DOI: 10.13382/j.jemi.2017.11.024

掩蔽法减少谱减法去混响中的音乐噪声*

吴礼福 申 浩

(江苏省大气环境与装备技术协同创新中心 南京信息工程大学电子与信息工程学院 南京 210044)

摘 要:针对谱减法去混响中出现的音乐噪声问题,研究了一种基于频率归一化的掩蔽后处理方法。首先利用谱减法对混响信号去除混响,其次对混响信号与去混响信号进行频率归一化,然后在频域将去混响信号和混响信号对应相减,据此计算出掩蔽因子,最后运用掩蔽因子对去混响信号进行处理来减少音乐噪声。仿真实验中分别利用语谱图和语音质量感知评价进行了评估,语谱图表明后处理语音信号中音乐噪声明显减少,语音质量感知评价表明采用后处理方法提高了语音信号 0.55 的评分。

关键词: 谱减法;混响;音乐噪声;掩蔽

中图分类号: TN912.35 **文献标识码**: A **国家标准学科分类代码**: 510.4040

Reducing musical noise in dereverberation of spectral subtraction based on masking method

Wu Lifu Shen Hao

(Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: Aiming at the problem of musical noise in dereverberation of spectral subtraction, a post processing masking method based on frequency normalization is researched. Firstly, the reverberation signal is processed by spectral subtraction to remove the reverberation. Secondly, the reverberation signal and the dereverberation signal are normalized in frequency. Then the dereverberation signal and the reverberation signal are subtracted correspondingly in frequency domain to get the mask factor. Finally, the mask factor is used to deal with the dereverberation signal in order to reduce the musical noise. In the simulation experiment, the spectrogram and the perceptual evaluation of speech quality are used to evaluate the masking method. The spectrogram verifying the musical noise is reduced obviously in the post processed speech signal, and the perceptual evaluation of speech quality shows that the post processing method improves the score of the speech signal 0.55.

Keywords: spectral subtraction; reverberation; musical noise; mask

0 引 言

声音在传播的时候一般都会受到加性噪声和混响的影响,这不仅会影响信号中信息的提取也会影响对信号的处理,例如在语音识别中对说话人特征[1]和语音特征参数[2]的提取,当然这些影响也不完全是负面的,例如根据噪声的情况可以对声源的情况进行判断[3]。语音增强[4]就是从原来的含有噪声或混响的语音中尽可能地还原出干净的原始语音。目前,语音增强的算法主要包括谱减法[5]、维纳滤波法[6]、最小均方误差估计法[7](minimum mean squareerror estimation, MMSE)等,其中去混响效果较好的算法之一就是谱减法。谱减法具有约束条件少、运算简单的优点,但是经过谱减法处理之后的语

收稿日期:2017-04 Received Date: 2017-04

* 基金项目:国家自然科学基金(11504176)资助项目

音中会有残留的“音乐噪声”。文献[8]中提到,由于谱减法中用当前帧噪声的均值来代替当前帧的噪声,所以在当前帧噪声的随机谱峰处,经谱减后就会有残余噪声,这些语音增强产生的残余噪声在时域便会形成没有规律的突发音调,“音乐噪声”就是各帧上具有音乐特性的残余噪声的群体结果。

针对谱减法中的“音乐噪声”已经提出了很多的算法或方法。例如,通过在频域内平滑增益因子的后滤波方法[9]、使用 burg 谱先验信噪比估计来消除“音乐噪声”[10]、基于噪声谱结构特性的谱减法[11]等等。这些算法及方法都在一定程度上减少了“音乐噪声”。

由“音乐噪声”产生的原因可以得到:“音乐噪声”在原来的混响语音中是不存在的,而是经过算法处理之后才产生。由此,本文研究了一种后处理算法来减少“音乐噪声”:找到去混响信号中能量高的部分,然后相应减少这些部分的能量以达到减少“音乐噪声”的目的。

1　谱减法基本原理

基本谱减法[12]的原理就是假设噪声与带噪语音互不相关,从带噪语音中减去噪声得到干净语音信号。$y(n)$ 是输入的带噪语音信号,$x(n)$ 是干净语音信号,$d(n)$ 是噪声信号,则:

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

式(1)两边进行傅里叶变换得到:

$$Y(\omega) = X(\omega) + D(\omega) \tag{2}$$

式中:$Y(\omega)$、$X(\omega)$、$D(\omega)$ 分别为 $y(n)$、$y(n)$、$d(n)$ 的傅里叶变换。$Y(\omega)$ 用极坐标则:

$$Y(\omega) = |Y(\omega)|e^{j\varphi_y(\omega)} \tag{3}$$

式中:$|Y(\omega)|$ 是带噪语音信号的幅度谱,$\varphi_y(\omega)$ 是带噪语音信号的相位谱。相应的噪声信号也可以用极坐标表示:

$$D(\omega) = |D(\omega)|e^{j\varphi_d(\omega)} \tag{4}$$

由于人耳对相位不敏感[13],可以借用原带噪语音信号的相位谱。

由此可以得到干净语音信号的估计值 $\hat{X}(\omega)$ 为:

$$\hat{X}(\omega) = \lfloor |Y(\omega)| - |\hat{D}(\omega)| \rfloor e^{j\varphi_y(\omega)} \tag{5}$$

式中:$|\hat{D}(\omega)|$ 为噪声信号的的幅度谱估计。用在变量上加 $\wedge$ 的方式来表示估计量。

假定 $d(n)$ 是均值为 0 的高斯分布,则:

$$|Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2 + 2\mathrm{Re}\{X(\omega)D^*(\omega)\} \tag{6}$$

由式(6)得到:

$$|\hat{X}(\omega)|^2 = |Y(\omega)|^2 - |\hat{D}(\omega)|^2 \tag{7}$$

则干净信号的估计值为:

$$\hat{X}(\omega) = (|Y(\omega)|^2 - |\hat{D}(\omega)|^2)^{1/2} \tag{8}$$

基本谱减法流程如图 1 所示。

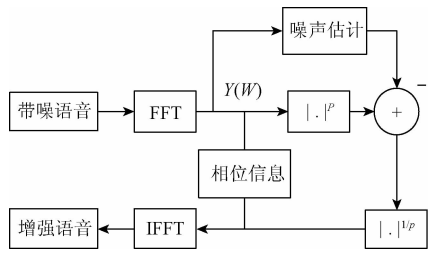


图 1　基本谱减法流程

Fig. 1　Flow chart of basic spectral subtraction

2　基于掩蔽的后处理方法

“音乐噪声”是经过算法处理之后才产生,原来的混响语音中是没有的,也就是说处理之后的信号要比原来的信号能量要高,所以可以找到去混响信号中能量高的部分,然后相应减少这些部分的能量就能达到减少“音乐噪声”的目的。

$$Z(\omega) = X(\omega) - Y(\omega) \tag{9}$$

式中:$X(\omega)$ 为去混响信号幅度的傅里叶变换,$Y(\omega)$ 为混响信号幅度的傅里叶变换。由于人耳对语音信号相位变化不敏感,所以在对信号进行处理时只处理幅度,相位用原信号相位代替即可。

但是,信号经过算法处理之后频谱会发生偏移,也就是说时间与频率不再对应。这是因为声音在传播时高频部分在遇到边界(比如:墙)的时候会被吸收而不是反射,导致低频部分更容易传播,进行算法处理时通常会弥补这种损失,这就导致了频谱的偏移。而掩蔽后处理算法的关键就是让处理之前和处理之后的信号进行对比,这种频谱的偏移对算法的影响是致命的。为了解决这个问题,需要对信号进行归一化处理,把线性幅度谱变成对数幅度谱后减去对应列对数谱的均值,然后再做对数的逆运算返回线性幅度谱。

$$Z_N(\omega) = X_N(\omega) - Y_N(\omega) \tag{10}$$

式中:$X_N(\omega)$ 和 $Y_N(\omega)$ 分别是经过一化处理的去混响信号和混响信号的幅度谱。因为只需要得到能量增加的部分所以有:

$$diff = \begin{cases} 0, & Z_N \leqslant 0 \\ Z_N, & Z_N \geqslant 0 \end{cases} \tag{11}$$

这样就得到了能量有差异的部分。

$$mask = \frac{1}{1 + \gamma \cdot diff} \tag{12}$$

式(12)得到 $mask$ 的值,γ 为加权系数,γ 的值根据情况决定使其达到最好的输出结果。

$$X_m(\omega) = X_N(\omega) \cdot mask \tag{13}$$

式中：$X_m(\omega)$ 即经过掩蔽处理之后的去混响信号。

掩蔽后处理算法流程如图 2 所示。这种后处理算法只会对经过谱减法处理之后的信号中能量多出来的部分进行掩蔽处理，避免了对信号其他部分的减少，从而在减少音乐噪声的同时一定程度上避免了信号的失真。

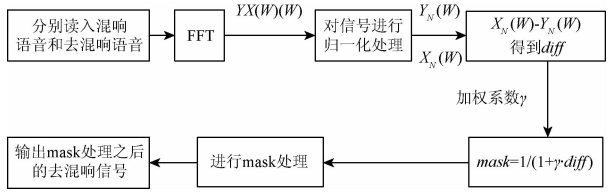


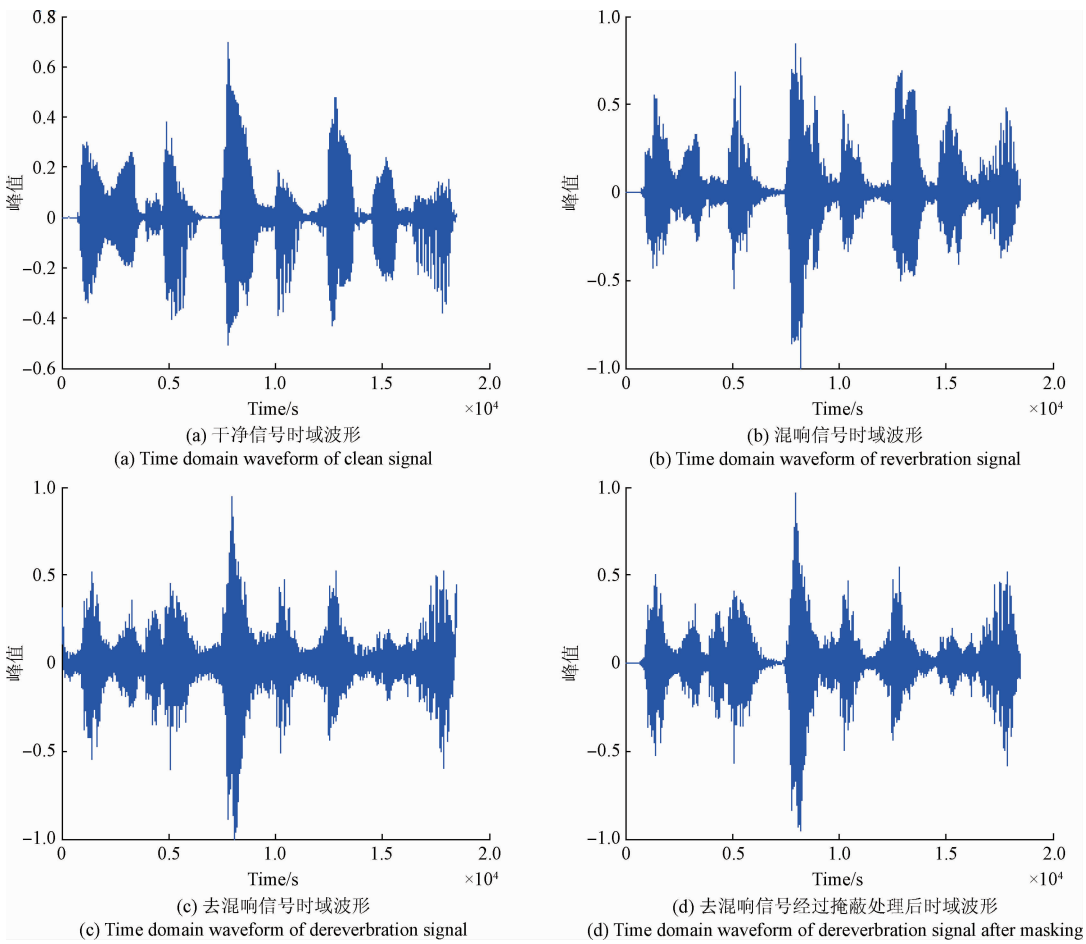
图 2　mask 后处理算法流程

Fig. 2　Flow chart of post processing algorithm based on mask

3　实验及分析

在实验仿真中，将一段采样率为 8 000 Hz 的干净语音信号与实际测得的房间的脉冲响应做卷积得到混响信号，混响信号经过谱减法处理得到去混响信号。进行掩蔽处理时，先把混响信号与去混响信号分帧之后变换到频域并把幅度和相位进行分离，接着在归一化处理后让去混响信号减去混响信号找到经过谱减法之后能量多出来的地方，最后利用掩蔽对去混响信号进行处理并输出处理之后的去混响信号。

图 3 所示为干净信号、混响信号、去混响信号以及经过掩蔽处理的去混响信号四者的时域波形。图 3(c)可以明显看出去混响信号比混响信号多出了一部分，多出的这部分就是要去掉的“音乐噪声”。图 3(d)可以看出经过掩蔽处理之后的去混响信号比原来的去混响信号就要好很多，看起来更加“干净”，表明经过掩蔽处理之后“音乐噪声”有了明显的减少。



(a) 干净信号时域波形
(a) Time domain waveform of clean signal

(b) 混响信号时域波形
(b) Time domain waveform of reverbration signal

(c) 去混响信号时域波形
(c) Time domain waveform of dereverbration signal

(d) 去混响信号经过掩蔽处理后时域波形
(d) Time domain waveform of dereverbration signal after masking

图 3　信号的时域波形

Fig. 3　Time domain waveform of signal

语谱图[14]是把语音信号用频率依次相连的窄带滤波器进行处理，每个滤波器的结果经整流均方后根据频率依次由低到高排列，用颜色的深浅来表示信号能量的强弱。如果某个滤波器输出信号的能量强，则颜色较深；反之，如果某个滤波器输出信号的能量弱，则颜色较浅。其横轴是时间，纵轴是频率，图上的颜色条纹表示各个时刻的语音在各个频率的能量情况。可以根据语谱图的特性来对语音信号进行评价。

图 4 所示为干净信号、混响信号、去混响信号以及经过掩蔽处理的去混响信号四者的语谱图。图 4(c)可以看出来经过谱减法处理之后，去混响信号在能量上多了很多，这些正是经过谱减后产生的“音乐噪声”。图 4(d)可以看出经过掩蔽处理后，能量少了很多，方框内的区域尤为明显，说明“音乐噪声”有了明显的减少。

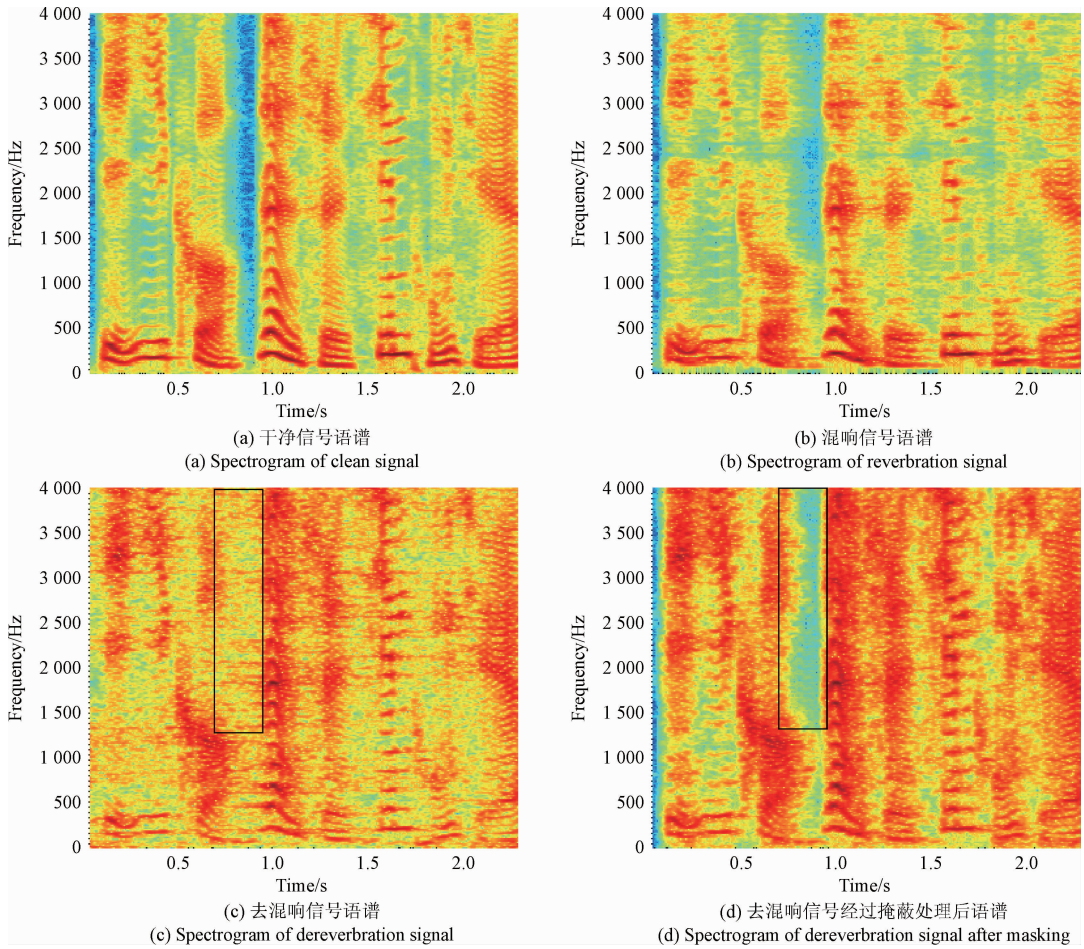


图 4　信号的语谱

Fig. 4　Spectrogram of signal

为了进一步评价掩蔽处理的效果，采用语音质量感知评价[15](perceptual evaluation of speech quality，PESQ)对掩蔽处理的效果进行评估，使效果量化，比较更加明显。语音质量感知评价 PESQ，已被 ITU-T(国际电信联盟电信标准化部)的相关资料证明，其对编码失真、传输丢失、多次变换编码、环境噪声、时间扭曲能够精确的给出令人满意的预测值。PESQ 得分的高低可以用来评价信号的好坏，PESQ 的得分通常在 1.0 ~ 4.5。但是当失真情况严重时，得分可能会低于 1.0。

图 5 所示为去混响信号和掩蔽处理之后的去混响信号两者 PESQ 得分情况。图 5 表明去混响信号在经过掩蔽处理之后，PESQ 评分有明显上升，由原来的 2.47 提高到了 3.02。

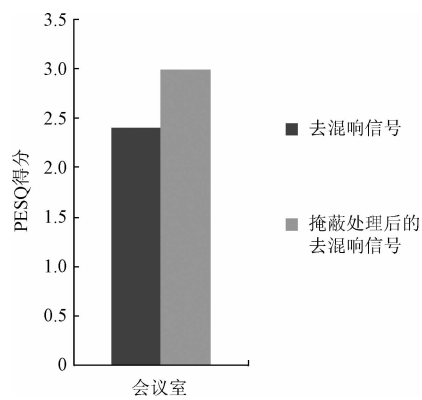


图 5　信号的 PESQ 得分

Fig. 5　PESQ scores of signal

4 结　论

本文在谱减法处理结果的基础上,提出一种基于频率归一化的掩蔽后处理算法来减少音乐噪声。在仿真试验中,通过语谱图和语音质量感知评价两种方法,证明了音乐噪声有明显减少。

参考文献

[1] 酆勇,熊庆宇,石为人,等. 一种基于受限玻尔兹曼机的说话人特征提取算法[J]. 仪器仪表学报,2016,37(2):256-262.
FENG Y, XIONG Q Y, SHI W R, et al. S-peaker feature extraction algorithm based on restricted Boltzmann[J]. Chinese Journal of Scientific Instrument, 2016,37(2):256-262.

[2] 张贺,沈天飞,滕秋霞. 小词汇孤立词语音识别系统多种特征组合参数的选择方法研究[J]. 电子测量技术,2015,38(3):48-53.
ZHANG H, SHEN T F, TENG Q X. Research about the selection methods on various c-ombination characteristic parameters in speechrecognition system based in small vocabularyand isolated words[J]. Electronic Measurement Technology,2015,38(3):48-53.

[3] 郭小青,李东新,田正宏,等. 基于噪声信号的振捣棒工作状态判定方法[J]. 国外电子测量技术,2016,35(7):15-18.
GUO X Q, LI D X,TIAN ZH H, et al. Vibrator rod state determining method based on the noise signal [J], Foreign Electronic Measurement Technology, 2016, 35(7):48-53.

[4] BENESTY J, MAKINO S, CHEN J. Speech Enhancement[M]. New York:Speinger,2005:25-46.

[5] KINOSHITA K, NAKATANI T, MIYOSHI M. Spectral subtraction steered by multistep forward linear prediction for signal channel speech dereverbration [C]. IEEE International Conference on Acoustics Speech and Signal Processing Preceedings,2006: 817-820.

[6] CHEN J D, BENESTY J, HUANG Y T, et al. New insights into the noise reductionwiener filter[J]. IEEE Transactions on Audio Speech and Language Processing, 2006,14(4):1218-1234.

[7] MARTIN R. Speech enhancement based on minimum mean-square error estimation and s-upergaussian priors[J]. IEEE Transactions on Speech and Audio Precessing,2005,13(5):845-856.

[8] 张仁志,崔慧娟. 谱相减法语音增强技术中“音乐噪声”的抑制[J]. 电声技术,2005(5):35-38.
ZHANG R ZH, CUI H J. Suppression of mus-ical noise in speech enhancement using spectral subtraction [J]. Audio Engineering,2005(5):35-38.

[9] THOMAS E, PETER V. Efficient musical noise suppression for speech enhancement system [C]. Acoutics, Speech and Signal Processing, IEEE, 2009: 4409-4412.

[10] 徐耀华,郭英,范海宁. 语音增强:使用 burg 谱先验信噪比估计消除“噪声语音”[J]. 信号处理,2009,25(1):141-146.
XU Y H, GUO Y, FAN H N. Speech ehancement using burg-based on a priori SNR es-timator to eliminate musical noise phenomenon [J]. Signal Prrocessing, 2009,25(1):141-146.

[11] 郑程诗,胡笑浒,周翊,等. 基于噪声谱结构特性的谱减法[J]. 声学学报,2010,35(2):215-222.
ZHENG C SH, HU X H, ZHOU Y, et al. Spectral subtraction based on the structure of noise power spectral density[J]. ACTA Acustica,2010,35(2):215-222.

[12] PHILIPOS C L. Speech Enhancement: Theory and Practice[M]. Bocaraton:CRC Press,2013:97-101.

[13] LANGARANI M S E, VEISI H, SANETI H. The effect of phase information in speech e-nhancement and speech recognition[C]. IEEE 11th International Conference on Information Science, Signal Precessing and their Applications,2012:1446-1447.

[14] 李富强,万红,黄俊杰. 基于 MATLAB 的语谱图显示与分析[J]. 微计算机信息,2005,21(10-3):172-174.
LI F Q, WAN H, HUANG J J. The display and analysis of spectrogram based on MATLAB [J]. Contral & Automation,2005,21(10-3):172-174.

[15] HU Y, LOIZOU P C. Evaluation of objectivequality measures for speech enhancement [J]. IEEE Transactions on Audio Speech & Lang-uage Processing, 2008,16(1):229-238.

作者简介

吴礼福,2005 年于中国科学技术大学获得硕士学位,2012 年于南京大学获得博士学位,现为南京信息工程大学副教授,主要研究方向为音频信号处理。
E-mail:wulifu@ nuist. edu. cn

Wu Lifu received M. Sc. from University of Science and Technology of China in 2005 and Ph. D. from Nanjing University in 2012, respectively. Now he is an associate professor in Nanjing University of Information Science and Technology. His main research interest includes audio signal processing.

申浩,2015 年于华北水利水电大学获得学士学位,现为南京信息工程大学硕士研究生,主要研究方向为音乐噪声处理。
E-mail:13140023858@163. com

Shen Hao received B. Sc. from North China University of Water Resources and Electric Power in 2015. Now he is a M. Sc. candidate in Nanjing University of Information Science and Technology. His main research interest includes musical noise processing.