

文章编号:1004-9037(2014)02-0211-11

基于核 Fisher 判别和加权码书映射的音频信号削波修复方法

邓 峰 鲍长春 鲍 枫

(北京工业大学电子信息与控制工程学院,北京,100124)

摘要:基于核 Fisher 判别(Kernel fisher discriminant, KFD)和加权码书映射(Weighted codebook mapping, WCBM),提出了一种 MDCT(Modified discrete cosine transform)域的音频信号削波修复方法。首先根据音频信号的 MDCT 系数提取子带包络等四种削波特征参数;其次,利用这些特征参数训练检测音频信号出现削波的核 Fisher 分类器;最后,利用子带包络的 WCBM 来修复音频信号的削波。测试结果表明,本文所提方法能有效修复音频信号的削波,其性能优于现有的几种削波修复方法。

关键词:核 Fisher 判别;加权码书映射;修正离散余弦变换;削波修复;削波检测

中图分类号:TN912.3

文献标志码:A

Clipping Restoration of Audio Signals Based on Kernel Fisher Discriminant and Weighted Codebook Mapping

Deng Feng, Bao Changchun, Bao Feng

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing, 100124, China)

Abstract: A new clipping restoration method of audio signals is proposed based on kernel Fisher discriminant (KFD) and weighted codebook mapping (WCBM) in modified discrete cosine transform (MDCT) domain. Firstly, the four clipping features are extracted according to the MDCT coefficients of audio signals. Secondly, the clipping feature parameters are used to train the optimal kernel Fisher classifier which is employed to detect the clipping. Finally, the WCBM of sub-band envelop is adopted to restore the clipping of audio signals. The test results indicate that the proposed method can effectively remove the clipping distortion of audio signals and obviously outperforms the existing clipping restoration methods.

Key words: kernel Fisher discriminant; weighted codebook mapping; modified discrete cosine transform; clipping restoration; clipping detection

引 言

在将老旧磁带或唱片转录成数字信号的过程中,数字音频信号经常会发生削波现象^[1-6],为了恢复原始音频的听觉效果,需对削波的数字音频信号进行修复。

削波主要分为两类^[5-6]:数字削波和模拟削波。数字削波是指音频信号幅度超过了录音存储设备的幅值动态范围,从而会在峰值处产生削波现象,

其削波的幅值严格相等;模拟削波是指在模拟系统中,由于高电平信号超出了模拟器件的线性区而产生削波现象,此时削波处的幅值不完全相等,会有一些微小抖动。本文主要考虑数字削波的修复,图 1 给出了一段音频信号削波前后的波形对比图。从图 1 可以看出,削波信号的峰-峰值受到了严重损坏,峰-峰值趋于平坦。这种削波将严重影响听觉质量。

传统的削波修复方法都在时域完成,即首先检测削波的起始位置,然后修复削波。如,文献[1]采

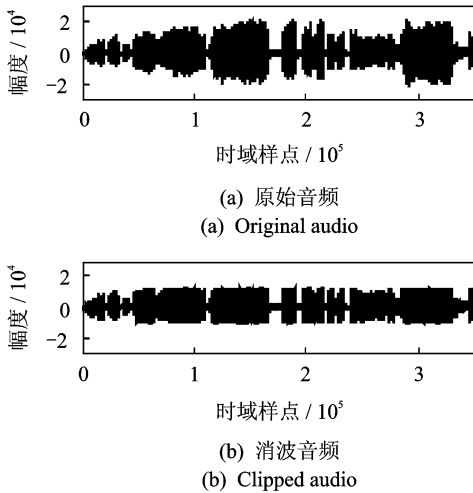


图 1 削波前后的音频信号波形对比

Fig. 1 Waveform comparison between original and clipped audios

用了基于概率模型的贝叶斯决策理论来检测消波的起始位置,然后利用 AR(Auto regressive)模型对削波信号进行平滑,以达到修复削波的目的。文献[2]则将音频信号的削波样点看作是缺失样点,并在带限信号的约束条件下,将削波修复问题构建为 l_2 -范数的最小化问题,通过非削波样点来重建削波样点,但是该方法仅适用于带限信号,不具有普适性。文献[3]利用最大后验概率准则,提出一种解卷积算法来修复削波,该算法虽然对削波的修复具有一定的效果,但是其计算复杂度较高,不适合实际的工程应用。文献[4]采用三次样条插值的方法进行削波修复,该方法可以较好地修复削波音频信号的时域波形,但是其不能有效消除因削波而引入的多余频谱成份,从而导致修复音频的听觉质量没有得到有效改善。文献[5]根据削波发生在音频信号峰值附近的特点,提出时域峰值检测算法检测削波的起始位置,然后假设这些削波样点为缺失样点,采用最小二乘自回归(Least square auto regressive, LSAR)模型进行预测插值来恢复这些削波样点,该方法的性能严重依赖削波样点的数目多少,普适性较差。文献[6]也采用了时域峰值检测方法来检测削波的起始位置,然后提出软阈值函数和谱加权函数相结合的方法来修复削波,该方法对音频信号的削波具有一定的修复效果,但是存在较大的音频失真。

上述削波修复方法的性能易受削波样点数目的影响,削波样点数越多,其修复效果越差。为了解决这个问题,本文在 MDCT 域提出了一种基于核 Fisher 判别(Kernel Fisher discriminant, KFD)

和加权码书映射(Weighted codebook mapping, WCBM)的削波修复方法,该方法首先利用 KFD,判别检测削波的发生,其次利用 WCBM 修复削波。该方法有效提高了削波修复的性能。

1 本文削波修复方法

实验发现,削波不仅影响音频信号的时域波形,而且对其频谱也有严重影响。图 2 为一段音频信号的原始音频和削波音频时频分析对比图。

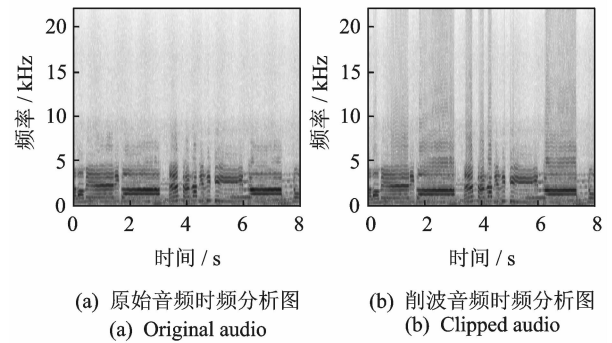


图 2 原始音频与削波音频时频分析对比图

Fig. 2 Spectrogram comparison between original and clipped audios

从图 2 可以看出,削波音频时频分析图的中、高频部分引入了多余的频谱成分,这些成分使得原始音频信号出现扰人的声音,严重影响原始音频信号的听觉质量。此外,本文还可以看出削波只影响部分音频帧的频谱,所以在进行削波修复前,首先需要检测音频信号是否存在削波,然后再对其进行修复处理。图 3 给出了本文提出的削波修复流程。

本文提出的削波修复方法包括削波检测和削

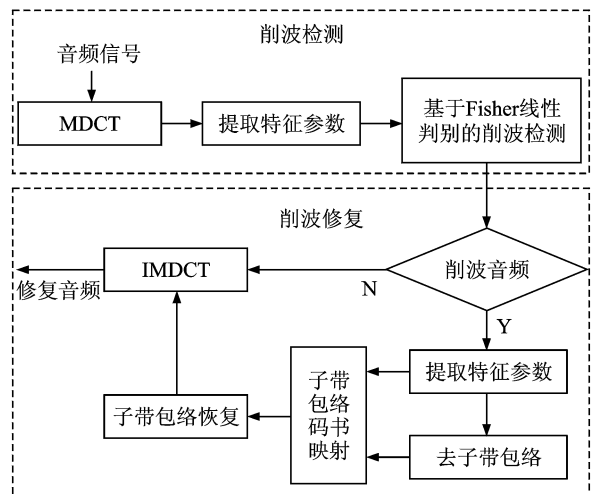


图 3 本文削波修复方法流程图

Fig. 3 Flow chart of proposed restoration method

波修复两个模块。首先对一帧输入音频信号进行 MDCT 变换,并根据得到的 MDCT 系数进行特征参数的提取;然后利用核 Fisher 非线性判别方法^[7-9]检测削波帧,如果当前帧为削波音频帧,则利用子带包络 WCBM 修复削波,得到修复的 MDCT 系数;如果当前帧为非削波音频帧,则直接输出 MDCT 系数;最后将输出的 MDCT 系数进行 MDCT 逆变换得到修复音频的时域信号。

1.1 基于核 Fisher 判别的削波检测方法

对于一帧音频信号的削波检测问题,可以把他看作是一种只有两个类别的分类问题,一类是削波音频帧,表示为 CLIP 类,另一类是非削波音频帧,表示为 UNCLIP 类。由于核 Fisher 判别是一种很好的两类问题分类方法,因此,本文结合核 Fisher 判别提出了一种 MDCT 域的削波检测方法,其基本原理如图 4 所示。

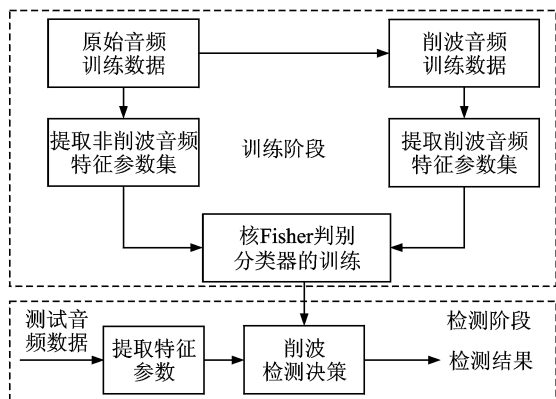


图 4 削波检测原理

Fig. 4 Block diagram of clipping detection method

削波检测包含分类器训练和削波检测两个阶段。在训练阶段,首先将原始音频训练数据集进行削波处理,得到削波音频训练数据集;其次对该训练数据集进行 MDCT 变换,并根据得到的 MDCT 系数按帧提取削波音频和非削波音频的特征参数;最后利用这些参数进行核 Fisher 分类器的训练,确定核 Fisher 分类器的最优投影向量和分类阈值。

在检测阶段,首先对测试音频数据进行 MDCT 变换;然后在 MDCT 域提取特征参数;最后利用训练好的核 Fisher 最优投影向量和分类阈值进行削波的分类决策,检测出削波类和非削波类,从而完成削波检测。

1.1.1 特征参数的提取与选择

特征参数提取和选择的基本任务就是从许多

特征中找出那些最有效的区别性特征。本文从传统听觉感知特征^[10]和 MPEG-7^[11-12]音色特征来考虑削波特征参数的选取,主要包括归一化自相关函数系数、MDCT 系数过零率、梯度指数、谱质心、谱重心、子带均方根能量、子带通量、音频扩展度、谱熵、MDCT 系数幅度方差和 MDCT 系数幅度均值参数。通过统计直方图分析发现,子带均方根能量、子带通量、MDCT 系数幅度方差和 MDCT 系数幅度均值特征能有效区分削波,其他参数的区分性较小,从而可以得到各个特征参数用于削波检测的能力,如表 1 所示。图 5 给出了具有削波检测的能力的特征参数的统计直方图。

表 1 削波特征参数的检测性能

Table 1 Feature selection for clipping detection

特征名称	能否检测削波	特征名称	能否检测削波
过零率	No	归一化自相关系数	No
梯度指数	No	子带均方根能量	Yes
谱质心	No	子带通量	Yes
音频谱重心	No	幅度均值	Yes
音频扩展度	No	幅度方差	Yes
谱熵	No		

从表 1 和图 5 可以看出,本文选取的子带均方根能量、子带通量、MDCT 系数幅度均值和 MDCT

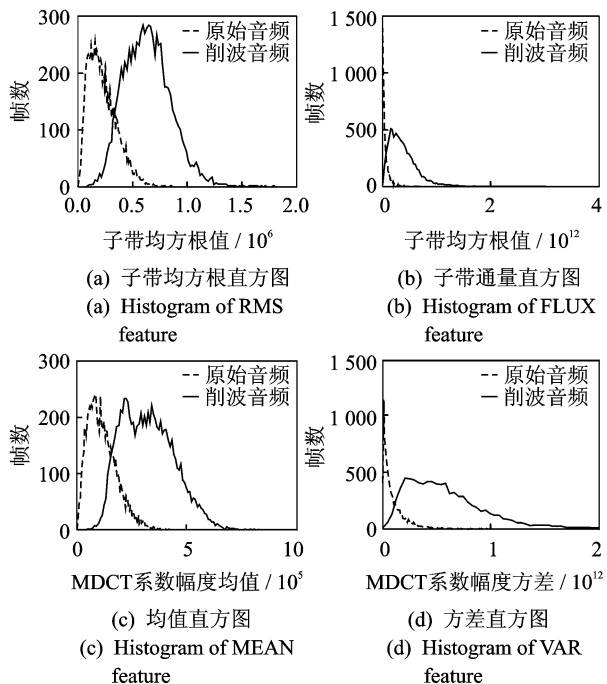


图 5 各特征参数统计直方图

Fig. 5 Statistical histograms of the most discriminable features

系数幅度方差特征参数具有很好的削波辨别能力。

下面将基于一帧音频 1 024 点的 MDCT, 介绍上述 4 个特征参数^[11-12]的具体提取方法。

(1) 子带均方根能量 $F_{\text{rms}}(b)$: 将一帧音频信号划分 $N_{\text{subband}} = 16$ 个子带, 每个子带的频点数为 $L_{\text{subband}} = 64$, 然后分别计算各子带的均方根能量 $F_{\text{rms}}(b)$ 为

$$F_{\text{rms}}(b) = \sqrt{\frac{\sum_{k=B_{\text{low}}(b)}^{B_{\text{up}}(b)} |Y(k)|^2}{L_{\text{subband}}}} \quad (1)$$

式中: b 为子带索引; k 为 MDCT 系数频点索引; $Y(k)$ 为削波音频信号的 MDCT 系数; $B_{\text{up}}(b)$ 和 $B_{\text{low}}(b)$ 分别为第 b 子带对应的频率上界和下界。

(2) 子带通量 F_{flux} : 表示相邻子带能量局部变化的总和, 其计算式如下

$$F_{\text{flux}} = \sum_{b=1}^{N_{\text{subband}}-1} |F_{\text{rms}}(b) - F_{\text{rms}}(b-1)|^2 \quad (2)$$

(3) MDCT 系数幅度均值 F_{mean} : 根据一帧音频信号的 MDCT 系数, 得

$$F_{\text{mean}} = \frac{1}{N} \sum_{k=0}^{N-1} |Y(k)| \quad (3)$$

式中: N 为帧长, 即一帧音频的 MDCT 系数的个数。

(4) MDCT 系数幅度方差 F_{var} : 根据一帧音频信号的 MDCT 系数, 得

$$F_{\text{var}} = \sum_{k=0}^{N-1} (|Y(k)| - F_{\text{mean}})^2 \quad (4)$$

这样, 本文提取了 $N_{\text{subband}} = 16$ 维的子带均方根能量、1 维子带通量、1 维 MDCT 系数幅度均值、1 维 MDCT 系数幅度方差特征, 构成 19 维的特征参数向量, 用于后面的核 Fisher 分类器的训练、削波检测以及修复处理。

1.1.2 核 Fisher 判别^[7-9]

KFD 就是将 Fisher 线性判别方法与核函数相结合实现一种非线性判别。它的核心思想是通过一个变换函数 Φ , 将原始特征空间 \mathbf{R}^n 中的非线性特征向量 \mathbf{x} 变换到一个高维特征空间 \mathbf{H} 中, 得到一个新的线性特征向量 $\Phi(\mathbf{x})$, 然后在新的高维特征空间 \mathbf{H} 中使用 Fisher 进行线性判别。也就是说, 它使原始特征空间 \mathbf{R}^n 中的非线性不可分问题转化为高维特征空间 \mathbf{H} 中的线性可分问题。核 Fisher 非线性判别分析的步骤如下。

(1) 核函数

变换函数 $\Phi(\mathbf{x})$ 通常利用满足 Mercer 条件^[7]

的核函数 $K(\mathbf{x}, \mathbf{y})$ 来完成。核函数 $K(\mathbf{x}, \mathbf{y})$ 是一种 $\Phi(\mathbf{x})$ 的内积运算, 即对于所有的原始特征向量 $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, 满足 $K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y})$ 。目前常用的核函数有线性核函数、多项式核函数、双曲正切核函数、径向基(Radial basis function, RBF)核函数等。其中, RBF 核函数最为常用。另外, 通过实验发现, 采用 RBF 核函数获得的削波修复性能优于其他几种核函数, 因此, 本文采用 RBF 核函数, 其计算式为^[7]

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad \sigma > 0 \quad (5)$$

式中: $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ 为原始特征向量; σ 为尺度参数, 控制 RBF 核函数的性能, 本文取 σ 为 1.0。

(2) 核 Fisher 最优投影方向

令 \mathbf{X} 为包含两类样本集 ψ 中的某个样本矢量, 样本集 ψ 由 N 个样本组成, 其中, 削波类 ψ_1 的样本数为 N_1 , 非削波类 ψ_2 的样本数为 N_2 , 每个样本 \mathbf{X} 为前一小节提取的 $d = 19$ 维削波特征向量。通过变换函数 Φ , 将两类样本向量 \mathbf{X} 变换到高维特征空间 H , 得到新的向量 $\Phi(\mathbf{X})$, 则 H 空间中两类样本向量的均值分别为

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} \Phi(\mathbf{X}_n^i) \quad i = 1, 2 \quad (6)$$

式中: n 为样本索引; \mathbf{X}_n^i 表示第 i 类的第 n 个样本。

设样本向量 \mathbf{X} 在 H 空间中的投影向量为 \mathbf{W} , 将 $\Phi(\mathbf{X})$ 投影到 \mathbf{W} , 得到投影样本 \mathbf{y}_x , 即 $\mathbf{y}_x = \mathbf{W}^T \Phi(\mathbf{X})$, 这样, 可以得到类 ψ_i 的投影样本 \mathbf{y}_x 的类内均值 μ_i 和类内方差 v_i^2 为

$$\mu_i = \frac{1}{N_i} \sum_{\mathbf{x} \in \psi_i} \mathbf{y}_x = \frac{1}{N_i} \sum_{\mathbf{x} \in \psi_i} \mathbf{W}^T \Phi(\mathbf{X}) = \mathbf{W}^T \mathbf{m}_i \quad i = 1, 2 \quad (7)$$

$$v_i^2 = \sum_{\mathbf{x} \in \psi_i} (\mathbf{y}_x - \mu_i)(\mathbf{y}_x - \mu_i)^T \quad i = 1, 2 \quad (8)$$

为了区分两类样本, KFD 需使两类样本的类内均值之差最大, 并使两类样本的类内方差最小。因此, 定义 KFD 准则函数为

$$\max J(\mathbf{W}) = \frac{(\mu_1 - \mu_2)^2}{v_1^2 + v_2^2} \quad (9)$$

将式(7,8)代入式(9), 得

$$\max J(\mathbf{W}) = \frac{\mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\mathbf{W}^T \mathbf{S}_w \mathbf{W}} \quad (10)$$

式中: \mathbf{S}_b 为投影样本类间离散度矩阵, \mathbf{S}_w 为投影样本类内离散度矩阵, 即

$$\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \quad (11)$$

$$S_w = \sum_{i=1}^2 \sum_{X \in \Psi_i} (\Phi(X) - m_i)(\Phi(X) - m_i)^T \quad (12)$$

S_b, S_w 和 $J(W)$ 的计算依赖于变换函数 Φ 的获取,根据再生核理论^[7],在高维线性特征空间 H 中,任何一个向量 $W \in H$ 都可表示为

$$W = \sum_{j=1}^N \alpha_j \Phi(X_j) \quad (13)$$

根据式(6,13),可得

$$W^T m_i = \sum_{j=1}^N \alpha_j \Phi(X_j)^T \cdot \frac{1}{N_i} \sum_{n=1}^{N_i} \Phi(X_n^i) = \frac{1}{N_i} \sum_{j=1}^N \sum_{n=1}^{N_i} \alpha_j K(X_j, X_n^i) = \alpha^T M_i \quad (14)$$

式中: n 为第 i 类样本矢量的索引; $M_i = (1/N_i) \cdot \sum_{n=1}^{N_i} K(X_j, X_n^i)$ 为样本集 ψ 中的第 j 个样本矢量 X_j 与第 i 类样本中各矢量内积核的均值; α_j 为向量 α 的第 j 个元素。

根据式(11,14),式(10)的分子可表示为

$$W^T S_b W = \alpha^T M \alpha \quad (15)$$

式中 $M = (M_1 - M_2)(M_1 - M_2)^T$ 。

根据式(12,14),式(10)的分母可表示为

$$W^T S_w W = \alpha^T F \alpha \quad (16)$$

式中: $F = \sum_{i=1}^2 K_i(I_i - L_i)K_i^T$, K_i ($i=1,2$) 为第 i 类的核矩阵,即 $(K_i)_{n,j} = K(X_j, X_n^i)$, $j \in [1 \dots N]$, $n \in [1 \dots N_i]$, I_i 为 N_i 阶单位矩阵, L_i 为所有元素为 $1/N_i$ 的 N_i 阶方阵。

将式(15,16)代入式(10),可得

$$\max J(\alpha) = \frac{\alpha^T M \alpha}{\alpha^T F \alpha} \quad (17)$$

利用拉格朗日乘法求式(17)的最大值,可得到高维线性特征空间 H 中的最优投影向量为

$$\alpha = F^{-1}(M_1 - M_2) \quad (18)$$

在实验中,为了防止 F 非正定,通常引入一个正则化参数 $\lambda=0.001$,即用 $F_\lambda = F + \lambda I$ 代替 F ,其中 I 为 N 阶单位矩阵。

(3) 分类阈值 y_0

利用式(18)得到的最优投影向量 α ,将原始特征向量 X 投影到 α 上,则投影样本值 y_x 为

$$y_x = \sum_{j=1}^N \alpha_j \Phi(X_j)^T \Phi(X) = \sum_{j=1}^N \alpha_j K(X_j, X) \quad (19)$$

这样,前面提及的 19 维样本矢量的分类问题转化成了一维分类问题,因此只需要确定一个分类阈值 y_0 ,将投影样本 y_x 与 y_0 相比较,就可进行分类决

策。本文利用先验知识来选定一维空间中的分类阈值 y_0 ,即它可以用两类样本投影到向量 α 上的投影样本的均值来表示

$$y_0 = \frac{\rho N_1 \mu_1 + \nu N_2 \mu_2}{N_1 + N_2} \quad (20)$$

式中: ρ, ν 为经验常数,用于控制分类阈值的选定,满足 $\rho + \nu = 2$ 。 μ_i ($i=1,2$) 表示将 ψ_i 类的样本矢量投影到向量 α 上得到的投影样本 y_x 的均值

$$\mu_i = \frac{1}{N_i} \sum_{X \in \psi_i} y_x \quad (21)$$

(4) 分类决策规则

对于任意待分类的未知特征样本向量 X ,根据式(19)得到其投影样本值 y_x ,这样,结合分类阈值 y_0 ,得到 KFD 的决策规则为

$$X \in \begin{cases} \Psi_1 & y_x \geq y_0 \\ \Psi_2 & y_x < y_0 \end{cases} \quad (22)$$

1.1.3 分类器的训练

本文选用 44.1 kHz 采样的音频数据作为非削波类训练数据集,然后将这些音频进行削波处理,得到削波类训练数据集,同时把发生削波的帧索引号保存到削波帧索引文件中,用于提取削波音频的特征参数。根据削波和非削波训练数据集,本文可得 KFD 分类器的训练流程如图 6 所示。

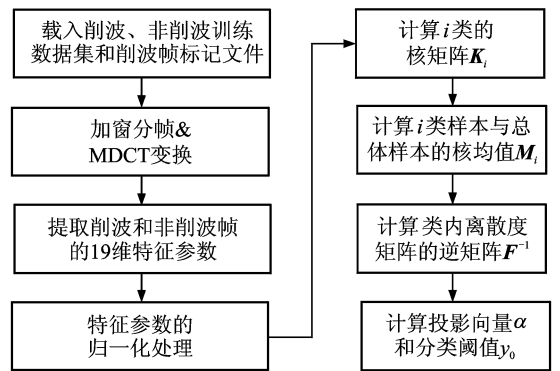


图 6 核 Fisher 分类器训练流程图

Fig. 6 Training flow chart of kernel Fisher analysis

从图 6 可知 KFD 分类器的训练流程为:首先,将削波和非削波训练音频数据分别加窗分帧并进行 MDCT 变换,帧长为 1 024 个样点(约为 23 ms);其次,根据削波帧索引文件,利用式(1~4)提取削波帧和非削波帧的子带均方根能量、子带通量、MDCT 系数幅度均值和 MDCT 系数幅度方差四类特征参数,这样得到削波音频和非削波音频的 19 维特征参数集;然后,将这些特征参数集进行归一化处理;最后,训练核 Fisher 分类器,确定高维线

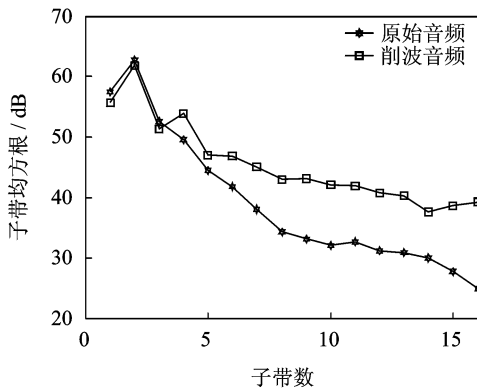
性特征空间 H 中的最优投影向量 α 和分类阈值 y_0 。

1.1.4 削波检测

首先,将待检测音频进行加窗分帧并进行 MDCT 变换;其次,提取一帧音频信号的子带均方根、子带通量、MDCT 系数幅度均值和 MDCT 系数幅度方差特征参数向量 \mathbf{X} ,并对其进行归一化处理;然后,利用式(19)计算其在核 Fisher 分类器的最优投影向量 α 上的投影值 y_x ;最后,根据式(22)进行削波检测。

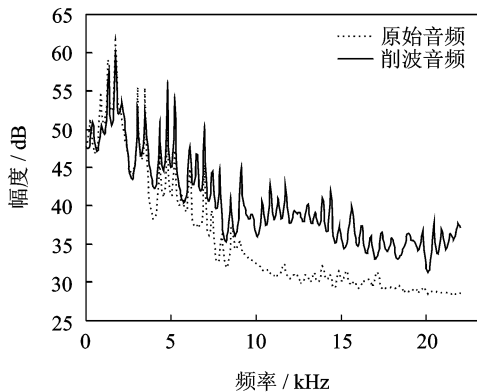
1.2 基于加权码书映射的削波修复方法

本文实验结果表明,削波在音频时频分析图的中、高频处引入了多余的频谱成分,使其谱包络得到提升。图 7 给出了原始音频和削波音频的包络对比图示例。



(a) 子带包络对比

(a) Comparison of sub-band envelope



(b) LPA 谱包络^[13]对比

(b) Comparison of LPA spectrum envelope

图 7 原始音频与削波音频包络对比示例

Fig. 7 Envelope comparison between original and clipped audios

从图 7 中可以看出,削波音频的两种包络在中、高频处要明显高于原始音频谱包络。也就是说,对于削波的修复,可以通过处理削波音频的子

带包络来实现,因此,本文提出一种基于子带包络的 WCBM 方法来修复削波。

首先,根据一帧音频信号的 MDCT 系数计算子带均方根能量(即子带包络)、子带通量、MDCT 系数幅度均值和方差 4 个特征参数;然后,对当前帧音频信号的 MDCT 系数进行去子带包络处理,得到去包络的 MDCT 系数;接着,利用 WCBM 方法得到修复音频的子带包络;最后,根据该子带包络结合去包络的音频信号的 MDCT 系数,进行子带包络的恢复处理,得到修复音频信号的 MDCT 系数,从而实现削波的修复。

1.2.1 去子带包络处理

首先根据式(1)计算削波帧的子带包络 $F_{\text{clip-rms}}(b)$,其中, $b \in [0, N_{\text{subband}})$ 为子带索引;然后利用 $F_{\text{clip-rms}}(b)$ 对削波音频信号的 MDCT 系数 $Y(b, k)$ 进行去子带包络处理,得到去包络的 MDCT 系数 $\tilde{Y}(b, k)$

$$\tilde{Y}(b, k) = \frac{Y(b, k)}{F_{\text{clip-rms}}(b)} \quad \forall k \in [0, N), b = \left\lfloor \frac{k}{L_{\text{subband}}} \right\rfloor \quad (23)$$

式中: k 表示 MDCT 系数索引号; $N = 1024$ 为 MDCT 系数的个数, $N_{\text{subband}} = 16$ 为子带个数; $L_{\text{subband}} = 64$ 为每个子带的频点数。

1.2.2 子带包络加权码书映射

本文采用 WCBM 方法来修复子带包络,其基本原理如图 8 所示^[14]。该方法包含削波特征参数和子带包络 2 个码书,它们的码字是一一对应的。削波特征参数码矢量为 \mathbf{F}_{clip} ,它是由削波帧的 19 维特征参数构成;子带包络码矢量为 \mathbf{F}_{rms} ,它是由原始音频的 16 维子带包络构成。削波特征参数和子带包络码书由 LBG 算法^[13]离线训练得到,码书大小为 1024。

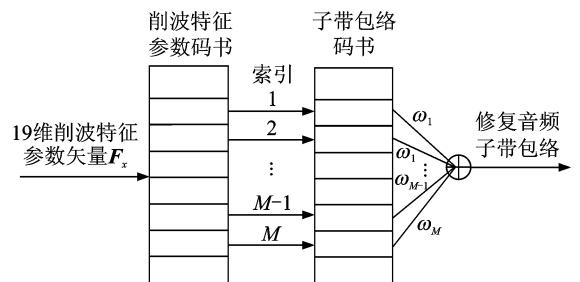


图 8 加权码书映射原理

Fig. 8 Block diagram of WCBM

WCBM 的具体步骤如下:

(1)对于输入的每一帧削波音频信号,首先在

MDCT 域提取 19 维削波特征参数矢量,这里定义为 \mathbf{F}_x 。

(2) 根据欧式距离测度,计算削波特征参数矢量 \mathbf{F}_x 和削波特征参数码书各个码矢量 \mathbf{F}_{clip} 间的距离,并选出 M 个欧式距离最小的码矢量,将其欧式距离分别记作 $\{d_1, d_2, \dots, d_M\}$,并且这 M 个码矢量对应的子带包络码矢量分别记作 $\{\mathbf{F}_{\text{rms}}(1), \mathbf{F}_{\text{rms}}(2), \dots, \mathbf{F}_{\text{rms}}(M)\}$ 。

(3) 根据上述步骤得到的 M 个欧式距离 $\{d_1, d_2, \dots, d_M\}$,计算削波特征参数矢量 \mathbf{F}_x 与 M 个码矢量之间的隶属度^[14]为

$$\rho_i = \left\{ \sum_{j=1}^M \left(\frac{d_i}{d_j} \right)^2 \right\}^{-1} \quad \forall i \in [1 \ M] \quad (24)$$

本文 M 取为 3。

(4) 利用隶属度 ρ_i 计算 M 个子带包络码矢量 $\mathbf{F}_{\text{rms}}(i)$ 的权值 ω_i 为

$$\omega_i = \frac{\rho_i}{\sum_{j=1}^M \rho_j} \quad \forall i \in [1 \ M] \quad (25)$$

(5) 利用权值 ω_i 对 M 个子带包络码矢量 $\mathbf{F}_{\text{rms}}(i)$ 进行加权求和,得到修复音频信号子带包络矢量 $\tilde{\mathbf{F}}_{\text{rms}}$

$$\tilde{\mathbf{F}}_{\text{rms}} = \sum_{i=1}^M \omega_i \cdot \mathbf{F}_{\text{rms}}(i) \quad \forall i \in [1 \ M] \quad (26)$$

1.2.3 子带包络修复

利用式(26)得到的修复音频信号子带包络矢量 $\tilde{\mathbf{F}}_{\text{rms}}$,结合式(23)得到的去包络的 MDCT 系数 $\tilde{Y}(b, k)$ 来修复削波,得到修复音频信号的 MDCT 系数 $\hat{Y}(b, k)$ 为

$$\hat{Y}(b, k) = \tilde{Y}(b, k) \cdot \tilde{\mathbf{F}}_{\text{rms}}(b) \quad \forall k \in [0 \ N] \quad (27)$$

式中: $b \in [0, N_{\text{subband}})$ 为子带索引; $N_{\text{subband}} = 16$ 为子带数目; k 为 MDCT 系数索引; $N = 1\ 024$ 为 MDCT 系数个数。

2 实验结果及分析

为了验证本文所提方法的性能,本文将从削波检测和削波修复两个方面分别进行测试。

在测试实验中,测试音频数据取自于 MPEG 标准音频测试数据库,采样率为 44.1 kHz。它们不同于训练数据集,是训练数据集之外的音频数据,共计 19 段不同风格的音频数据,包括大提琴、流行音乐、鼓乐、交响乐和混合乐等。将这些测试数据集进行削波处理,得到削波测试数据集。本文的削波比例分别为其幅度的 30%、40% 和 50%,即对音频信号的幅度削去其幅度最大值的 30%,

40% 和 50%。

2.1 削波检测性能测试

为了验证本文方法对削波检测的有效性,本文采用虚警率 R_{FAR} 、漏检率 R_{MDR} 以及准确率 R_{RIGHT} 来衡量,其定义式分别为^[6]

$$R_{\text{FAR}} = \frac{N_{\text{false}}}{N_{\text{nonclip}}} \times 100\% \quad (28)$$

$$R_{\text{MDR}} = \frac{N_{\text{miss}}}{N_{\text{clip}}} \times 100\% \quad (29)$$

$$R_{\text{RIGHT}} = \left(1 - \frac{N_{\text{false}} + N_{\text{miss}}}{N_{\text{total}}} \right) \times 100\% \quad (30)$$

式中: N_{false} 表示未削波类被判为削波类的帧数目; N_{nonclip} 表示实际未削波类帧数目; N_{miss} 表示削波类被判为未削波类的帧数目; N_{clip} 表示实际 CLIP 类削波类帧数目; N_{total} 表示总的帧数目,满足 $N_{\text{total}} = N_{\text{clip}} + N_{\text{nonclip}}$ 。

根据 1.1.4 节介绍的削波检测方法,分别对 3 种不同削波比例的测试音频信号进行削波检测测试,并统计不同音频类别的检测准确率、虚警率和漏检率,统计结果如表 2 所示。

从表 2 可以看出,本文算法在 3 种不同削波比例下,对不同类别音频数据的削波检测准确率都达到 90% 以上。虽然有些音频类别的虚警率或漏检率较大,但也都低于 10%。另外,从表中还可以看

表 2 削波检测结果

Table 2 Clipping detection test results		%		
音频类别	削波比例	准确率	虚警率	漏检率
大提琴类	30	92.78	9.75	4.78
	40	92.86	6.92	8.93
	50	93.61	8.17	5.26
流行音乐类	30	91.81	8.37	5.28
	40	92.52	8.42	5.18
	50	92.67	8.06	6.82
古乐类	30	91.64	9.94	7.50
	40	92.42	8.06	7.18
	50	95.21	8.05	4.22
交响乐类	30	90.91	7.59	4.24
	40	92.95	6.32	7.64
	50	93.60	5.73	6.55
混合乐类	30	90.69	7.55	8.35
	40	91.87	7.66	7.85
	50	93.12	6.36	7.15

出,随着削波比例的增大,检测准确率会增加,且虚警率和漏检率都有所减小。这是因为随着削波比例的增大,削波失真对音频信号产生的影响也会增大,即会引入更多的多余频谱成分,这样就会造成削波与非削波帧的特征差别增大,从而使得检测准确率提高,虚警率和漏检率减小。

传统的削波检测方法都是在时域进行^[5-6],检测音频信号的时域波形峰值,通过峰值来确定是否出现削波。对于数字削波,这些方法的检测准确率可以达到 100%。而对于模拟削波,由于时域波形峰值存在微小抖动,所以时域峰值检测方法的检测准确率会大大下降^[5-6],并且其检测准确率与抖动大小直接相关,抖动越大,检测准确率越低。但是,本文提出了一种频域削波检测方法,其不依赖音频信号的时域波形,而是在 MDCT 域检测音频信号是否出现削波,也就是说,本文方法对数字削波和模拟削波会具有相同的检测效果。因此,本文提出的削波检测方法更具有通用性。

2.2 削波修复性能测试

在测试实验中,参考方法为三次样条插值方法^[4]、软阈值和谱加权函数相结合的方法^[6]。为了简化起见,这两种参考算法分别记为三次插值法和软阈值法。本文分别从音频时域波形及时频分析、谱包络对比分析、对数谱失真(Log-spectral distortion, LSD)^[15]和时域分段信噪比(SNR_{seg})4 个方面与参考算法进行修复性能比较。

2.2.1 音频时域波形及时频分析

图 9 给出了原始音频信号、削波音频信号、本文算法和两种参考算法的削波修复音频信号的时域波形对比图。从图中可以看出,软阈值法没有明显修复削波音频的时域波形;三次插值法与本文方法能较好的修复时域波形。但是,本文方法修复音频的时域波形与原始音频最为相似,从而可以说明本文方法在修复时域波形方面优于两种参考算法。

为了进一步说明本文方法的性能优于参考算法,图 10 给出了图 9 中各音频信号对应的音频时频分析图。

通过对比图 10 的时频分析图可以发现,削波音频频谱因削波引入了大量的多余成分;软阈值修复方法修复音频频谱损失较大,且其削波修复效果较差;三次插值法对削波虽然具有一定的修复效果,但是其修复音频频谱仍然残留不少的多余成分;本文方法能很好地消除削波引入的多余频谱成分,并且音频频谱的失真较小。因此,从时频分析图上也可

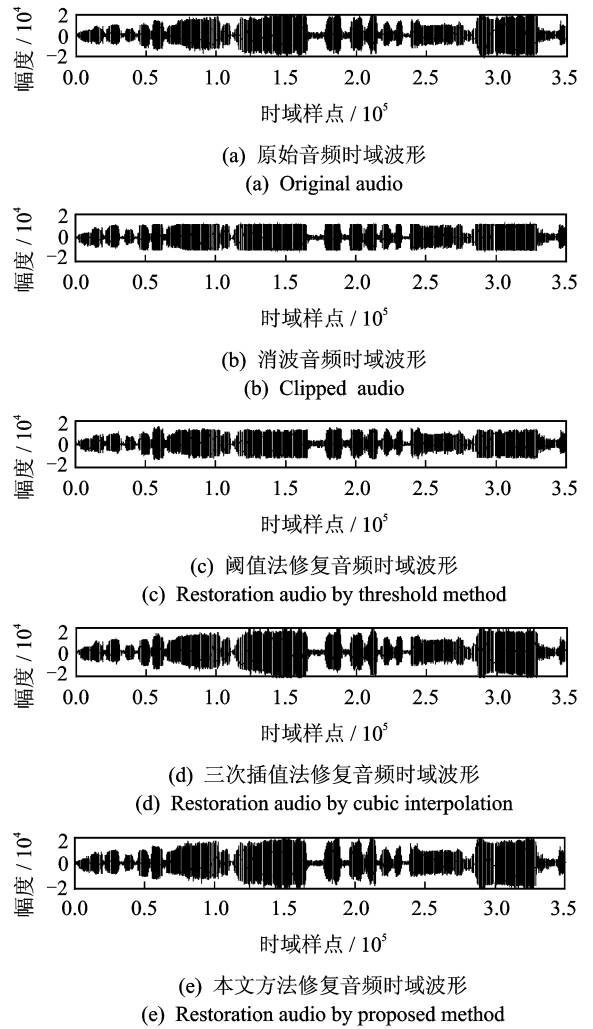


图 9 音频信号时域波形对比图

Fig. 9 Waveform comparison between original and clipping restoration audios

以说明本文方法的修复性能优于两种参考算法。

2.2.2 谱包络对比分析

图 11,12 分别给出了原始音频、削波音频以及本文算法和两种参考算法的削波修复音频的子带包络和 LPC 谱包络对比图,测试音频片段与图 7 相同。

从图 11,12 可以看出,两种参考算法在一定程度上降低了削波音频中、高频处的谱包络,但是与原始音频的谱包络还具有较大的偏差。而本文算法修复音频的子带包络和 LPC 谱包络都能较好跟踪原始音频的包络趋势,偏差较小,即更好地消除了削波导致的扰人声音。因此,可以说明本文方法的性能优于两种参考算法。

2.2.3 对数谱失真测试

本文采用 LSD 来衡量削波修复算法修复后音

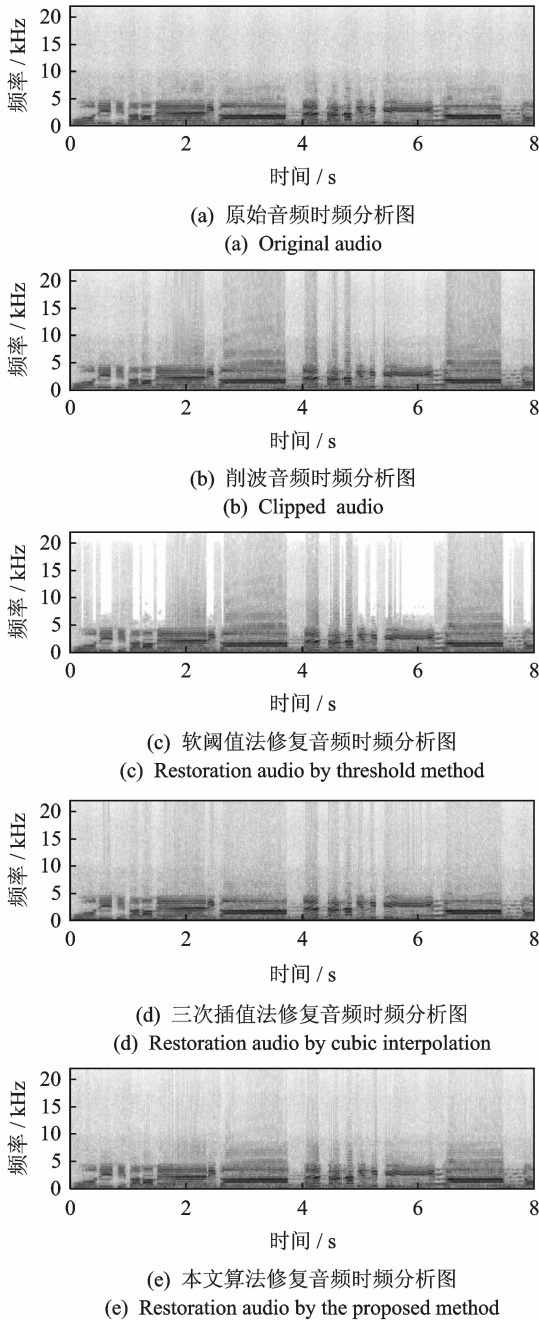


图 10 音频时频分析对比图

Fig. 10 Spectrogram comparison between original and clipping restoration audios

频的频谱相似性,其计算公式如下^[15]

$$d_{LSD} = \frac{1}{L} \sum_{l=0}^{L-1} \sqrt{\frac{1}{N} \sum_{k=0}^{N-1} \left[10 \log_{10} \frac{|\hat{X}(l,k)|^2}{|X(l,k)|^2} \right]^2} \quad (31)$$

式中: l 为帧索引; k 为频点索引; L 为总帧数; N 为 FFT 长度; $X(l,k)$ 和 $\hat{X}(l,k)$ 分别为原始音频和修复后音频第 l 帧、第 k 频点的傅里叶变换系数。

根据各削波修复算法得到的修复音频,在不同削波比例下分别进行 LSD 测试,测试结果如表 3

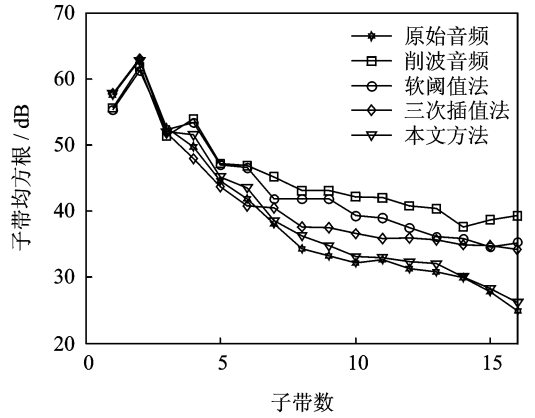


图 11 子带包络对比图

Fig. 11 Sub-band envelope comparison between kinds of audios

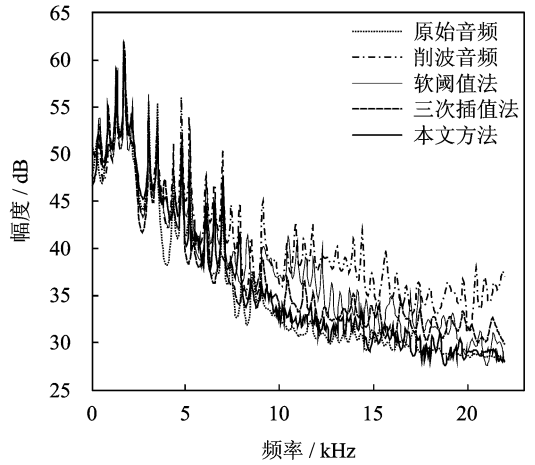


图 12 LPA 谱包络对比图

Fig. 12 LPA comparison between kinds of audios

所示。

从表 3 可以看出,软阈值法和三次插值法的 LSD 失真高于本文所提方法的 LSD 失真,即本文方法修复音频的 LSD 失真最小,从而进一步说明本文方法修复削波的有效性,其性能优于参考算法。

2.2.4 时域分段信噪比测试

本文采用时域分段信噪比来衡量不同算法得到的修复音频在时域波形上的差异程度。分段信噪比反应了音频序列帧信噪比的几何平均,定义为

$$SNR_{seg} = \frac{1}{L} \sum_{l=0}^{L-1} 10 \cdot \log_{10} \left(\frac{\sum_{n=Nl}^{Nl+N-1} x^2(n)}{\sum_{n=Nl}^{Nl+N-1} [x(n) - y(n)]^2} \right) \quad (32)$$

表 3 LSD 失真测试结果比较

Table 3 Test results of LSD

音频类别	削波比例/%	软阈值法	三次插值法	本文方法
大提琴类	30	21.552	10.785	2.216
	40	21.610	10.818	2.417
	50	21.637	10.858	2.779
流行音乐类	30	17.933	7.239	3.237
	40	17.708	7.469	3.418
	50	17.360	7.929	3.778
鼓乐类	30	24.032	8.599	2.132
	40	23.978	8.634	2.218
	50	23.935	8.681	2.383
交响乐类	30	13.670	7.356	2.335
	40	13.590	7.884	3.440
	50	13.380	8.549	3.719
混合乐类	30	17.127	7.908	2.069
	40	17.145	8.188	2.161
	50	16.974	8.686	2.444

式中： L 为测试音频信号的总帧数； N 为帧长； $x(n)$ 表示原始时域音频信号； $y(n)$ 表示削波时域音频信号或修复后时域音频信号。表 4 给出了削波前后分段信噪比提高的结果。

表 4 分段信噪比提高测试结果比较

Table 4 Test results of segmental SNR

音频类别	削波比例/%	SNR _{seg} /dB		
		软阈值法	三次插值法	本文方法
大提琴类	30	0.126	2.286	3.203
	40	0.375	3.879	5.180
	50	0.330	4.671	6.417
流行音乐类	30	-0.310	1.903	4.964
	40	-0.714	3.481	6.579
	50	-0.945	3.832	6.803
鼓乐类	30	-0.341	1.262	3.031
	40	-0.497	1.902	5.053
	50	-0.781	2.825	7.203
交响乐类	30	-0.271	2.439	5.329
	40	-0.335	3.621	5.822
	50	-0.227	3.431	8.645
混合乐类	30	-0.013	0.557	5.172
	40	-0.010	0.983	6.596
	50	0.045	2.169	6.021

从表 4 可以知道,本文算法的分段信噪比提高程度明显好于两种参考算法,从而再次说明本文算法比参考算法具有更好的削波修复性能。

2.2.5 计算复杂度分析

本文的计算量主要集中在特征提取以及加权码书映射两部分,而特征提取的计算复杂度为 N (帧长)次加法操作,加权码书映射的计算复杂度为码书大小与码书向量维数的乘积。因此,本文的计算复杂度较小,实时性较好,可以应用于实时性要求较高的场合。

3 结束语

对于削波的检测和修复,传统的方法都是在时域进行,本文在 MDCT 域提出了一种新的削波检测与修复方法。首先是基于 KFD 提出一种 MDCT 域的削波检测方法。该方法根据音频信号的 MDCT 系数提取子带包络等削波特征参数,然后利用这些特征参数训练核 Fisher 分类器,用于检测削波;最后根据检测结果,提出一种基于子带包络的 WCBM 方法来修复削波,得到修复音频信号的 MDCT 系数,并将其变换到时域,实现削波的修复。测试结果表明,本文所提方法能有效修复音频信号中的削波(削波比例分别为 30%,40%和 50%),并且其性能优于参考修复方法。

参考文献:

- [1] Godsill S J, Rayner p J W. Digital audio restoration — A statistical model-based approach[M]. London: Springer-Verlag, 1998:191-204.
- [2] ABEL J S, Smith J O. Restoring a clipped signal [C]//ICASSP. [S.l.]: IEEE, 1991:1745-1748.
- [3] Olofsson T. Deconvolution and model-based restoration of clipped ultrasonic signals[J]. IEEE Trans on Instrumentation and Measurement, 2005, 54 (3): 1235-1240.
- [4] Audacity. Cubic interpolation method for declipping in clipfix plug in [EB/OL]. <http://www.gacrecords.org.uk/audacity.html>, 2013.
- [5] 刘春花,付强,杨家玮,等. 音频信号截幅失真的检测与修复[J]. 声学技术, 2009, 28(4): 498-502.
Liu Chunhua, Fu Qiang, Yang Jiawei, et al. The clipping detection and restoration for audio signal[J]. Technical Acoustics, 2009, 28(4): 498-502.
- [6] Zhang D W, Bao C C, Deng F. et al, A restoration method of the clipped audio signals based on MDCT [C]//ISSPIT 2011. Bilbao, Spain; IEEE, 2011:253-

- 257.
- [7] Li Z G, Wang F L, Zhu W Z. An optimal kernel Fisher nonlinear discriminant analysis method and applied on face recognition[C] //CIS 2008. Suzhou, China: IEEE, 2008:233-237.
- [8] Zhou X, Wu Y. Application of kernel fisher discriminant analysis to digital signal classification[J]. Journal of Beijing University of Posts and Telecommunications, 2011,34(2):35-39.
- [9] Zhu J, Hoi S C H, Lyu M R. Face annotation using transductive kernel Fisher discriminant [J]. IEEE Trans on Multimedia, 2008,10(1):86-96.
- [10] Larsen E R, Aarts R M. Audio bandwidth extension—Application of psychoacoustics, signal processing and loudspeaker design[M]. UK: John Wiley & Sons Ltd, 2004:145-234.
- [11] Deng J D, Simmermacher C, Cranefield S. A study on feature analysis for musical instrument classification[J]. IEEE Trans on Systems, Man, and Cybernetics, 2008,38(2):429-438.
- [12] MPGE Audio Group. Information technology—Multimedia content description interface—Part 4: Audio [S]. ISO/IEC15938-4, 2001.
- [13] 鲍长春. 数字语音编码原理[M]. 西安:西安电子科技大学出版社,2007:109-153.
Bao Changchun. The principles of digital speech coding[M]. Xi'an: Xidian University Press, 2007:109-153.
- [14] 张勇,胡瑞敏. 基于高斯混合模型的语音频带扩展所发的研究[J]. 声学学报,2009,34(5):471-480.
Zhang Yong, Hu Ruimin. Speech wideband extension based on gaussian mixture model [J]. Acta Acustica, 2009,34(5):471-480.
- [15] Pulakka H, LAaksonen L, Vainio M, et al. Evaluation of an artificial speech bandwidth extension method in three languages [J]. IEEE Trans on Audio, Speech and Language Processing, 2008,16(6):1124-1137.

作者简介:邓峰(1983-),男,博士研究生,研究方向:语音和音频增强,E-mail:dengfeng@emails.bjut.edu.cn;鲍长春(1965-),男,博士,教授,博士生导师,研究方向:语音和音频信号处理;鲍枫(1990-),男,硕士研究生,研究方向:语音和音频增强。