音频信号截幅失真的检测与修复

刘春花[1]，付 强[2]，杨家玮[1]，颜永红[2]
(1．西安电子科技大学通信工程学院，西安 710071；2。中国科学院声学研究所，北京 100190)

摘要：音频信号截幅失真的修复是信号修复工作的重要组成部分，目的是使得信号恢复完备性，以便于对其进行进一步研究。提出了一种音频信号截幅失真的检测与修复方法。首先通过研究截幅失真特性，描述了基于音频信号峰值大小而进行的检测算法，并创造性地将基于 AR 模型以及最小均方误差的 LSAR 算法应用于修复算法中。随后给出应用该算法后得到的实验结果，以及在截幅失真测试集上得到的信噪比及 PESQ 值改善的统计，由上述实验数据可以验证，该方法对不同采样率截幅失真的修复均有一定效果，达到了音频修复的要求，提高了音频质量。

关键词：数字截幅；模拟截幅；AR 模型；LSAR
中图分类号：TB556 **文献标识码**：A **文章编号**：1000-3630(2009)-04-0498-05
DOI 编码：10.3969/j.issn1000-3630.2009.04.014

Detection and restoration of clipping distortion in audio signal

LIU Chun-hua[1], FU Qiang[2], YANG Jia-wei[1], YAN Yong-hong[2]
(1. *School of Telecommunication Engineering, Xidian University, Xi'an* 710071, *China*;
2. *Institute of Acoustics, The Chinese Academy of Science, Beijing* 100190, *China*)

Abstract: Clipping distortion in audio signal is an important part of signal restoration, whose aim is to make the audio integrative in order to do further research on it. This paper presents a method for detection and restoration of clipping distortions in audio signals. It begins with a description of the method for the detection of clipping distortions based on the audio's amplitude peak value, and creatively uses a least square AR (LSAR) repairing method based on the AR modeling and the theory of the minimum mean square error for the restoration of the clipping distortions. Subsequently, the results obtained by this method are shown, and the statistical results acquired by the method effecting on the clipping distortions test collection - the signal to noise ratio (SNR) and the value of the Perceptual evaluation of speech quality (PESQ) are presented. Obviously, the method is efficient when it processes on the clipping distortions with different sample rates, achieves the demands of audio restoration, and improves the audio quality.

Key words: digital clipping; analogue clipping; AR modeling; LSAR

1 引 言

由于录音设备或者环境噪声的影响，实际音频信号处理系统中经常有截幅发生。截幅会给原始音频信号带来刺耳的响声，如果截幅密度很大，将会严重影响信号的质量，不利于对音频信号的进一步研究。

信号插值与外推的方法在信号处理应用领域得到广泛的应用。文献[1]提出一种基于重叠正弦函数来进行信号截幅修复的方法，该方法的出发点是将截幅点视为丢失点，从信号重建的角度来考虑问题，但是该方法的局限在于其仅适用于带限信号，不具有普适性。

另外，近些年起到了用于音频信号截幅修复的算法还有解卷积算法[2]，标准解卷积算法为维纳滤波，但维纳滤波属于典型的线性求解法，而截幅失真属于非线性失真，因而应用维纳滤波后，可能会能引入新的失真，甚至给原始音频信号带来更坏的影响。

文献[2]针对该算法的不足，提出了 MAP 算法及其简化算法。虽然该方法对于截幅修复虽有一定的效果，但是算法复杂度很高，不利于实际工程的实现。

本文在总结他人算法的基础上，根据音频信号截幅失真特性，从检测到修复提出新的思路。

针对截幅发生在信号峰值附近的特点，提出了

收稿日期: 2008-07-22; **修回日期**: 2008-10-17
基金项目: 国家高技术研究发展计划 (2006AA010102, 2006AA01Z195), 国家重点基础研究发展规划项目计划 (2004CB318106), 国家自然科学基金(10574140, 60535030) 资助。
作者简介: 刘春花(1982-), 女, 河北唐山人, 硕士研究生, 研究方向为语音/音频处理中的信号分析、噪声消除。
通讯作者: 刘春花, E-mail: liuchunhua2008@gmail.com

基于峰值的检测算法；考虑现有 LSAR[3,4]算法(一种双边线性预测的方式)的优越性，将其应用到截幅失真修复领域，作为核心修复算法。与现有专业音频处理软件 Cooledit 中对截幅的修复效果对比可以看出，该检测与修复方法对音频信号截幅失真有较好的修复效果。

2　截幅失真特性及其检测原理

研究音频信号的截幅失真特性，可将其分类为数字截幅和模拟截幅[5,6]。

产生于信号处理中的音频信号，如数字转录过程，当信号幅度超出录音设备动态范围时，会在峰值点处产生截幅，此时截幅处数值严格相等，称之为数字截幅；而在实际模拟系统中，由于高信号电平不能存储，超出模拟器件线性区的信号被截幅，此时截幅处数值不完全相等，具有小抖动[5]，称之为模拟截幅。

针对上述特点，设计其检测系统如图 1 所示：

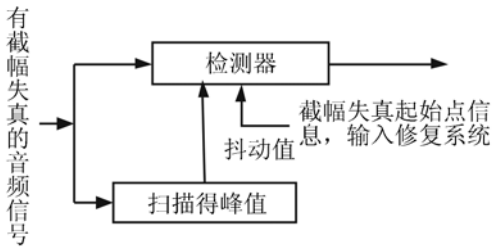


图 1　截幅失真检测原理图

Fig.1　Principle of the detection of clipping distortions

检测过程如下：将有截幅失真的音频信号输入到检测系统中，首先对其进行全局扫描得到峰值大小，其次根据实际截幅类型选择适当的抖动值(数字截幅抖动值为零，模拟截幅则根据其截幅程度选择合适值)，最后由检测器得出截幅失真起始点信息，输入到修复系统。

实验证明，该方法可以简单有效地完成截幅失真位置的检测。对于数字截幅，由于检测原理是根据其产生特性而定的，所以检出率近 100%；而对于模拟截幅，其检出率则与实验采用的抖动值的大小直接相关，实验者应根据实际修复效果调整该值大小。

由于漏警和虚警是一对矛盾，一般情况下，实验者尽量保证漏警为零(即最大检出率)，因为后面修复算法的有效性可弥补虚警的存在，从而可以更有效地修复截幅信号。

3　截幅失真修复

在介绍 LSAR 修复算法之前，先回顾一下音频信号的 AR[7]模型与最小均方误差准则(MMSE)。

根据线性预测的原理，音频信号用 AR 模型表示，预测样点值与过去 p 个值的关系式如式(1)：

$$\hat{x}_n=\sum_{k=1}^{p}a_k x_{n-k} \tag{1}$$

其中，$\hat{x}_n$ 为待预测样点，x_n 为信号实际采样点，a_k (k=1，2，...，p)为 AR 模型系数。

由式(1)可以得到预测误差表达式如式(2)：

$$e_n=x_n-\hat{x}_n=x_n-\sum_{k=1}^{p}a_k x_{n-k} \tag{2}$$

当使得 $e_n^T e_n$ 取得最小值时，认为预测样点值最好地逼近原始值，即 MMSE 准则，也即 LSAR 算法基于的基本原理。

在具体阐述该算法之前，先给出真实截幅失真数据的细节图如图 2 所示：

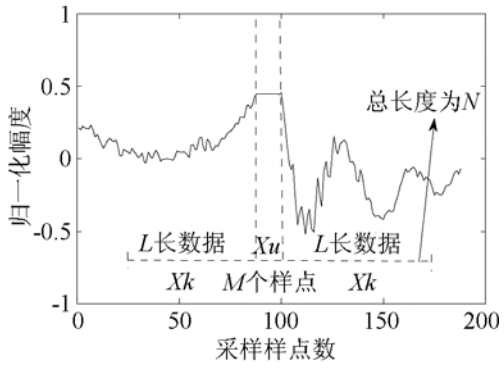


图 2　真实截幅失真数据的细节图

Fig.2　The detail description of the actual clipping distortion

其中 $\boldsymbol{x}_{U_n}(\boldsymbol{Xu})$为截幅失真数据，左右 $\boldsymbol{x}_{Kn}(\boldsymbol{Xk})$为未受影响数据，总的数据长度为 N。LSAR 详细算法如下：

图 2 中信号为 x，依据图后说明，可以将信号表示如式(3)所示。

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_{Kn1}\\ \boldsymbol{x}_U\\ \boldsymbol{x}_{Kn2}\end{pmatrix}=\begin{pmatrix}\boldsymbol{x}_{Kn1}\\ \boldsymbol{0}\\ \boldsymbol{x}_{Kn2}\end{pmatrix}+\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{x}_U\\ \boldsymbol{0}\end{pmatrix}=\boldsymbol{K}\boldsymbol{x}_{Kn}+\boldsymbol{U}\boldsymbol{x}_U \tag{3}$$

对于长度为 N 的数据，如果截幅失真位置开始于第 k 个采样点，截止于第 $k+M-1$ 个点处，即截幅段长度为 M 个采样点，取 AR 模型预测阶数为 p(p>=M)。

对于该 N 长数据，应用式(2)计算其预测误差并写成联合方程式如式(4)所示。

$$\begin{pmatrix} e(p) \\ e(p+1) \\ \vdots \\ e(k-1) \\ \text{--------} \\ e(k) \\ e(k+1) \\ e(k+2) \\ \vdots \\ e(k+M+p-2) \\ e(k+M+p-1) \\ \text{--------} \\ e(k+M+p) \\ e(k+M+p+1) \\ \vdots \\ e(N-1) \end{pmatrix} = \begin{pmatrix} x(p) \\ x(p+1) \\ \vdots \\ x(k-1) \\ \text{--------} \\ x_{Uk}(k) \\ x_{Uk}(k+1) \\ x_{Uk}(k+2) \\ \vdots \\ x(k+M+p-2) \\ x(k+M+p-1) \\ \text{--------} \\ x(k+M+p) \\ x(k+M+p+1) \\ \vdots \\ x(N-1) \end{pmatrix} - \begin{pmatrix} x(p-1) & x(p-2) & \cdots & x(0) \\ x(p) & x(p-1) & \cdots & x(1) \\ \vdots & \vdots & \ddots & \vdots \\ x(k-2) & x(k-3) & \cdots & x(k-p-1) \\ \text{----} & \text{----} & \text{----} & \text{----} \\ x(k-1) & x(k\text{-}2) & \cdots & x(k-p) \\ x_{Uk}(k) & x(k\text{-}1) & \cdots & x(k-p+1) \\ x_{Uk}(k+1) & x_{Uk}(k) & \cdots & x(k-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ x(k+M+p-3) & x(k+M+p-4) & \cdots & x_{Uk}(k+M-2) \\ x(k+M+p-2) & x(k+M+p-3) & \cdots & x_{Uk}(k+M-1) \\ \text{----} & \text{----} & \text{----} & \text{----} \\ x(k+M+p-1) & x(k+M+p-2) & \cdots & x(k+M) \\ x(k+M+p) & x(k+M+p-1) & \cdots & x(k+M+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-p-1) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{pmatrix} \quad (4)$$

将其用矢量表示如式(5)所示：

$$\boldsymbol{e}(\boldsymbol{x}_{Uk},\boldsymbol{a}) = \boldsymbol{x} - \boldsymbol{X}\boldsymbol{a} \quad (5)$$

对式(5)应用 MMSE 均方最小准则，得到：

$$\boldsymbol{e}^T\boldsymbol{e}(\boldsymbol{x}_{Uk},\boldsymbol{a}) = \boldsymbol{x}^T\boldsymbol{x} + \boldsymbol{a}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a} - \boldsymbol{a}^T\boldsymbol{X}^T\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{X}\boldsymbol{a} \quad (6)$$

要使该均方误差最小，由于式(6)中含 $\boldsymbol{a}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a}$ 四阶因子以及 $\boldsymbol{a}^T\boldsymbol{X}^T\boldsymbol{x}$ 和 $\boldsymbol{x}^T\boldsymbol{X}\boldsymbol{a}$ 三阶因子，对其进行微分后会有非线性方程产生，所以分成如下两步：

(1) 通过式(4)中虚线上下的已知数据，形成如下方程：

$$\hat{\boldsymbol{a}} = (\boldsymbol{X}_{Kn}^T\boldsymbol{X}_{Kn})^{-1}(\boldsymbol{X}_{Kn}^T\boldsymbol{x}_{Kn}) \quad (7)$$

从而求出 AR 模型预测系数 $\boldsymbol{a}$ 的值。

(2) 将第一步得到的 $\boldsymbol{a}$ 值代入式(4)中虚线间未知数据部分，并按照 $\boldsymbol{x}_{Kn}$、$\boldsymbol{x}_{Uk}$ 数据进行整理，最终得到预测值误差的矢量形式如式(8)所示：

$$\boldsymbol{e} = \boldsymbol{A}_1\boldsymbol{x}_{Uk} + \boldsymbol{A}_2\boldsymbol{x}_{Kn} \quad (8)$$

其中 $\boldsymbol{A}_1$ 和 $\boldsymbol{A}_2$ 是由预测系数 $\boldsymbol{a}$ 组成的矩阵。最后整理可以得到：

$$\hat{\boldsymbol{x}}_{Uk}^{LSAR} = (-\boldsymbol{A}_1^T\boldsymbol{A}_1)^{-1}(\boldsymbol{A}_1^T\boldsymbol{A}_2)\boldsymbol{x}_{Kn} \quad (9)$$

$\hat{\boldsymbol{x}}_{Uk}$ 即为 LSAR 算法得到的截幅修复值。

该方法充分利用了截幅点左右两边的数据信息，从而有效地避免了单向预测产生的结果数据单调增加或减少。当截幅点数过多时，截幅段中间数据的修复是有一定误差存在的，但基于双边线性预测的 LSAR 算法所达到的修复效果显示，这种误差不会给处理效果造成很大的影响。

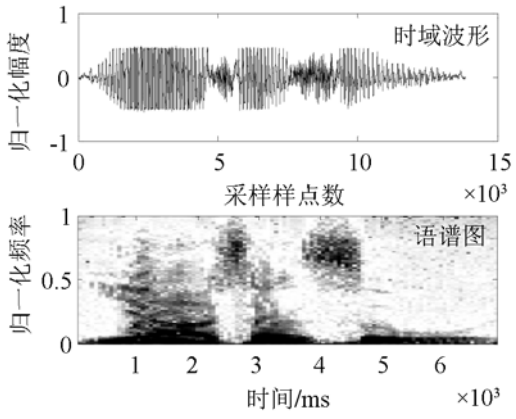
修复过程如下：由检测算法得到截幅失真起始点信息，$\boldsymbol{x}_{Uk}$ 取为截幅段数据，$\boldsymbol{x}_{Kn}$ 取其左右相同长度数据(由实验可得，该长度约 7.5ms)，总长为 N(变长)。该取法使得截幅位于未污染音频信号中间，保证有最多的有用信息可以利用，两侧的 7.5ms 保证了预测所需长度。另外，当音频信号采样率比较低时，相对截幅点数较少，选择截幅段长度 M 作为预测阶数 p 可能无法完成修复要求，实际算法中，若使 p 取 M 和 10 中的最大值，即可保证对原始音频信号至少 10 阶的线性预测，从而使低采样率数据修复结果具有高的可靠性。

不可避免的是，如果信号截幅程度很深，其修复效果会略差一些，但改进的听觉效果还是很明显的。究其略差的原因，对截幅很深的数据，实际波形并非整体连成一片被截幅，而是时间上有很多样点受到影响。本文的检测算法，只定位单个截幅片段，然后对该段截幅进行修复。如果截幅段很密集时，取左右两边数据必然会引入一定的误差(把相邻的截幅段数据当成未受影响数据引入算法中)，也即影响了修复效果。

4 实验结果与分析

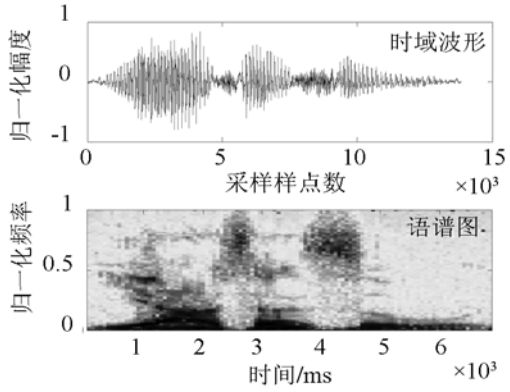
通过将该算法应用于不同采样率的模拟截幅以及数字截幅，来验证其截幅修复效果。

实验 1：算法应用于模拟截幅的音频信号中，信号采样率为 16kHz，其截幅修复前后的时域波形以及语谱图见图 3(时间单位为 ms)。由实验 1 可见，对于模拟截幅的音频信号，该算法修复效果明显。不同的是，由图 3 可以看出，修复后信号幅度有所压缩。这是由于数据显示采用归一化方法，对修复后数值，如果其值超出 1，在波形显示上依然处于截幅状态，因此需经过一个压缩处理过程才可见其效果。



(a)　模拟截幅的时域波形及其语谱图

(a)　Waveform and spectrogram of analogue clipping



(b)　模拟截幅修复后的时域波形及其语谱图

(b)　Waveform and spectrogram of analogue clipping after being restored

图 3　模拟截幅修复前后的时域波形及其语谱图

Fig.3　Waveform and spectrogram of analogue clipping before and after being restored

实验 2：算法应用于数字截幅的音频信号中，信号采样率为 44.1kHz，其截幅修复前后的时域波形以及语谱图见图 4(时间单位为 ms)。由实验 2 可以看出，对于数字截幅的音频信号，该算法修复效果明显。由时域波形可见，截幅地方的波形得到较好修复；由语谱图可见，原截幅处表现为一片竖直亮条，修复后的语谱接近真实信号。

综上所述，对于不同采样率的数据，无论模拟截幅或者数字截幅，从主观上讲，该算法均可以达到预期的处理效果。

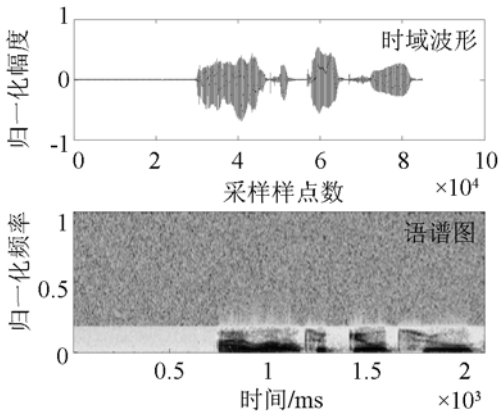
下面分别从信噪比及听觉感知模型角度，给出测试集文件(1-12)在本文提出的修复算法以及 Cooledit 中截幅修复模块处理下的主客观评测统计结果，由于两个角度评价均需要原始音频信号作为参考，所以统计只针对人工产生的数字截幅，选择不同采样率的数据，是为了验证算法的鲁棒性。

(1) 信噪比：任何失真都可归结为噪声来看待，对于截幅失真用信噪比表示如式(10)、(11)：

$$\text{截幅信噪比}=10*\lg\left(\frac{\text{原始音频幅度}^2}{(\text{原始音频幅度}-\text{截幅音频幅度})^2}\right) \tag{10}$$

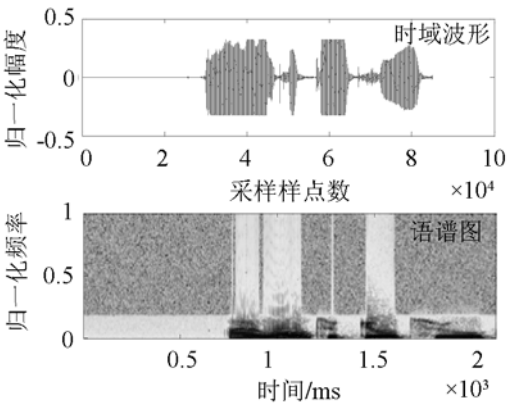
$$\text{修复后信噪比}=10*\lg\left(\frac{\text{原始音频幅度}^2}{(\text{原始音频幅度}-\text{修复后音频幅度})^2}\right) \tag{11}$$

依据公式(10)和(11)，得到音频信号在对应两种



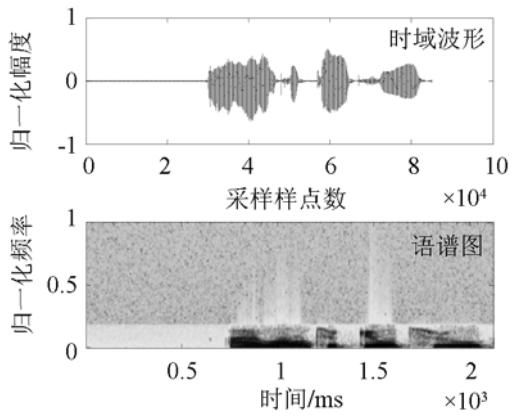
(a) 原始信号的时域波形及其语谱图

(a) Waveform and spectrogram of original signal



(b) 数字截幅的时域波形及其语谱图

(b) Waveform and spectrogram of digital clipping



(c)　数字截幅修复后的时域波形及其语谱图

(c)　Waveform and spectrogram of the digital clipping after being restored

图 4　原始信号及数字截幅修复前后的时域波形及其语谱图

Fig.4　Waveform and spectrogram of original signal and digital clipping before and after being restored

处理方式下的截幅修复前后信噪比改善统计表如表 1 所示。

表 1 音频信号截幅修复前后信噪比改善统计表
Table 1 The statistical results of SNR's improvement

原始音频/wav	原采样率/kHz	截幅信噪比/dB	本文算法修复后信噪比/dB	Cooledit 修复后信噪比/dB
1	44.1	15.741	30.809	27.275
2	44.1	10.961	25.404	24.120
3	44.1	10.898	20.763	12.169
4	44.1	11.860	22.783	18.622
5	44.1	12.724	20.682	12.711
6	44.1	10.121	21.852	17.439
7	8	15.547	16.202	16.832
8	8	15.158	16.959	16.905
9	8	16.127	20.649	23.228
10	16	17.355	19.783	17.934
11	16	18.822	27.546	20.820
12	16	12.255	16.356	13.590

由该统计信息可以看出，对于不同采样率的截幅数据，应用本文算法之后，其截幅修复后的信噪比得到了明显改善。与 Cooledit 修复后的结果对比可以看出，对于采样率高的截幅信号，本文算法具有较好的修复效果；而对于采样率低的截幅信号(如 7.wav 和 9.wav)，由于本文采用的 LSAR 算法中等效的可用信息样点数较少，其信噪比提升较差，不如 Cooledit 的处理效果。

(2) PESQ[8]评分方法：ITU-T 推荐 P.862 中的 PESQ 算法作为重建音频质量的主观评价，能够比较待测试音频信号与指定参考信号之间的听觉距离，提供一个类似于 MOS 的语音质量打分 PESQ MOS。

PESQ 所得分数与主观语音质量 MOS 取值域接近，且关联程度非常高，实践证明，在噪声干扰和波形编码等方面，PESQ 可以给出令人满意的预测值。然而目前算法做到的 PESQ 值计算只能固定两个采样率——16kHz 和 8kHz，所以将其应用于本文测试集时，有如下的约定：

对于原始采样率为 44.1kHz 的数据统一转化到 16 kHz 进行计算，而原始采样率为 16 kHz 或者 8 kHz 的数据，保持原采样率不变进行计算。

在此约定下，应用 PESQ 算法，得到音频信号在对应两种处理方式下的截幅修复前后 PESQ 值改善统计表如表 2 所示。

由该统计信息可以看出，对于不同采样率的截幅数据，应用本文算法之后，其截幅修复后 PESQ 值也得到了提高。与 Cooledit 修复后的结果对比可以看出，采样率越高的截幅信号，本文算法的修复效果越是明显，这与前面从信噪比的角度的分析结果是不相冲突的。表 2 中文件 5.wav 的异常，可能是在采样率转换时的一些偏差，导致了该异常的发生，但由表 2 可知，大多数的情况是较为理想的。

表 2 音频信号截幅修复前后 PESQ 值改善统计表
Table 2 The statistical results of PESQ's improvement

原始音频/wav	原采样率/kHz	截幅 PESQ 值	本文算法修复后 PESQ 值	Cooledit 修复后 PESQ 值	PESQ 采样率/kHz
1	44.1	2.559	3.825	3.513	16
2	44.1	1.977	3.333	3.240	16
3	44.1	1.941	3.107	2.857	16
4	44.1	2.098	3.112	2.953	16
5	44.1	2.366	3.525	2.569	16
6	44.1	2.179	3.223	3.171	16
7	8	3.370	3.889	3.771	8
8	8	3.308	3.390	3.413	8
9	8	3.255	3.599	3.752	8
10	16	3.760	4.159	3.750	16
11	16	3.374	4.314	3.645	16
12	16	3.833	4.190	3.920	16

由上面的分析可以知道，无论是用信噪比还是 PESQ 值来衡量该算法对音频信号的修复，均达到了一定的效果。通过对不同采样率的截幅失真数据进行修复表明，该算法具有一定的鲁棒性，可不同程度地提高音频信号质量。

5 结 论

本文提出的方法，对典型的截幅失真具有较好的修复效果，但现实环境是复杂多变的，要使其具有更强的自适应性，仍需进一步的努力。

参 考 文 献

[1] Jonathan S. Abel, Julius O. Smith. Restoring a clipped signal[A]. Acoustics, Speech, and Signal Processing, 1991. ICASSP-91[C]. 1991 International Conference on, 1745-1748.

[2] Tomas Olofsson. Deconvolution and model-based restoration of clipped ultrasonic signals[J]. Instrumentation and Measurement, IEEE Transactions, 2005, **54**(3): 1235-1240.

[3] Simon J. Godsill and Peter J. W. Rayner. Digital audio restoration[M]. US: Springer US. 1998, 106-121.

[4] Saeed V. Vaseghi. Advanced digital signal processing and noise reduction[M]. New York: Wiley. 2000, 313-331.

[5] Multitask Noisy Speech Enhancement System[DB/OL]. http://sound.eti.pg.gda.pl/denoise/clipping.html.

[6] CEDAR Tools[DB/OL]. www.cedar-audio.com/products/cedartools/cedartools.html.

[7] 赵力. 语音信号处理[M]. 北京: 机械工业出版社. 2005. 56-62.
ZHAO Li. Speech signal processing[M]. Beijing: China Machine Press. 2005. 56-62.
Antony W. Rix', John G. Beerends', Michael P. Hollier', etc. Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs[A]. Acoustics, Speech, and Signal Processing, 2001. Proceedings. ICASSP-01, 2001 IEEE International Conference[C]. 2001, 749-752.