

基于参数估计和感知提升的语音增强降噪算法

王晶* 尹栋 蒋涉权 杨立东 谢湘
(北京理工大学信息与电子学院 北京 100081)

摘要: 为了提高单通道语音增强降噪算法的整体质量,该文从噪声消除和语音感知两个角度出发对传统语音增强算法进行改进,通过引入多种处理手段来达到最佳优化效果。首先在参数估计方面,把基于弱语音出现的平滑算法加入到基于固定先验信噪比的软判决方法中来解决噪声谱过估计问题,并根据语音帧存在概率动态调整平滑因子,从而提高先验信噪比的跟踪效果。其次在语音质量感知提升方面,采用谐波恢复的方法重建语音段的高频谐波分量,并采用相位补偿和增益平滑的方法消除静默段和语音段的音乐噪声。实验结果表明,相比传统算法,该文算法通过引入参数估计改进模块和感知质量提升模块,在消噪效果和语音质量两方面均得到了较大的提高,并适用于多类噪声环境和信噪比条件。

关键词: 语音增强; 噪声功率密度估计; 先验信噪比; 谐波恢复; 相位补偿

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2016)01-0174-06

DOI: 10.11999/JEIT150504

Speech Enhancement Denoising Algorithm Based on Parameters Estimation and Perception Improvement

WANG Jing YIN Dong JIANG Shequan YANG Lidong XIE Xiang
(School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China)

Abstract: In order to enhance the whole quality of single channel speech enhancement denoising algorithm, both noise reducing and speech perception are considered to improve the traditional speech enhancement algorithm and many kinds of processing methods are taken to achieve the best optimization effect. Firstly, in the view of parameters estimation, spectrum smoothing algorithm based on weak speech presence is added to the soft decision method based on fixed prior signal-to-noise ratio in order to solve the problem of noise spectrum overestimation. Moreover, the smoothing parameter is dynamically controlled by the speech presence probability in order to enhance the tracing effect of prior signal-to-noise ratio. Secondly, in the view of the speech perception improvement, the harmonic reconstruction method is used to reconstruct the harmonic components in high frequencies of speech section. Phase compensation method and gain smoothing method are also employed to remove the annoying musical noise in speech and silence segment. The experimental results show that compared with the traditional algorithm, the proposed algorithm obtains good performance in both denoising effect and speech quality by introducing parameter estimation improvement module and perceived quality improvement module, and it is suitable for many kinds of noise environment and signal-to-noise ratio conditions.

Key words: Speech enhancement; Noise power density estimation; Prior signal-to-noise ratio; Harmonic reconstruction; Phase compensation

1 引言

单通道语音增强降噪算法因具有算法复杂性小、便于实现的特点在带噪语音信号处理领域被广

泛采用,以期获得较好的噪声抑制效果,同时保持较好的语音质量。单通道的增强算法通常在频域进行,一般涉及到噪声谱和先验信噪比的估计。两个参数估计的准确性直接影响到增强后的语音质量。噪声谱估计的最简单方法是用端点检测算法(Voice Activity Detection, VAD)检测语音信号的有声无声部分,然后在无声部分对噪声谱进行估计和更新,在平稳噪声环境下此方法能得到很好效果,但在非平稳环境下不能很好地工作。文献[1]提出了最小统计算法(Minimum Statistics, MS),文献[2]提出

收稿日期: 2015-04-30; 改回日期: 2015-09-06; 网络出版: 2015-11-18

*通信作者: 王晶 wangjing@bit.edu.cn

基金项目: 国家自然科学基金(61473041, 11461141004, 61571044), 北京市高等学校青年英才计划(YETP1202)

Foundation Items: The National Natural Science Foundation of China (61473041, 11461141004, 61571044), The Beijing Higher Education Young Elite Teacher Project (YETP1202)

了最小控制递归平均 (Minimum Controlled Recursive Averaging, MCRA) 算法, 还有很多在此基础上的改进算法如文献[3]的 IMCRA 算法。上述算法都能对每帧的噪声进行持续更新, 在跟踪效果上与 VAD 算法相比, 都得到了很好的提升, 但在噪声水平上升时, 上述算法都不能进行很好的跟踪。文献[4]的直接判决法 (Decision-Directed, DD) 被广泛用于先验信噪比的估计, 但 DD 算法存在一帧的延时和平滑因子不能与噪声水平自适应的问题, 文献 [5] 提出了两步消噪算法 (Two Step Noise Reduction, TSNR), 解决了帧延时问题。传统的增强算法往往把高频点的谐波成分当作噪声滤除, 降低了增强后的语音质量, 由于参数的估计准确度不高, 增强后的语音往往存在音乐噪声, 影响了语音的感知质量。

本文采用文献[6,7]提出的软判决方法, 根据语音时频点出现的概率来决定噪声谱的更新, 并加入基于弱语音出现的谱平滑算法, 保护强语音出现后的弱语音; 为提高 DD 算法的跟踪效果, 本文根据语音帧出现概率动态调整平滑因子; 最后, 为提升增强后语音的整体感知质量, 引入了高频谐波恢复、增益平滑和相位补偿 3 个模块。

2 算法总体描述

对输入的带噪语音信号进行分帧和傅里叶变换得到频谱数据, 在频域先对噪声谱和先验信噪比两个参数进行估计, 根据后验信噪比和先验信噪比计算得到增益函数, 利用得到的增益函数对带噪语音进行一级增强滤波处理(即预处理), 增益函数采用文献[4]中基于对数的最小均方误差(log-MMSE)算法, 根据预处理的结果对语音进行感知质量提升, 感知质量提升模块包括谐波恢复算法、低通滤波和相位补偿算法。进行感知质量提升首先需要对语音进行活动检测, 经过预处理后, 仅依靠每帧的信噪比就可以对语音活动做出较为准确的判断, 若为语音段则采用谐波恢复的方法重新计算增益函数, 增

益函数同样采用 log-MMSE 算法, 若为静默段则进行低通滤波, 滤除高频的尖峰来减小音乐噪声。为进一步消除残留噪声, 采用相位补偿的算法进一步消除整段语音的残留噪声。最后, 对增强处理后的频谱数据进行反傅里叶变换和重叠相加得到增强后的时域信号, 具体算法流程图如图 1 所示。

3 算法改进模块

3.1 参数估计改进模块

3.1.1 噪声谱估计 带噪语音信号 $x(n) = s(n) + d(n)$, 假定 $s(n)$ 和 $d(n)$ 统计独立且频谱数据 (FFT 的实部和虚部) 服从高斯分布, 分别用 $X(k, m) = |X(k, m)| e^{j\theta_{xk}}$, $D(k, m)$ 和 $S(k, m) = |S(k, m)| e^{j\theta_{sk}}$ 表示 $x(n)$, $s(n)$ 和 $d(n)$ 的第 m 帧第 k 个频率点的频谱。功率谱的导出基于式(1)两个假设。

$$\left. \begin{aligned} H_0(k, m): X(k, m) &= D(k, m) \\ H_1(k, m): X(k, m) &= S(k, m) + D(k, m) \end{aligned} \right\} \quad (1)$$

$H_0(k, m)$, $H_1(k, m)$ 分别表示 m 帧的第 k 个频率点语音存在和不存在, 基于式(1)假设, 功率谱的更新可以写成式 (2) 的形式, 其中 $P(H_1 | X)$, $P(H_0 | X)$, $\hat{\lambda}_d(k, m) = E[|D(k, m)|^2]$ 分别表示频率点语音存在, 不存在的概率和估计得到的噪声功率谱。

$$\hat{\lambda}_d(k, m) = P(H_1 | X) \hat{\lambda}_d(k, m - 1) + P(H_0 | X) |X(k, m)|^2 \quad (2)$$

两种假设下语音信号的概率密度分别如式 (3) 所示, 其中 $\lambda_s = E[|X(k, m)|^2]$ 。

$$\left. \begin{aligned} P(X(k, m) | H_0) &= \frac{1}{\pi \lambda_d(k, m)} \exp\left\{-\frac{|X(k, m)|^2}{\lambda_d(k, m)}\right\} \\ P(X(k, m) | H_1) &= \frac{1}{\lambda_d(k, m) + \lambda_s(k, m)} \cdot \exp\left\{-\frac{|X(k, m)|^2}{\lambda_s(k, m) + \lambda_d(k, m)}\right\} \end{aligned} \right\} \quad (3)$$

根据式(3)计算得到语音存在概率为

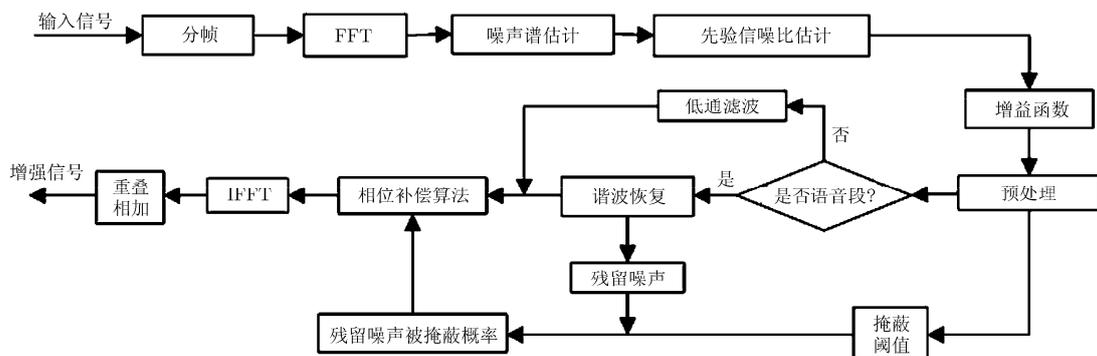


图 1 本文算法流程图

$$P(H_1 | X(k, m)) = \left\{ 1 + \frac{P(H_0)}{1 - P(H_0)} (1 + \hat{\xi}(k, m)) \exp(-V(k, m)) \right\}^{-1} \quad (4)$$

$$V(k, m) = \frac{\hat{\gamma}(k, m) \hat{\xi}(k, m)}{1 + \hat{\xi}(k, m)} \quad (5)$$

其中, $\hat{\gamma}(k, m)$, $\hat{\xi}(k, m)$ 分别表示后验信噪比估计和先验信噪比估计。当处于强噪声环境时, 分析式(3)知, 两个假设下的概率分布基本相同, 难以对语音存在概率做出判断, 文献[7]在假定语音先验不存在概率 $P(H_0) = P(H_1) = 0.5$ 的情况下, 取使 $P_e = P(H_1)P_m + P(H_0)P_f$ 取得最小值的固定 $\hat{\xi}$ 值来反应语音的存在与否, 仅根据后验信噪比 $\hat{\gamma}$ 的值计算语音存在概率, 文献[6,7]中对 $\hat{\xi}_{\text{opt}}$ 取固定值为 15 dB。

$$P(H_1 | X(k, m)) = \left\{ 1 + \left(1 + \hat{\xi}_{\text{opt}} \exp \left(-\frac{\hat{\xi}_{\text{opt}} \hat{\gamma}(k, m)}{1 + \hat{\xi}_{\text{opt}}} \right) \right) \right\}^{-1} \quad (6)$$

后验信噪比 $\hat{\gamma}$ 的计算, 在实际中根据式(7)计算。

$$\hat{\gamma}(k, m) = \frac{|X(k, m)|^2}{\hat{\lambda}_d(k, m-1)} \quad (7)$$

为保护强语音后的弱语音, 防止噪声的过估计, 采用文献[8]的方法, 根据弱语音的出现概率对带噪语音的功率谱进行平滑。

$$S_f(k, m) = \sum_{i=-\omega}^{i=\omega} b(i) |X(k-i, m)|^2 \quad (8)$$

$$S_{\text{pf}} = \begin{cases} a_p S_{\text{pf}}(k, m-1) + (1 - a_p) S_f(k, m), & \text{出现弱语音} \\ S_f(k, m), & \text{其它} \end{cases} \quad (9)$$

其中, a_p 和 b 分别是平滑因子和滑动窗, 根据文献[8]中方法建议的弱语音的出现概率 $P_w(k, m)$, 式(9)可以写成式(10)的平滑形式。

$$S_{\text{pf}}(k, m) = P_w(k, m) a_p S_{\text{pf}}(k, m-1) + (1 - a_p P_w(k, m)) S_f(k, m) \quad (10)$$

后验信噪比的计算改为

$$\hat{\gamma}(k, m) = \frac{S_{\text{pf}}(k, m)}{\hat{\lambda}_d(k, m-1)} \quad (11)$$

3.1.2 先验信噪比估计 在增强算法中对先验信噪比的估计广泛采用直接判决(DD)方法, 如式(12)所示。

$$\hat{\xi}_{\text{DD}}(k, m) = \beta \frac{|\hat{S}(k, m-1)|^2}{\hat{\lambda}_d(k, m)} + (1 - \beta) \max[\hat{\gamma}(k, m) - 1, 0] \quad (12)$$

DD 算法中采用了固定的平滑因子 β , 折中取

值为 0.98。 β 越大平滑越大, 先验信噪比估计方差越小, 音乐噪声越小, 但降低了先验信噪比的跟踪性能, 本文根据语音存在概率自适应调整 β 的值, 当前帧为语音时减小 β 提高跟踪速度, 为噪声帧时增大 β 减小方差。采用文献[9]中的方法, 对 m 帧定义如式(13)所示的语音存在概率。

$$\text{SP}(m) = 1 - \frac{1}{1 + \left(\alpha \left(\prod_{k=1}^N \Lambda(k, m) \right)^\theta \right)} \quad (13)$$

其中, $\alpha = 0.0625$, $\theta = 1/128$ 。 $\Lambda(k, m)$ 由后验信噪比 $\hat{\gamma}(k, m)$ 和先验信噪比 $\hat{\xi}(k, m)$ 计算得到, 实际运算时 $\hat{\xi}(k, m)$ 可由上一帧的计算结果代替。

$$\Lambda = \frac{1}{1 + \hat{\xi}(k, m)} \exp \left[\frac{\hat{\gamma}(k, m) \hat{\xi}(k, m)}{1 + \hat{\xi}(k, m)} \right] \quad (14)$$

定义新的平滑因子:

$$\beta(m) = \min(0.98, 0.98 - 0.05 \text{SP}(m)) \quad (15)$$

得到改进的 DD 算法为

$$\hat{\xi}_{\text{DD}}(k, m) = \beta(m) \frac{|\hat{S}(k, m-1)|^2}{\hat{\lambda}_d(k, m)} + (1 - \beta(m)) \max[\hat{\gamma}(k, m) - 1, 0] \quad (16)$$

DD 算法存在一帧的延时, 利用文献[5]中的 TSNR 算法解决了上述问题。

3.2 感知质量提升模块

3.2.1 谐波恢复 语音增强算法中很难对噪声谱做出精确的估计, 较多的高频谐波成分被当作噪声滤除, 影响了增强后的语音质量, 本文采用文献[10,11]中的谐波恢复方法, 对谐波进行恢复。

首先对语音进行一级增强, 然后对增强后语音进行非线性变换, 实质是对信号加一周矩形窗, 当语音信号的谐波分量与矩形窗周期相同时, 谐波分量得到加强, 因语音信号是准周期信号, 为防止频谱混叠, 进行分段平滑处理, 最后用处理后频谱计算先验信噪比。

$$\hat{s}(n) = \text{IFFT} \left(G_{\text{mmse}} |X(k, m)| \exp(j\theta_{\text{sk}}) \right) \quad (17)$$

$$\hat{s}_{\text{harmonic}}(n) = \begin{cases} \hat{s}(n), & \hat{s}(n) > 0 \\ 0, & \text{其它} \end{cases} \quad (18)$$

$$\hat{S}_{\text{harmonic}}(k, m) = \text{FFT}(\hat{s}_{\text{harmonic}}(n)) \quad (19)$$

$$\hat{S}_{\text{harmonic}}(k, m)$$

$$= \begin{cases} \hat{S}(k, m), & k \leq k_0 \\ a\hat{S}(k, m) + 2(1-a)\hat{S}_{\text{harmonic}}(k, m), & k_0 < k < k_1 \\ b\hat{S}(k, m) + 2(1-b)\hat{S}_{\text{harmonic}}(k, m), & k \geq k_1 \end{cases} \quad (20)$$

$$\hat{\xi}_{\text{HRNR}}(k, m) = \frac{\rho(k, m) |\hat{S}(k, m)|^2 + (1 - \rho(k, m)) |\hat{S}_{\text{harmo}}(k, m)|^2}{\hat{\lambda}_d(k, m)} \quad (21)$$

$$G_{\text{HRNR}}(k, m) = f(\hat{\xi}_{\text{HRNR}}(k, m), \hat{\gamma}(k, m)) \quad (22)$$

其中 a, b, k_0, k_1, ρ 按照经验值分别设置为 0.5, 0.4, 500 Hz, 2500 Hz, G_{mmse} 。

3.2.2 音乐噪声的消除 为在保证语音质量的前提下, 进一步减小增强后残留的噪声以减小音乐噪声, 用相位补偿的算法并结合听觉掩蔽对残留噪声进行消除, 经过预处理后可以对语音存在与否做出比较准确的估计, 此时可以在静默段进行增益平滑, 尽最大可能地消除音乐噪声。

(1) 增益平滑: 在静默段可以对增益函数加一平滑窗, 消除静默区的音乐噪声, 具体过程如下:

$$\zeta(m) = \frac{\sum_{k=0}^{N-1} |\hat{S}(k, m)|^2}{\sum_{k=0}^{N-1} |X(k, m)|^2} \quad (23)$$

其中 $\zeta(m)$ 作为检测静默段的参数, 当 $\zeta(m)$ 大于 0.3 时认为是语音段。根据 $\zeta(m)$ 值的大小来设定滑动窗的大小。

$$L(m) = \begin{cases} 1, & \zeta(m) \geq 0.3 \\ 2 \left\lfloor 10 \left(1 - \frac{\zeta(m)}{0.3} \right) \right\rfloor + 1, & \text{其它} \end{cases} \quad (24)$$

音乐噪声是语音增强后凸起的小尖峰造成, 尖峰在频域中代表高频成分, 对增益进行平滑消除尖峰, 可以用滤波的思想进行, 设置一低通滤波器消除高频成分, 窗函数具有低通滤波特性, 采用文献[12]定义的窗函数, 增益函数的幅度谱与窗函数的谱幅度相乘。

(2) 相位补偿: 相位补偿算法参考文献[13~15]中的方法, 利用实信号傅里叶变换后系数的共轭对称性, 对信号进行傅里叶变换后, 在保持幅值不变的情况下, 根据噪声水平的大小改变共轭向量对之间的相位。具体过程如下:

$$\sigma(k, m) = \eta(k, m) \psi(k) \sqrt{\hat{\lambda}_R(k, m)} \quad (25)$$

$$\hat{\lambda}_R(k, m) = \hat{\lambda}_d(k, m) - [1 - G_{\text{HRNR}}(k, m)]^2 |X(k, m)|^2 \quad (26)$$

$$\psi(k) = \begin{cases} 1, & 0 < k/N < 0.5 \\ -1, & 0.5 < k/N < 1 \\ 0, & \text{其它} \end{cases} \quad (27)$$

$$X_\sigma(k, m) = X(k, m) + \sigma(k, m) \quad (28)$$

$$\angle X_\sigma(k, m) = \text{ARG}(X_\sigma(k, m)) \quad (29)$$

$$|\hat{S}_\sigma(k, m)| = G_p |X(k, m)| \exp(j \angle X_\sigma(k, m)) \quad (30)$$

$$\eta(k, m) = 1.5 (1 - G_{\text{HRNR}}(k, m)) (1 - P_{\text{mask}}(k, m)) \quad (31)$$

其中, $P_{\text{mask}}(k, m)$ 参考文献[16,17]中的方法, 表示残留噪声被掩蔽的概率。

4 实验和分析

本文中的增益函数选用文献[4]中的基于对数的 log-MMSE 算法, 为测试本文组合算法的增强效果, 设置一参照算法, 增益函数同样选用 log-MMSE 算法, 噪声谱估计选用 Gerkmann 提出的原始软判决算法, 先验信噪比选用原始的 TSNR 算法, 去掉感知质量提升模块。采用文献[18,19]提到的 3 种客观指标, 分别如式(32), 式(33), 式(34)所示, 测试参照算法和本文算法在不同噪声源和不同信噪比的增强效果。

噪声源是从 NOISEX-92 噪声库中获得的, 选择了高斯白噪声(white)和办公室噪声(office), 纯净语料由 20 条语料构成, 包括 10 条男声语料和 10 条女声语料。制作采样率为 8 kHz, 信噪比为 -2 dB, 3 dB, 8 dB, 13 dB, 18 dB 的测试语料。

$$C_{\text{sig}} = 3.093 - 1.029 S_{\text{LLR}} + 0.603 S_{\text{PESQ}} - 0.009 S_{\text{WSS}} \quad (32)$$

$$C_{\text{bak}} = 1.634 + 0.4785 S_{\text{PESQ}} - 0.007 S_{\text{WSS}} + 0.063 S_{\text{segSNR}} \quad (33)$$

$$C_{\text{ovl}} = 1.594 + 0.8055 S_{\text{PESQ}} - 0.512 S_{\text{LLR}} - 0.007 S_{\text{WSS}} \quad (34)$$

其中 $S_{\text{PESQ}}, S_{\text{LLR}}, S_{\text{WSS}}$ 和 S_{segSNR} 分别代表语音质量的感知评估(PESQ)值, 对数似然比(Log Likelihood Ratio, LLR), 加权谱斜率(Weighted Spectral Slope, WSS)和分段信噪比(Segmental Signal to Noise Ratio, SegSNR)。 $C_{\text{sig}}, C_{\text{bak}}$ 和 C_{ovl} 是由前面的客观指标加权出的 5 分制度量指标(无量纲), 分别表示信号的失真程度、背景噪声的干扰程度和语音的综合质量, 值越大信号质量越好。

从表 1, 表 2 可以看出, 本文算法在两种噪声环境的不同信噪比下都取得了较好的效果, 并且在噪声抑制、语音保真以及语音的综合质量 3 种客观指标衡量下, 本文算法相比传统的 log-MMSE 语音增强算法均有明显的改进。从两表格分析, 两种算法的 C_{sig} 相差较大, 主要是因为本文算法与 log-MMSE 对比, 在 PESQ 得分相差微乎其微的情况下, 本文中采用了谐波恢复和自适应平滑因子, 降低了谱的失真。在静默段采用了低通滤波, 静默段的残留噪声很小, 在结合听觉掩蔽效应保证语音质量的前提下, 对语音段的残留噪声进行了消除, 所以本文中的噪声干扰程度与 log-MMSE 相比较小, C_{bak} 值较大。

图 2 是一段 SNR=5 dB 带白噪声语音信号经本

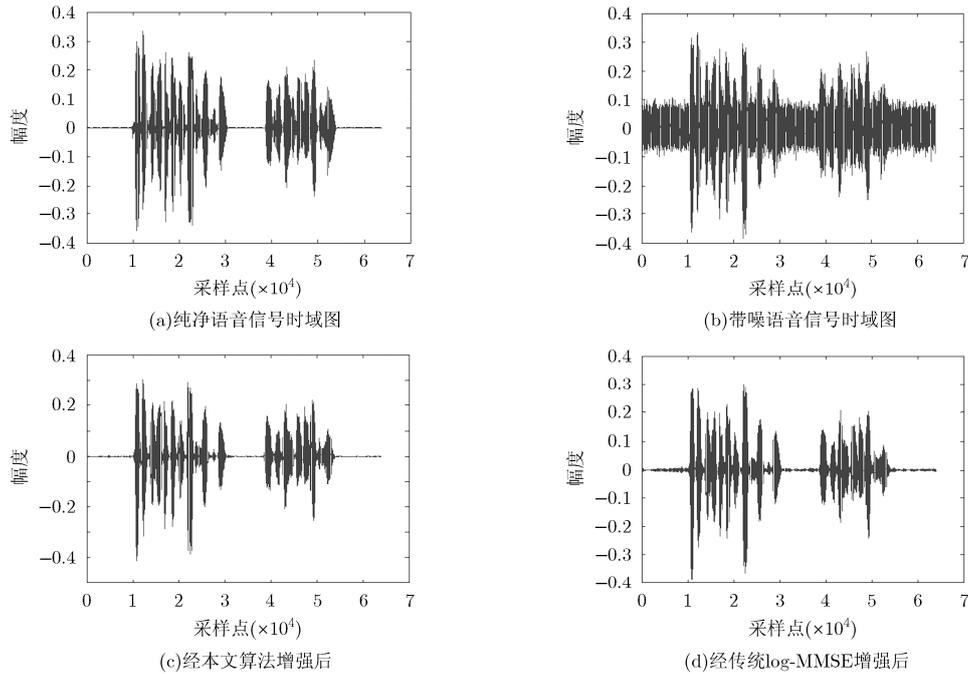


图2 经本文算法和 log-MMSE 算法增强后语音信号的时域波形图对比

表1 office 噪声环境下的客观指标

客观指标	增强算法	输入信噪比(dB)				
		-2	3	8	13	18
C_{sig}	Log-MMSE	0.979	1.594	2.182	2.673	3.052
	本文算法	1.642	2.160	2.681	3.144	3.518
C_{bak}	Log-MMSE	0.981	1.425	1.860	2.252	2.606
	本文算法	1.243	1.626	2.017	2.387	2.735
C_{ovl}	Log-MMSE	0.927	1.499	2.045	2.514	2.900
	本文算法	1.362	1.844	2.329	2.767	3.143

表2 white 噪声环境下的客观指标

客观指标	增强算法	输入信噪比(dB)				
		-2	3	8	13	18
C_{sig}	Log-MMSE	0.803	1.211	1.547	1.850	2.190
	本文算法	2.339	2.795	3.166	3.547	3.898
C_{bak}	Log-MMSE	1.749	2.056	2.369	2.691	3.019
	本文算法	1.761	2.212	2.579	2.917	3.235
C_{ovl}	Log-MMSE	1.180	1.582	1.931	2.233	2.549
	本文算法	1.861	2.359	2.744	3.110	3.467

文算法和 log-MMSE 算法处理后的时域波形图的对比, 由时域波形图分析可知, 本文在消噪效果和语音保真方面做到了很好的折中, 取得了较好的语音增强效果。

除上述的客观评价外, 还在本文算法和传统 log-MMSE 算法之间进行了主观听音效果对比评价, 从语音质量和残留噪声大小两方面考量两种算

法性能的优劣。主观评测时选用了 10 条 SNR=5 dB 带粉红噪声的语料, 5 条男声语料, 5 条女声语料, 5 名听音人, 相当于总共 50 条语料, 在语音保真度的指标下, 统计选择增强算法语音质量较好的语料条数, 然后计算其占总数的比例, 所占比例越大, 说明总体效果较好。在噪声抑制度的指标下, 以同样的方法进行, 统计增强算法残留噪声较少的语料条数占总数的比例。

从表 3 所示的主观评测结果可以看出, 在语音质量方面(语音保真度)本文算法较 Log-MMSE 稍微有所提高, 在残留噪声大小方面(噪声抑制度)改善较为明显。

表3 主观测试结果(%)

增强算法	语音保真度	噪声抑制度
Log-MMSE	43	36
本文算法	57	63

5 结束语

本文采用了频域参数估计与感知质量提升相结合的语音增强降噪算法, 通过优化噪声谱估计和先验信噪比估计模块改善参数估计性能, 通过设计谐波恢复和音乐噪声消除模块进一步提升听觉感知质量, 从而尽可能地降低了背景噪声并提高了增强后的语音质量。通过合理设计算法改进模块, 本文算法相比传统算法在噪声抑制和语音保真两方面均得

到较大的提升效果。单通道增强算法的关键是对噪声谱和先验信噪比做出尽可能准确的估计, 需要从这两方面做出进一步的改进, 音乐噪声的存在是增强算法中不可避免的, 如何保证在语音失真尽可能小的情况下消除音乐噪声, 也是今后研究的方向。

参 考 文 献

- [1] MARTIN R. Noise power spectral density estimation based on optimal smoothing and minimum statistics[J]. *IEEE Transactions on Speech Audio Processing*, 2001, 9(5): 504-512.
 - [2] COHEN I. Noise estimation by minima controlled recursive averaging for robust speech enhancement[J]. *IEEE Signal Processing Letters*, 2002, 9(1): 12-15.
 - [3] COHEN I. Noise spectrum estimation in adverse environment: improved minima controlled recursive averaging[J]. *IEEE Transactions on Speech Audio Processing*, 2003, 11(5): 466-475.
 - [4] EPHRAIM Y and MALAH D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator[J]. *IEEE Transactions on Acoustics Speech and Signal Processing*, 1985, 33(2): 443-445.
 - [5] CYRIL P, CLAUDE M, and PASCAL St. Improved signal-to-noise ratio estimation for speech enhancement[J]. *IEEE Transactions on Speech and Language Processing*, 2006, 14(6): 2098-2108.
 - [6] GERKMANN T and HENDRIKS R C. Noise power estimation based on the probability of speech presence[C]. Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New York, 2011: 145-148.
 - [7] GERKMANN T, BREITHAUPT C, and MARTIN R. Improved a posteriori speech presence probability estimation based on a likelihood ratio with fixed priors[J]. *IEEE Transactions on Audio Speech and Language Processing*, 2008, 16(5): 910-919.
 - [8] FENG Y and AN B. Noise power spectrum estimation based on weak speech protection for speech enhancement[C]. Proceedings of 12th International Conference on Signal Processing (ICSP), Hangzhou, 2014: 484-487.
 - [9] 袁文浩. 基于噪声估计的语音增强方法研究[D]. [硕士学位论文], 华东理工大学, 2013.
YUAN Wenhao. The study of speech algorithms based on noise power spectrum estimation[D]. [Master dissertation], East China University of Science and Technology, 2013.
 - [10] PLAPOUS C, MARRO C, and SCALART P. Speech enhancement using harmonic regeneration[C]. Proceedings of International Conference on Acoustics Speech and Signal Processing, Pennsylvania, 2005: 157-160.
 - [11] 颜丽君. 基于噪声估计和掩蔽效应的语音增强[D]. [硕士学位论文], 西南交通大学, 2014.
YAN Lijun. Speech enhancement based on noise estimation and masking effect[D]. [Master dissertation], Southwest Jiaotong University, 2014.
 - [12] ESCH T and VARY P. Efficient musical noise suppression for speech enhancement systems[C]. Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), Taipei, 2009: 4409-4412.
 - [13] WOJCICKI K, MILACIC M, STARK A, *et al.* Exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement[J]. *IEEE Signal Processing Letters*, 2008, 15: 461-464.
 - [14] ISLAM Md T and SHAHNAZ C. Speech enhancement based on noise-compensated phase spectral[C]. Proceedings of International Conference on Electronic Engineering and Information & Communication Technology (ICEEICT), Yichang, 2014: 1-5.
 - [15] PALIWAL K, WÓJCICKI K, and SHANNON B. The importance of phase in speech enhancement[J]. *Speech Communications*, 2011, 53(4): 465-494.
 - [16] 卜凡亮, 王为民, 戴启军. 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.
BU Fangliang, WANG Weimin, and DAI Qijun. Optimizing speech enhancement based on noise masked probability[J]. *Journal of Electronics & Information Technology*, 2005, 27(5): 753-756.
 - [17] ALAYA S, ZOGHLAMI N, and LACHIRI Z. Speech enhancement based on perceptual filter bank improvement[J]. *International Journal of Speech Technology*, 2014, 17(3): 253-258.
 - [18] HU Y and LOIZOU P. Evaluation of objective measures for speech enhancement[C]. Proceeding of Interspeech, Pittsburgh, 2006: 1447-1450.
 - [19] ZHANG Jie, ZHAO Xiaoqun, and XU Jingyun. Suitability of speech quality evaluation measures in speech enhancement[C]. 2014 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, 2014: 22-26.
- 王 晶: 女, 1980 年生, 副教授, 研究方向为语音与音频信号处理、移动通信等。
尹 栋: 男, 1990 年生, 硕士, 研究方向为语音信号处理、语音增强。
蒋涉权: 男, 1989 年生, 硕士, 研究方向为语音信号处理、阵列信号处理。
杨立东: 男, 1979 年生, 副教授, 研究方向为语音与音频信号处理、模式识别。
谢 湘: 男, 1976 年生, 副教授, 研究方向为语音信号处理、人机交互等。