间进行调整。

2 基于声音场景分类的数字助听器降噪算法

在上述算法得到数字助听器所处声音场景之后,可以通过调节噪声抑制系数 λ 对输入信号进行降噪,系统框图如图2所示。图2所示的多通道数字助听器中,经过A/D转换的数字声信号首先进行声音场景分类。在通过子带增益进行多通道响度补充之后,增加了噪声抑制单元,通过抑制系数 λ 实现降噪功能,其中 λ 取值在 $0 \sim 1$ 之间,根据声音场景和信噪比进行调整,具体如图3所示。在不同的声音场景中,参数 $K_0, K_1, K_2, A_0, A_1, A_2, B_0$ 和 B_1 取值不同,其取值可以根据实验进行确定,1组典型取值如表1所示。

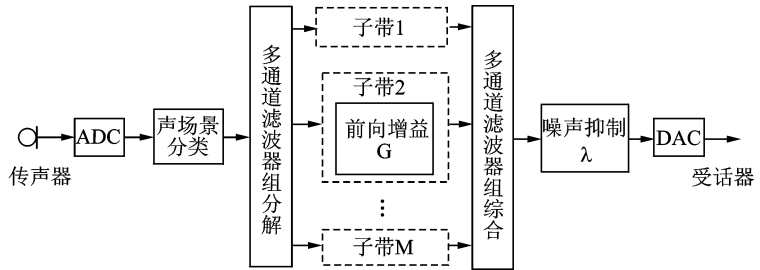


图2 含有降噪单元的数字助听器系统

Fig. 2 Block diagram of digital hearing aid comprising the unit of noise reduction

降噪算法具体步骤为:

表1 不同声音场景下的系数取值表

Tab. 1 Coefficient table with different sound scenes

声音场景	B_0 /dB	B_1 /dB	K_0	A_0	K_1	A_1	K_2	A_2
纯语音	9	18	0.032 8	-0.444 4	0.016 4	-0.296 3	0.005 7	-0.101 3
含噪语音	9	18	0.049 2	-0.666 7	0.024 6	-0.444 4	0.016 4	-0.175 4
噪声	9	18	0.0656	-0.889 9	0.032 8	-0.592 6	0.019 8	-0.237 8

(1) 对每帧信号取绝对值,并对绝对值求平均值 S_{avg} ;

(2) 将平均值转到 dB 域, 得到 $S_{dB} = 20\lg(S_{avg})$;

(3) 计算信号包络

$$l_s = \eta \cdot l'_s + (1 - \eta) \cdot S_{dB} \quad (7)$$

式中: l'_s 为前一帧信号包络, η 为平滑因子, 取值在 0 至 1 之间。

(4) 计算噪声包络

$$\begin{cases} l_n = \xi \cdot l'_n + (1 - \xi) \cdot l_s & l_s > l_n \\ l_n = l_s & \text{其他} \end{cases} \quad (8)$$

式中: l'_n 为前一帧噪声包络, ξ 为平滑因子, 取值在 0 至 1 之间。

(5) 计算信噪比

$$SNR = l_s - l_n \quad (9)$$

根据声音场景选择噪声抑制函数的参数并计算衰减值, 即有

$$\lambda = K_n \cdot SNR + A_n \quad (10)$$

(6) 执行衰减, 抑制噪声。

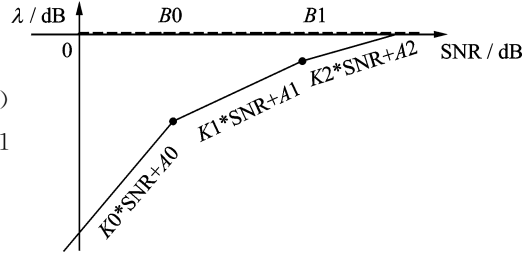


图 3 噪声抑制系数 λ 调节曲线

Fig. 3 λ adjustment curve of noise suppression coefficient

3 实验与仿真

3.1 测试系统组成

测试系统由计算机、测试用助听器算法实时测试电路板、示波器等组成。测试用源信号由计算机软件产生, 通过计算机声卡的 Line out(线路输出)接口输出至测试电路板的 Line in(线路输入)接口, 经测试电路板运行的助听器声音场景分类及降噪算法程序处理后, 由测试电路板的 Line out(线路输出)接口输出, 通过计算机声卡的 Line in(线路输入)接口输入至计算机软件进行分析, 并可以通过示波器观察并测量波形。具体如图 4 所示。

3.2 声音场景分类实验

采用系统事先录制的音频数据库进行了声音场景分类实验。数据库内有纯语音、含噪语音、纯噪音 3 种类型音频文件, 其中纯语音选用人民卫生出版社出版的《汉语普通言语测听 CD》中的句表, 纯噪音采用 NOISEX-92 噪声库中的白噪声(White noise)、坦克车内噪声(Tank noise)、餐厅噪声(speech babble)和 高频信道噪声(HF channel noise), 含噪语音由 SurroundRouter 专业声场景仿真软件进行合成并播放, 信噪比可以调节。其中, 信噪比为 0 dB 情况下的声音场景分类正确率如表 2 所示, 无论是在纯语音、纯噪音还是含噪语音场景, 其分类正确率都超过了 95%。

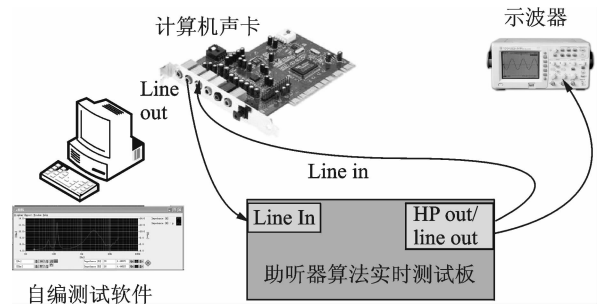


图 4 响度补偿与降噪测试系统组成

Fig. 4 System composition of loudness compensation and noise test

3.3 噪声抑制实验

对本文提出的降噪算法进行测试,并与传统维纳滤波降噪法进行对比。实验在 NOISEX-92 噪声库中的 White noise, Tank noise, Speech babble 和 HF channel noise 4 种噪声情况下,在输入信噪比为 0 dB, 5 dB 和 10 dB 下分别测试。实验中语音信号的采样率为 16 kHz, 帧长为 256 点, 帧移 50%。降噪后的语音从输出信噪比、MOS 得分、语谱图等方面来评价增强效果。MOS 得分采用 40 人在静音室进行试听, 每人试听 3 种信噪比下的输出语音样本, 每种信噪比下的样本试听 50 句。MOS 得分采用 5 级评分标准, 得分越高表示语音质量越好。输出信噪比和 MOS 得分结果见表 3。

表 3 本文方法与传统维纳滤波测试结果

Tab. 3 Test results of the proposed method and the conventional Wiener filtering

噪声源	输入信噪比/dB	传统维纳滤波法		本文方法	
		输出信噪比/dB	MOS 得分	输出信噪比/dB	MOS 得分
White noise	0	6.9	3.3	9.3	3.7
	5	10.1	3.7	14.8	4.2
	10	14.9	4.4	19.7	4.8
Tank noise	0	6.7	3.3	9.2	3.7
	5	10.1	3.9	14.5	4.1
	10	14.8	4.3	19.2	4.7
HF channel noise	0	6.6	2.8	9.5	3.4
	5	11.2	3.4	13.1	3.8
	10	13.5	4.1	19.4	4.5
Speech babble	0	5.3	2.7	8.2	3.2
	5	9.4	3.2	11.0	3.6
	10	13.0	4.0	14.0	4.1

在输入为含 Tank noise 噪声且信噪比为 0 dB 情况下, 维纳滤波法输出信号语谱图与本文方法语谱图如图 5 所示。从图中可以看出, 两种降噪算法都取得了比较好的效果, 但本文方法输出信号语谱图背景更干净, 声纹更清晰, 说明输出语音质量更高。经真耳试听, 背景噪声明显变小, 语音更清晰。同样条件下, White noise, HF channel noise 和 Speech babble 类型噪声也取得了良好的输出语音质量。

4 结束语

本文针对数字助听器应用, 提出了一种基于声音场景分类的噪声抑制算法。算法使用调制滤波法对纯语音、纯噪声和含噪声语音 3 种场景进行分类, 并根据分类结果调整噪声抑制算法参数集, 得到不同的抑制系数。本文方法在助听器测试系统中取得了良好的实验效果, 场景分类正确率在 95% 以上。在不同的噪声类型情况下, 经过本文算法处理的输出语音信号取得了良好的信噪比和 MOS 评分的提升。本文算法不仅具有良好的信噪比提升效果, 而且实时性好, 可以应用于需要实时语音增强的其他声学系统, 如会议系统、语音通信等。

表 2 不同声音场景下的分类正确率

Tab. 2 Accuracy of classification with different sound scenes

噪声源	纯语音	SNR=0 dB	纯噪声
White noise	100	99.1	100.0
Tank noise	100	98.4	100.0
Speech babble	100	95.7	98.2
HF channel noise	100	97.3	100.0

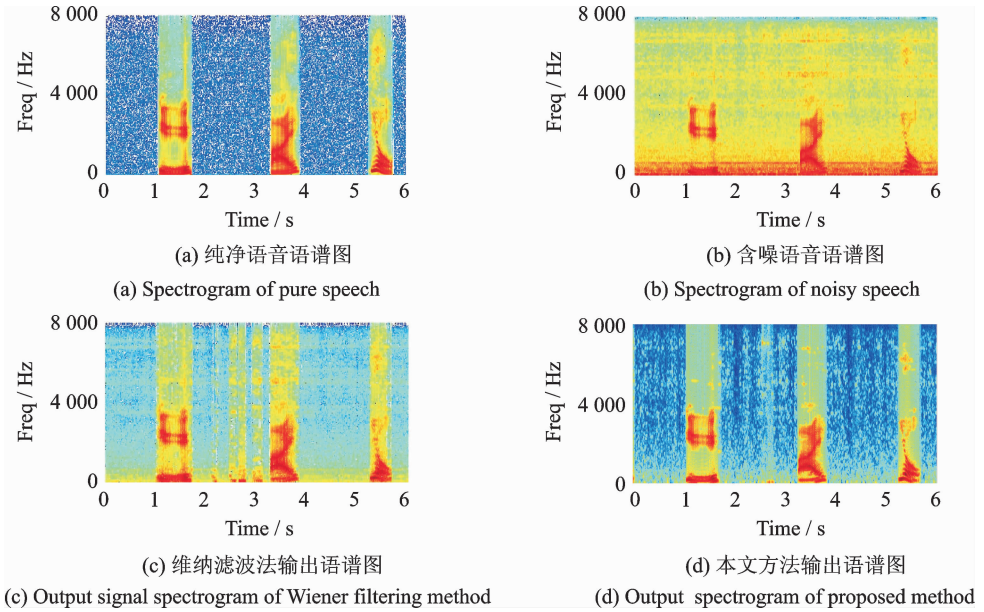


图5 Tank noise 噪声、信噪比为 0 dB 时维纳滤波法与本文方法输出信号语谱图

Fig. 5 Output signal spectrogram of Tank noise with 0 dB SNR Wiener filtering method and the proposed method

参考文献:

- [1] Scharrer R, Vorlander M. Sound field classification in small microphone arrays using spatial coherences[J]. *Audio, Speech, and Language Processing, IEEE Transactions on*, 2013, 21(9): 1891-1899.
- [2] Gil-Pita R, Ayllon D, Ranilla J, et al. A computationally efficient sound environment classifier for hearing aids[J]. *Biomedical Engineering, IEEE Transactions on*, 2015, 62(10): 2358-2368.
- [3] Büchler M C. Algorithms for sound classification in hearing instruments[D]. Zurich: Swiss Federal Institute of Technology, 2002.
- [4] Alexandre E, Cuadra L, Rosa M, et al. Feature selection for sound classification in hearing aids through restricted search driven by genetic algorithms[J]. *Audio, Speech, and Language Processing, IEEE Transactions on*, 2007, 15(8): 2249-2256.
- [5] Maas A L, Le Q V, Oneil T M, et al. Recurrent neural networks for noise reduction in robust ASR[J]. *INTERSPEECH, 2012, Citeseer*: 22-25.
- [6] 蔡明琦, 凌震华, 戴礼荣. 基于隐马尔可夫模型的中文发音动作参数预测方法[J]. *数据采集与处理*, 2014, 29(2): 204-210. Cai Mingqi, Ling Zhenhua, Dai Lirong. Hidden-Markov-model-based articulatory movement prediction for Chinese[J]. *Journal of Data Acquisition and Processing*, 2014, 29(2): 204-210.
- [7] Sen I, Saraclar M, Kahya Y P. A comparison of SVM and GMM-based classifier configurations for diagnostic classification of pulmonary sounds[J]. *Biomedical Engineering, IEEE Transactions on*, 2015, 62(7): 1768-1776.
- [8] Greenberg S, Kingsbury B E. The modulation spectrogram; In pursuit of an invariant representation of speech[J]. *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 1997, 3(1): 1647.
- [9] Dubey R K, Kumar A. Comparison of subjective and objective speech quality assessment for different degradation/noise conditions[C]// *Signal Processing and Communication (ICSC), 2015 International Conference on*. [S. l.]: IEEE, 2015: 261-266.

作者简介:



汪家冬(1990-),男,硕士研究生,研究方向:数字语音信号处理, E-mail: 1304943689@qq.com.



邹采荣(1963-),男,教授,博士生导师,研究方向:声信号与语音信号处理。



蒋本聪(1992-),男,硕士研究生,研究方向:数字语音信号处理。

王青云(1972-),女,博士,研究方向:语音信号处理。

