

# 一种用于截幅音频修复中的自适应一致迭代硬阈值算法

邹霞 吴彭龙\* 孙蒙 张星昱

(陆军工程大学指挥控制工程学院 南京 210007)

**摘要:** 一致迭代硬阈值(CIHT)算法在处理音频截幅失真中具有较好的性能。但是,在截幅程度较大时音频截幅修复的性能会下降。因此,该文提出一种基于自适应门限的改进算法。该算法自动估计音频信号截幅程度,根据估计的截幅程度信息,自适应调整算法中的截幅程度因子。与近年来提出的CIHT算法和一致字典学习算法(CDL)相比,该文所提算法能更好地重建音频信号,特别在音频信号截幅失真严重的情况。该算法的运算复杂度与CIHT相近,与CDL相比,拥有更快的运行速度,有利于实时实现。

**关键词:** 音频信号处理; 截幅失真; 自适应门限; 一致迭代硬阈值

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2019)04-0925-07

DOI: 10.11999/JEIT180543

## An Adaptive Consistent Iterative Hard Thresholding Algorithm for Audio Declipping

ZOU Xia WU Penglong SUN Meng ZHANG Xingyu

(The Army Engineering University of PLA, Nanjing 210007, China)

**Abstract:** Audio clipping distortion can be solved by the Consistent Iterative Hard Thresholding (CIHT) algorithm, but the performance of restoration will decrease when the clipping degree is large, so, an algorithm based on adaptive threshold is proposed. The method estimates automatically the clipping degree, and the factor of the clipping degree is adjusted in the algorithm according to the degree of clipping. Compared with the CIHT algorithm and the Consistent Dictionary Learning (CDL) algorithm, the performance of restoration by the proposed algorithm is much better than the other two, especially in the case of severe clipping distortion. Compared with CDL, the computational complexity of the proposed algorithm is low like CIHT, compared with CDL, it has faster processing speed, which is beneficial to the practicality of the algorithm.

**Key words:** Audio signal processing; Clipping distortion; Adaptive threshold; Consistent Iterative Hard Thresholding (CIHT)

### 1 引言

现实生活中,由于录音设备和环境的影响,音频信号会出现不同程度的失真。其中,截幅失真是一种较为常见的失真类型。截幅的发生,会给信号造成刺耳的噪声。如果截幅程度过大,会严重影响音频的听觉质量,且不利于音频信号的进一步处理。

如何根据前后未被截幅的信号来恢复被截幅的信号是一项有意义的研究。目前针对截幅信号修复

的方法包括:自回归模型<sup>[1]</sup>、受限带宽模型<sup>[2]</sup>和贝叶斯估计<sup>[3]</sup>等。近年来,基于稀疏表示的信号修复技术引起了广泛关注。这类方法的基本思想在于音频信号可以用字典进行稀疏表示。

基于音频信号的稀疏表示思想,文献<sup>[4]</sup>通过舍弃被截幅的样本点,将其描述为未被截幅样本点上的稀疏分解问题。在重建音频信号过程中,使用额外的有用信息有助于改善重建信号的质量。根据截幅样本点的有效阈值这一已知信息,通过增加对截幅样本点的幅度约束,能改善修复效果。基于这一约束,文献<sup>[5-7]</sup>提出了振幅约束的稀疏分解,运用两步聚类算法或使用凸优化工具箱进行求解。同样,使用其它有用信息也会使修复效果有所改善,如使用人耳听觉信息<sup>[6]</sup>、多通道数据<sup>[8]</sup>、先验知识<sup>[9]</sup>等。

在文献<sup>[4,5]</sup>中,利用约束的正交匹配追踪算法对截幅信号进行分解并重构。在信号的修复上,虽

收稿日期: 2018-06-04; 改回日期: 2018-12-04; 网络出版: 2018-12-13

\*通信作者: 吴彭龙 17551050128@163.com

基金项目: 国家自然科学基金(61402519), 江苏省优秀青年基金(BK20180080)

Foundation Items: The National Natural Science Foundation of China (61402519), The Natural Science Foundation of Jiangsu Province for Excellent Young Scholars (BK20180080)

然有良好的修复效果,但算法的运算复杂度高,需要较长时间重构音频信号,不利于算法的实时实现。为使算法在不影响修复效果的前提下,能拥有较快的运算速度,文献[10]通过加入正则项,构造出一个统一的目标函数,就可以使用已知算法进行优化,如迭代硬阈值(Iterative Hard Thresholding, IHT)算法。此类算法使用梯度下降的迭代方法,能大幅度减少信号修复所需的时间,使得适用场景更加丰富。

以上方法都是基于固定字典对信号进行稀疏分解,再通过重构信号达到修复目的。文献[11]提出一种一致字典学习(Consistent Dictionary Learning, CDL)的方法重构信号。此算法首先使用一致迭代硬阈值(Consistent Iterative Hard Thresholding, CIHT)算法,在过完备离散余弦变换(Discrete Cosine Transformation, DCT)字典上对信号进行稀疏分解,得到稀疏表示系数。然后,固定稀疏表示系数对字典进行更新,迭代若干次后得到一个训练字典和信号的稀疏表示系数,最终重构信号。此算法的修复效果要优于CIHT算法,但是由于存在字典学习过程,此算法需要较长的运行时间。但是,这些算法在音频信号截幅程度较大的情况下,对信号的修复效果还不够理想。

综上所述,当前的截幅信号修复还存在以下问题。一是有些算法需要的修复时间太长,二是针对截幅程度不同的信号,修复算法的自适应能力差,特别当截幅程度较大时,大多数算法不能取得好的修复效果。针对这些问题,本文对文献[10]中的CIHT算法提出改进,在继承CIHT算法良好的修复效果和较快的运算速度的基础上,提出自适应一致迭代硬阈值算法(Adaptive Consistent Iterative Hard Thresholding, ACIHT),此算法同时利用了截幅门限的信息和截幅程度的有用信息,来提高修复效果。

## 2 截幅失真

截幅失真可以看作是动态范围压缩的一种特例,是当音频信号幅值超出设备允许的峰值时产生的一种失真现象。如图1所示,当样本点的幅值大于允许的最大幅值时,此样本点将会被截幅。

令  $\mathbf{x} \in \mathbb{R}^N$  为原始音频信号,  $\mathbf{y} \in \mathbb{R}^N$  为受到截幅的音频信号,因此可将  $\mathbf{y}$  表示为

$$y_i = \begin{cases} \theta^+, & x_i \geq \theta^+ \\ \theta^-, & x_i \leq \theta^- \\ x_i, & \theta^- < x_i < \theta^+ \end{cases} \quad (1)$$

其中,  $\theta^+, \theta^-$  分别为正截幅门限和负截幅门限,

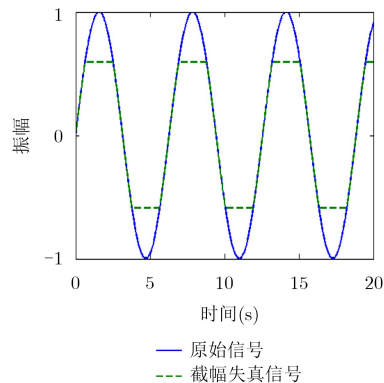


图1 信号截幅失真示意图

$x_i$  为原始样本点,  $N$  为样本点总个数。因此

$$\mathbf{y} = \mathbf{M}_r \mathbf{x} + \theta^+ \mathbf{M}_{c+} \mathbf{1} + \theta^- \mathbf{M}_{c-} \mathbf{1} \quad (2)$$

其中,  $\mathbf{1} \in \mathbb{R}^N$  是元素全为1的向量,  $\mathbf{M}_r \in \{0, 1\}^{N \times N}$ ,  $\mathbf{M}_{c+} \in \{0, 1\}^{N \times N}$ ,  $\mathbf{M}_{c-} \in \{0, 1\}^{N \times N}$  为掩蔽矩阵, 其中,  $\mathbf{M}_r$  是保留  $N$  阶单位矩阵中与  $\mathbf{x}$  的可靠部分对应的  $I_r$  行, 令  $I_{c+}$  和  $I_{c-}$  行的元素全为0得到的, 同理,  $\mathbf{M}_{c+}$  是保留  $N$  阶单位矩阵中与  $\mathbf{x}$  受到正截幅部分对应的  $I_{c+}$  行, 令  $I_r$  和  $I_{c-}$  行的元素全为0得到的,  $\mathbf{M}_{c-}$  是保留  $N$  阶单位矩阵中与  $\mathbf{x}$  受到负截幅部分对应的  $I_{c-}$  行, 令  $I_{c+}$  和  $I_r$  行的元素全为0得到的,  $\{I_r, I_{c+}, I_{c-}\}$  为音频信号矢量的支撑集  $I \triangleq \{1, 2, \dots, N\}$  的一个划分。下标  $r$  表示未受到截幅失真, 下标  $c+$ ,  $c-$  表示受到正, 负截幅。

## 3 修复算法

### 3.1 CIHT算法

一种简单的处理截幅问题的方法是直接使用 IHT<sup>[12]</sup> 算法, 依靠未被截幅的样本点对音频进行修复。但此方法没有考虑到受到截幅的样本点的有用信息, 导致修复效果并不理想。因此, 文献[10]对优化函数提出改进, 考虑问题如式(3)所示

$$\hat{\mathbf{a}} = \arg \min_{\tilde{\mathbf{a}} \in \mathbb{R}^N} \frac{1}{2} \|\beta(\mathbf{y} - D\tilde{\mathbf{a}})\|^2, \quad \text{s.t. } \|\tilde{\mathbf{a}}\|_0 \leq K \quad (3)$$

其中,  $\mathbf{y}$  为观测信号,  $D$  为过完备 DCT 字典,  $\beta(u) = \mathbf{M}_r u + (\mathbf{M}_{c+} u)_+ + (\mathbf{M}_{c-} u)_-$ ,  $K$  为稀疏度约束常数,  $\|\tilde{\mathbf{a}}\|_0 \leq K$  表示向量  $\tilde{\mathbf{a}}$  中非零元素的个数小于等于  $K$ 。最小化

$$\frac{1}{2} \|\beta(\mathbf{y} - D\tilde{\mathbf{a}})\|^2 \quad (4)$$

等价于

$$\min_{\tilde{\mathbf{a}} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{M}_r(\mathbf{y} - D\tilde{\mathbf{a}})\|^2 + \frac{1}{2} \|\mathbf{M}_{c+}(\theta^+ \mathbf{1} - D\tilde{\mathbf{a}})_+\|^2 + \frac{1}{2} \|\mathbf{M}_{c-}(\theta^- \mathbf{1} - D\tilde{\mathbf{a}})_-\|^2, \quad \text{s.t. } \|\tilde{\mathbf{a}}\|_0 \leq K \quad (5)$$

其中,  $(u)_+ = \max(0, u)$ ,  $(u)_- = \min(0, u)$ 。式(5)中的第2项和第3项使用了被截幅样本点的有用信息, 使用梯度下降的方法可求解 $\hat{\mathbf{a}}$ 。

### 3.2 ACIHT算法

现在大多数针对截幅音频信号的修复算法在截幅程度不大的情况下有较好的修复效果, 但对于截幅严重的音频信号, 修复效果并不理想, 有些算法甚至会引起更严重的失真。本文在式(5)的基础上进行了改进, 考虑到截幅程度越大(截幅门限的绝对值越小), 受到截幅失真影响的样本点就越多, 原始信号样本点的振幅与截幅门限的差值越大(正截幅)或越小(负截幅)。因此, 本文通过加入截幅程度因子 $\gamma(k)$ 对式(5)的一致性进行加强, 于是考虑模型如式(6)所示

$$\begin{aligned} \min_{\tilde{\mathbf{a}} \in \mathbb{R}^N} & \frac{1}{2} \|\mathbf{M}_r(\mathbf{y} - \mathbf{D}\tilde{\mathbf{a}})\|^2 \\ & + \frac{1}{2} \|\mathbf{M}_{c+}[\gamma(k)(\theta^+ \mathbf{1} - \mathbf{D}\tilde{\mathbf{a}})]_+\|^2 \\ & + \frac{1}{2} \|\mathbf{M}_{c-}[\gamma(k)(\theta^- \mathbf{1} - \mathbf{D}\tilde{\mathbf{a}})]_-\|^2, \\ \text{s.t.} & \|\tilde{\mathbf{a}}\|_0 \leq K \end{aligned} \quad (6)$$

其中,  $\gamma(k) = \exp(2.279 \times \exp(0.3815 \times \ln k))$ ,  $k$ 为截幅比 $N_c/N$ ,  $N_c$ 为受到截幅(正截幅和负截幅)的样本点个数,  $N$ 为总样本点个数。

截幅程度因子 $\gamma(k)$ 是利用纯净音频信号进行不同程度的人为截幅, 然后针对不同截幅程度的音频信号, 选择有最优修复效果的 $\gamma$ 值进行拟合得到的。实验阶段选用15段语音信号和15段音乐信号对 $\ln(\gamma)$ 和对数截幅比 $\ln(k)$ 的关系进行拟合。对每段信号进行不同程度的截幅, 并找出在每个截幅程度下, 能兼顾提高信号失真比和听觉质量的 $\gamma$ , 最后得到测试数据散点图, 如图2所示。

图2描述了 $\ln(\gamma)$ 与对数截幅比 $\ln(k)$ 的关系。其中, 不同形状的标记表示不同的音频信号, 即语音信号或音乐信号。由散点图可以发现, 当音频信号的截幅程度越大, 即截幅样本点越多,  $\gamma$ 的值就越

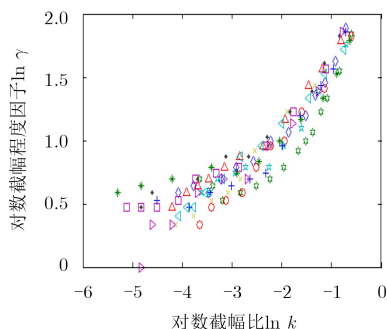


图2 测试数据散点图

大,  $\gamma$ 随 $k$ 递增。由图2可以发现, 在相同截幅比的情况下, 对于不同音频信号, 拥有最佳修复效果的 $\gamma$ 值不完全相同, 但却比较接近, 而且图2中所有样本点分布比较集中, 有较强的指数关系。因此, 对散点图进行拟合, 如图3所示, 得到拟合函数 $\gamma(k) = \exp(2.279 \times \exp(0.3815 \times \ln k))$ 。

本文所提算法是在式(5)的基础上加入截幅程度因子 $\gamma(k)$ 得到式(6), 其目的是为了修复后的音频信号更加接近于原始干净信号。当信号受到轻微截幅时, 信号本身蕴含的有用信息比较多, 使用CIHT算法或CDL算法重构信号就能较好地对截幅信号进行修复。但由于ACIHT算法在修复过程中使用了信号截幅程度的信息对式(5)的一致性进行加强, 且拟合函数 $\gamma(k)$ 是恒大于1的, 所以在信号受到轻微截幅的情况下, 使用式(6)来修复信号有可能会使得截幅样本点的幅值在经过重构后会大于原始样本点的幅值, 从而引起新的失真, 对听觉造成影响, 如图4所示。

通过实验发现, 当音频信号受到轻微截幅失真, 特别是当截幅比 $k < 0.02$ 时, 容易出现截幅样本点的幅值在经过重构后会大于原始样本点的幅值的情况。为解决此问题, 本文对截幅程度因子的拟合函数进行修正, 得到新的截幅程度因子 $\eta(k)$ 的拟合函数如式(7)

$$\eta(k) = [\gamma(k)]^p \quad (7)$$

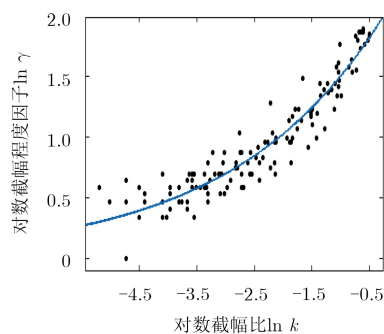


图3 测试数据拟合图

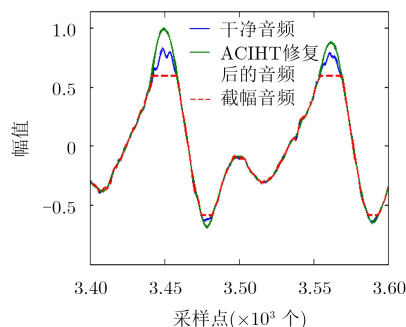


图4 失真示意图

其中,  $\gamma(k) = \exp(2.279 \times \exp(0.3815 \times \ln k))$ ,  $\rho$ 为sigmoid函数的变形, 即 $\rho = 1/(1 + e^{-2 \times 10^3 \times (k-0.02)})$ ,  $\rho$ 的函数曲线图如图5所示。

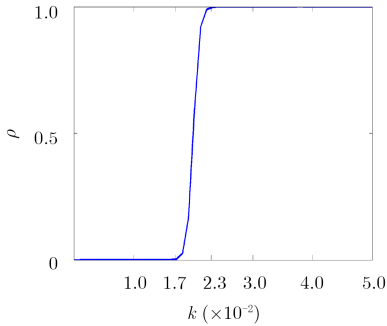


图5 函数曲线图

由图5可以发现, 当音频信号受到轻微截幅, 即截幅比 $k$ 满足 $0 < k < 1.7 \times 10^{-2}$ 时,  $\rho = 0$ , 此时 $\eta(k) = 1$ , 在信号的修复过程中不会使用截幅程度的信息; 当截幅比 $k$ 满足 $2.3 \times 10^{-2} < k < 1$ 时,  $\rho = 1$ , 此时 $\eta(k) = \gamma(k)$ , 截幅程度的信息将会应用到信号的修复过程中; 当截幅比 $k$ 满足 $1.7 \times 10^{-2} \leq k \leq 2.3 \times 10^{-2}$ 时,  $0 < \rho < 1$ , 即 $1 < \eta(k) < \gamma(k)$ , 此时截幅程度的信息会经过削弱再应用到信号的修复中。因此, 使用新的截幅程度因子 $\eta(k)$ 对式(5)的一致性进行加强可以避免截幅样本点的幅值在经过重构后会大于原始样本点的幅值的情况, 于是对式(6)进行修正后得到模型如式(8)所示

$$\begin{aligned} \min_{\tilde{\mathbf{a}} \in \mathbb{R}^N} & \frac{1}{2} \|\mathbf{M}_r(\mathbf{y} - D\tilde{\mathbf{a}})\|^2 \\ & + \frac{1}{2} \|\mathbf{M}_{c+}[\eta(k)(\theta^+ \mathbf{1} - D\tilde{\mathbf{a}})]_+\|^2 \\ & + \frac{1}{2} \|\mathbf{M}_{c-}[\eta(k)(\theta^- \mathbf{1} - D\tilde{\mathbf{a}})]_-\|^2, \\ \text{s.t.} & \quad \|\tilde{\mathbf{a}}\|_0 \leq K \end{aligned} \quad (8)$$

本文所提方法通过识别截幅音频信号中每个样本点的幅度值, 判断该样本点是否受到截幅失真, 从而得到掩蔽矩阵和截幅比。当音频信号 $\mathbf{y}$ 受到截幅失真时,  $\mathbf{y}$ 的幅度值满足 $\mathbf{y} \in [\theta^-, \theta^+]$ 。若令 $\mathbf{y}_i$ 为截幅音频信号的第 $i$ 个样本点, 若 $\theta^- < \mathbf{y}_i < \theta^+$ , 则判定该样本为可靠样本点, 即未受到截幅失真的样本点; 当 $\mathbf{y}_i = \theta^+$ 时, 则判定该样本点为受到正截幅的样本点; 当 $\mathbf{y}_i = \theta^-$ 时, 则判定该样本点为受到负截幅的样本点。由此可得掩蔽矩阵 $\mathbf{M}_r$ 、 $\mathbf{M}_{c+}$ 和 $\mathbf{M}_{c-}$ 。通过对所有样本点的判定可以得到截幅样本点的个数为 $N_c = N - N_r$ , 其中 $N_c$ 为受到截幅(正截幅和负截幅)的样本点个数,  $N_r$ 为可靠样本点个数,  $N$ 为总样本点个数。最后通过截幅比

$k = N_c/N$ 及拟合公式可以求得 $\gamma$ 和 $\eta$ 的值, 并运用于算法中。

## 4 实验结果与分析

采用两段语音信号和两段音乐信号对本文所提自适应算法模型(ACIHT)进行评估, 信号的采样率为16 kHz, 以便进行主观语音质量评估(Perceptual Evaluation of Speech Quality, PESQ)<sup>[13,14]</sup>测试。处理窗长为256, 使用75%的重叠。窗函数采用汉明窗, 过完备DCT字典为512点。稀疏度约束常数 $K$ 取32, 也就是在迭代过程中, 对于向量 $\tilde{\mathbf{a}}$ , 最多只保留最大的32个元素, 令剩余元素全为0。评价采用两个指标, 信号失真比<sup>[11]</sup>和PESQ值。

### 4.1 信号失真比

信号失真比定义为

$$\text{SDR} = 10 \lg \frac{\|\mathbf{x}\|^2}{\|\mathbf{x} - \mathbf{y}\|^2} \quad (9)$$

其中,  $\mathbf{x}$ 为无失真信号,  $\mathbf{y}$ 为失真信号。本文采用音频信号截幅样点的信号失真比的改善值对算法进行评价, 即

$$\begin{aligned} \Delta \text{SDR} = 10 \lg & \frac{\|(\mathbf{M}_{c+} + \mathbf{M}_{c-})\mathbf{x}\|^2}{\|(\mathbf{M}_{c+} + \mathbf{M}_{c-})(\mathbf{x} - \hat{\mathbf{x}})\|^2} \\ & - 10 \lg \frac{\|(\mathbf{M}_{c+} + \mathbf{M}_{c-})\mathbf{x}\|^2}{\|(\mathbf{M}_{c+} + \mathbf{M}_{c-})(\mathbf{x} - \mathbf{y})\|^2} \end{aligned} \quad (10)$$

其中,  $\mathbf{x}$ 为纯净音频信号,  $\hat{\mathbf{x}}$ 为修复后的音频信号,  $\mathbf{y}$ 为截幅失真的音频信号。

图6和图7给出了3种算法对截幅音频信号进行修复后SDR的改善值对比。截幅程度从严重(SDR=2 dB)到轻微(SDR=20 dB)。由图6和图7可以发现, CDL算法优于CIHT算法。相比于CIHT算法和CDL算法, ACIHT算法对截幅信号的修复效果有明显提升。特别是在信号受到严重截幅时, 经过CIHT算法和CDL算法修复后, 信号的质量提升并不明显, 但使用本文所提ACIHT算法对信号进行修复后, 质量有较为明显的提升。

与CIHT算法和CDL算法相比, 本文所提算法不仅能使截幅信号的信号失真比有较大的提升, 而且在重构音频信号波形上展现出优良的性能, 如图8和图9所示。

图8和图9给出了1段语音信号和1段音乐信号经过3种算法修复后的波形对比图, 其中两段截幅信号的信号失真比都为12 dB, 在语音信号中, 有5.8%的样本点受到截幅失真, 且 $\eta = 2.158$ , 在音乐信号中, 有3.3%的样本点受到截幅失真, 且 $\eta = 1.859$ 。由图8和图9可见, 无论是语音信号还是音乐信号,

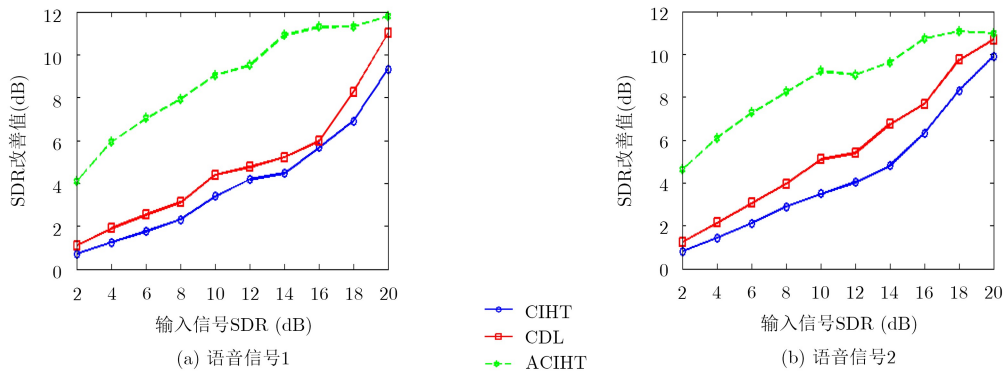


图6 截幅语音信号修复后的SDR提升对比图

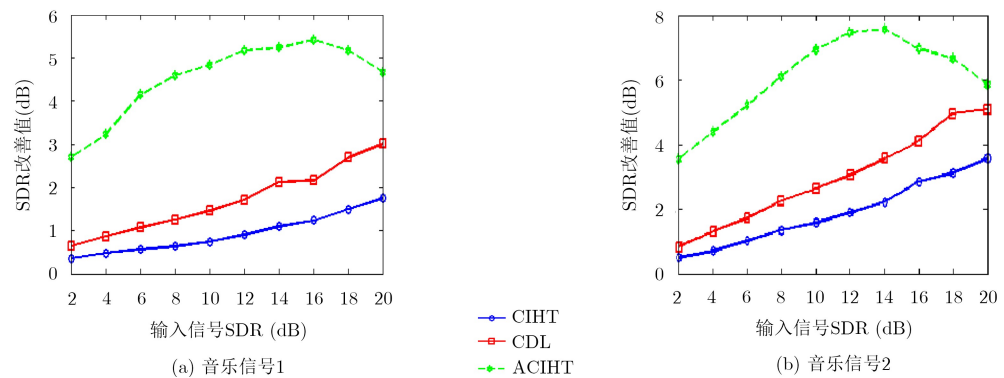


图7 截幅音乐信号修复后的SDR提升对比图

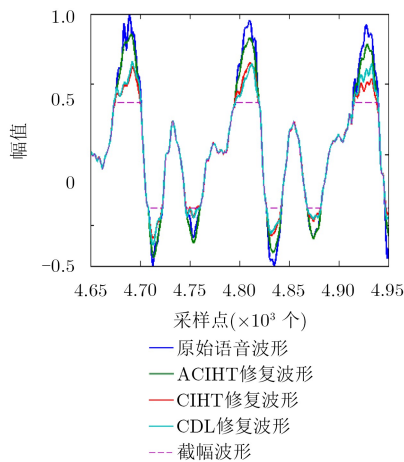


图8 截幅语音信号修复后波形对比图

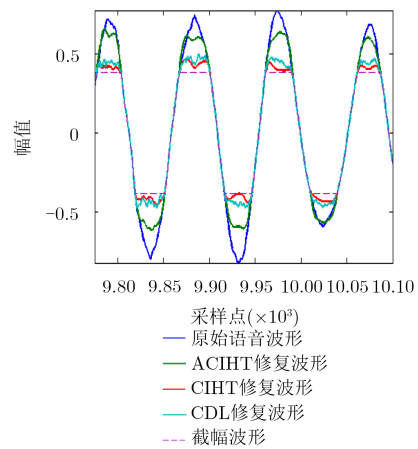


图9 截幅音乐信号修复后波形对比图

ACIHT算法都能较好地重构信号波形。与CIHT算法和CDL算法相比，经过ACIHT算法重构后的信号波形与原始信号波形更相似，且截幅样本点经过ACIHT算法修复后的幅值与原始幅值更接近。

在运行环境为Intel(R) Core(TM) i5-5200U CPU @ 2.20 GHz的条件下，对运算复杂度进行了测试。修复1段5 s的音频信号，当此音频信号受到严重截幅(SDR = 4 dB)时，CIHT算法需要7.891 s，ACIHT算法需要8.601 s，而CDL算法需要79.108 s。所提算法在没有大幅增加运算复杂度的情况下改进了CIHT算法。

### 4.2 PESQ打分

本文对两段语音信号进行不同程度的人为截幅，对每段语音信号使用CIHT算法、CDL算法和ACIHT算法进行修复。最后，对每种方法修复后的语音信号进行PESQ评分，得到PESQ值如表1、表2所示。

由表1和表2可以看出，对于不同语音信号，运用ACIHT算法之后，其修复后的PESQ值也得到明显提高，相比于CIHT算法和CDL算法，ACIHT算法展现出更加优越的修复效果。从表1和表2还可以发现，在某些截幅程度上，使用CDL算法修复后的

表1 截幅语音信号1修复前后PESQ得分比较

输入语音 SDR (dB)	截幅语 音	CIHT算法 修复后	CDL算法 修复后	ACIHT算法 修复后
2	1.8838	2.0877	2.1201	<b>2.2400</b>
4	2.2041	2.4411	2.4860	<b>2.6039</b>
6	2.3451	2.6239	2.6247	<b>2.8551</b>
8	2.5084	2.7974	2.7608	<b>3.1258</b>
10	2.6576	3.0501	3.0715	<b>3.2847</b>
12	2.7951	3.2538	3.3020	<b>3.6657</b>
14	2.9858	3.4915	3.5057	<b>3.8716</b>
16	3.1098	3.6881	3.6223	<b>4.1016</b>
18	3.2984	3.8203	3.7167	<b>4.2174</b>
20	3.4128	4.1934	4.2224	<b>4.2440</b>

表2 截幅语音信号2修复前后PESQ得分比较

输入语音 SDR (dB)	截幅语 音	CIHT算法 修复后	CDL算法 修复后	ACIHT算法 修复后
2	1.7080	1.8980	1.9802	<b>2.2026</b>
4	1.9977	2.3065	2.3451	<b>2.5566</b>
6	2.2115	2.6041	2.6537	<b>2.7818</b>
8	2.3900	2.8904	2.9176	<b>3.0617</b>
10	2.5946	3.1397	3.2116	<b>3.3242</b>
12	2.7625	3.4662	3.4546	<b>3.4784</b>
14	2.9359	3.7844	3.8324	<b>3.9217</b>
16	3.2481	3.9820	4.0386	<b>4.0463</b>
18	3.3362	4.2150	4.1375	<b>4.3005</b>
20	3.4004	4.3186	4.2845	<b>4.3500</b>

PESQ值低于使用CIHT算法修复后的PESQ值。CDL算法是在CIHT算法的基础上,使用未被截幅样本点的信息对字典进行更新,得到新的字典,在新的字典上,未被截幅的样本点能很好地被重构。但由于截幅样本点的真实幅度未知,相比于使用未更新的字典,使用新字典反而不能很好地重构受到截幅的样本点,从而影响听觉质量。

## 5 结论

本文提出一种能自适应不同截幅程度的音频修复算法,此算法对不同截幅程度的音频信号修复表现出了优良的性能,对典型的截幅失真具有较好的修复结果,能更好地适应于现实环境中的截幅情况。本文算法在没有大幅增加运算复杂度的情况下改进了CIHT算法,且拥有比CDL算法更低的运算复杂度,有利于实时实现。

## 参考文献

[1] JANSSEN A, VELDHUIS R, and VRIES L. Adaptive interpolation of discrete-time signals that can be modeled as

autoregressive processes[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1986, 34(2): 317–330. doi: [10.1109/TASSP.1986.1164824](https://doi.org/10.1109/TASSP.1986.1164824).

- [2] ABEL J S and ABEL J S. Restoring a clipped signal[C]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Canada, 1991, 3: 1745–1748. doi: [10.1109/ICASSP.1991.150655](https://doi.org/10.1109/ICASSP.1991.150655).
- [3] SIMON J G, PATRICK J, and WILLIAM N W. Statistical model-based approaches to audio restoration and analysis[J]. *Journal of New Music Research*, 2001, 30(4): 323–338. doi: [10.1076/jnmr.30.4.323.7489](https://doi.org/10.1076/jnmr.30.4.323.7489).
- [4] ADLER A, EMIYA V, and JAFARI M G. Audio inpainting[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(3): 922–932. doi: [10.1109/TASL.2011.2168211](https://doi.org/10.1109/TASL.2011.2168211).
- [5] ADLER A, EMIYA V, and JAFARI M G. A constrained matching pursuit approach to audio declipping[C]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Prague, Czech Republic, 2011: 329–332. doi: [10.1109/ICASSP.2011.5946407](https://doi.org/10.1109/ICASSP.2011.5946407).
- [6] DEFRAENE B, MANSOUR N, and HERTOOGH S D. Declipping of audio signals using perceptual compressed sensing[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2013, 21(12): 2627–2637. doi: [10.1109/TASL.2013.2281570](https://doi.org/10.1109/TASL.2013.2281570).
- [7] FOUCAERT S and NEEDHAM T. Sparse recovery from saturated measurements[J]. *Information and Inference: A Journal of the IMA*, 2017, 6(2): 196–212. doi: [10.1093/imaiai/iaw020](https://doi.org/10.1093/imaiai/iaw020).
- [8] OZEROV A, BILEN C, and PEREZ P. Multichannel audio declipping[C]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Shanghai, China, 2016: 659–663.
- [9] KAI S, KOWALSKI M, and DORFLER M. Audio declipping with social sparsity[C]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, 2014: 1577–1581. doi: [10.1109/ICASSP.2014.6853863](https://doi.org/10.1109/ICASSP.2014.6853863).
- [10] KITIC S, JACQUES L, and MADHU N. Consistent iterative hard thresholding for signal declipping[C]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, 2013: 5939–5943. doi: [10.1109/ICASSP.2013.6638804](https://doi.org/10.1109/ICASSP.2013.6638804).
- [11] RENCKER L, BACH F, WANG Wenwu, et al. Consistent dictionary learning for signal declipping[C]. *International Conference on Latent Variable Analysis and Signal Separation*, Guildford, UK, 2018: 446–455. doi: [10.1007/978-](https://doi.org/10.1007/978-)

- 3-319-93764-9\_41.
- [12] LECUE G and FOUCART S. An IHT algorithm for sparse recovery from sub-exponential measurements[J]. *IEEE Signal Processing Letters*, 2017, 24(3): 1280–1283. doi: [10.1109/LSP.2017.2721500](https://doi.org/10.1109/LSP.2017.2721500).
- [13] HINES A, SKOGLUND J, and KOKARAM A. Robustness of speech quality metrics to background noise and network degradations: Comparing ViSQOL, PESQ and POLQA[C]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, 2013: 3697–3701. doi: [10.1109/ICASSP.2013.6638348](https://doi.org/10.1109/ICASSP.2013.6638348).
- [14] 何孝月. 基于EPESQ的VoIP语音质量评估的研究与实现[D]. [硕士学位论文], 中南大学, 2008. doi: [10.7666/d.y1326186](https://doi.org/10.7666/d.y1326186).
- HE Xiaoyue. Speech Quality Evaluation of VoIP Based on EPESQ[D]. [Master dissertation], Central South University, 2008. doi: [10.7666/d.y1326186](https://doi.org/10.7666/d.y1326186).

邹霞：男，1979年生，副教授，研究方向为语音信号处理等。

吴彭龙：男，1995年生，硕士，研究方向为智能信息处理等。

孙蒙：男，1984年生，讲师，研究方向为机器学习和语音处理等。

张星昱：男，1994年生，硕士，研究方向为信息安全处理等。