

# 基于 AAC 比特流的音频信号 Hiss 噪声抑制方法

邓峰, 鲍长春, 鲍枫

(北京工业大学 电子信息与控制工程学院 语音与音频信号处理研究室, 北京 100124)

**摘要:** 基于 AAC 比特流, 提出了一种压缩域音频 Hiss 噪声抑制方法。该方法的输入为含噪音频的 AAC 比特流, 输出为增强音频的 AAC 比特流。首先, 利用修正的绝对中值标准差 (MMAD, modified median absolute deviation) 估计 Hiss 噪声, 其次, 利用修正的离散傅里叶变换 (MDFT, modified discrete Fourier transform) 计算听觉掩蔽阈值参数, 最后, 根据参数软阈值方法得到增强的 AAC 比特流, 并用于 AAC 解码器得到最终的增强音频信号。主观和客观测试结果表明, 所提出的方法能有效去除 AAC 解码音频信号中的 Hiss 噪声, 其性能明显优于现有的几种 Hiss 噪声消除方法。

**关键词:** AAC 比特流; Hiss 噪声; 压缩域; 修正的离散傅里叶变换; 噪声抑制

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-436X(2013)05-0020-11

## Hiss noise reduction method used for audio signals based on AAC bit-stream

DENG Feng, BAO Chang-chun, BAO Feng

(Speech and Audio Signal Processing Lab, School of Electronic Information and Control Engineering,  
Beijing University of Technology, Beijing 100124, China)

**Abstract:** A hiss noise reduction method used for audio signals based on AAC bit-stream in compressed domain was proposed. The input was AAC bit-stream derived from noisy audio signals, and the output was the enhanced AAC bit-stream used for decoder. First, Hiss noise was estimated by modified median absolute deviation (MMAD). Second, the auditory masking threshold and the masking parameters were calculated by modified discrete Fourier transform (MDFT). Finally, the parametric soft-thresholding method was employed to obtain the enhanced AAC bit-stream which was used to get the enhanced audio signals at AAC decoder. The objective and subjective quality testing results indicate that the proposed algorithm can effectively remove the Hiss noise derived from AAC bit-stream of audio signals and obviously outperforms the existing hiss noise reduction methods.

**Key words:** AAC bit-stream; Hiss noise; compressed domain; modified discrete Fourier transform; noise reduction

### 1 引言

随着 Internet 技术的不断发展, 人们经常用含 AAC 等编码器的录音设备录制各种场合的音频信号, 并上传到网上与人共享。这些音频经常会混入各种噪声 (如 Hiss 噪声), 使得音频信号的质量较

差, 严重影响其听觉效果。另外, 随着数字化进程的加快, 许多珍贵的老唱片等也采用数字化的存储方式, 然而在这个过程中经常会引入 Hiss 噪声, 严重影响其音质。如何增强这类音频的质量是目前亟待解决的课题。

人们在听音乐时, 经常会发现音频信号中混杂

收稿日期: 2012-08-10; 修回日期: 2013-01-20

基金项目: 北京市教育委员会科技发展计划重点基金资助项目 (KZ201110005005); 国家自然科学基金资助项目 (61072089)

**Foundation Items:** Scientific Research Key Program of Beijing Municipal Commission of Education (KZ201110005005); The National Natural Science Foundation of China(61072089)

着“滋滋”的声音，这就是广泛存在于音频信号中的 Hiss 噪声<sup>[1]</sup>，它严重影响音频信号的听觉质量。Hiss 噪声是一种全带（0~20kHz）加性噪声，其能量分布在信号所有频率成分中。Hiss 噪声主要是包括电子设备正常运行时发出的声音、不同的存储媒介（如磁带等）放音时夹杂的声音以及录音环境中的背景噪声等。

目前，人们对 Hiss 噪声消除的研究比较少，且主要是在线性域内对其进行消除。在文献[1]中，剑桥大学研究人员在 FFT 域提出利用维纳滤波以及谱减算法<sup>[1]</sup>来消除音频信号中的 Hiss 噪声。其中，维纳滤波的去噪效果较差，谱减法虽然能有效地消除音频信号中的 Hiss 噪声，但是会引入大量的“音乐噪声”，从而严重影响增强音频信号的听觉质量。1995 年，DONOHO D L 首先提出了小波阈值去噪方法<sup>[2]</sup>，并随着小波理论不断成熟，小波阈值去噪算法取得快速发展。为了克服传统硬阈值法去噪效果差和软阈值法由于过度光滑而使信号失真的缺点，文献[3,4]提出了 2 种新的软阈值方法<sup>[3,4]</sup>，文献[5]提出了一种自适应的小波阈值法<sup>[5]</sup>，这些算法都能获得较好的去噪效果。文献[6]提出了一种基于小波变换的自适应多阈值 Hiss 噪声抑制方法，能有效地消除珍贵旧唱片中的 Hiss 噪声。此外，由于离散余弦变换（DCT, discrete cosine transformation）具有更强的信号能量集中特性、更高的频谱分辨率和更好的相位保护特性，研究人员将阈值去噪方法引入到 DCT 域，提出了一系列阈值去噪方法<sup>[7-9]</sup>。文献[7,8]在 DCT 域分别提出了一种基于软硬阈值函数的去噪算法，这些算法虽然能在一定程度上消除噪声，但是会造成增强音频信号的严重失真。为了解决这个问题，文献[9]提出了一种改进的软阈值去噪方法，通过比较一帧音频信号功率和噪声功率，并将其划分成信号帧和噪声帧，然后分别对其采取不同的噪声衰减处理办法，这样能很好地去除背景噪声，并且音乐噪声大大减少，但是增强音频信号的高频成分损失比较大，使其主观听觉上比较沉闷，不透亮。

由于现有的 Hiss 噪声消除方法都是在线性域进行的，然而大量音频信号是以压缩比特流的形式存在的，所以现有方法不能满足消除音频比特流中 Hiss 噪声的要求。鉴于此，本文结合 MPEG 组织的 AAC 音频编码器<sup>[10]</sup>，提出了一种基于 AAC 比特流的 Hiss 噪声抑制方法。该方法的输入为含噪音频信

号的 AAC 比特流，输出为增强音频信号的 AAC 比特流，这样，将其用于 AAC 解码器得到最终的增强音频，从而使用户得到干净的音频信号。

## 2 AAC 音频编码器概述

AAC 音频编码器<sup>[10]</sup>被誉为“21 世纪的数据压缩方式”，它是 MPEG 的第三代音频编码标准，也是音频编码领域的一个技术里程碑，但对早期的标准不具有后向兼容性。AAC 编码器定义了 3 种框架（profile），分别是低复杂度（LC, low complexity）框架，主框架（main profile）和可分级采样率（SSR, scalable sampling rate）框架。其中，低复杂度 LC 框架能够在对硬件资源要求不高的场景下提供与主框架相近的音质，因此是最受欢迎的一种框架，本文就是基于 AAC 编码器的 LC 框架展开研究工作的。AAC 编码器是一种感知音频编码器，其一般结构如图 1 所示。

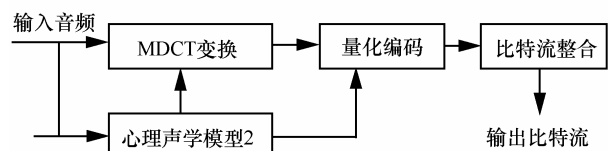


图 1 AAC 编码器结构框图

从图 1 中可知 AAC 编码器的基本编码过程。

1) 首先对输入音频进行分帧，并利用 2 048 个时域样点进行心理声学模型 2 的分析，计算出信掩比和掩蔽阈值，以及编码比特数。同时得到心理声学模型 2 的另一个重要参数为一帧音频信号的变换块类型，即长变换块或短变换块。为了有效地对平稳信号和暂态信号分别进行编码，AAC 编码器采用了变换块切换的方法，其切换示意图如图 2 所示。

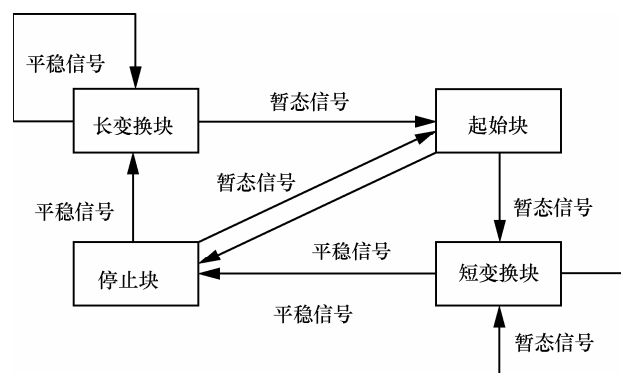


图 2 AAC 编码器变换块切换示意图

其中，长变换块表示当前帧为平稳信号，短变换块表示当前帧为暂态信号，平稳信号到暂态信号的过渡叫起始块，暂态信号到平稳信号的过渡叫停止块。

2) 根据变换块类型对一帧音频信号进行修正离散余弦变换 (MDCT, modified discrete cosine transformation)。对于平稳信号使用长变换块，即对 2 048 个样点加窗并进行 MDCT 变换，得到  $N=1 024$  个 MDCT 系数谱；对于暂态信号采用短变换块，即一帧信号被划分为 8 个短变换块，每个短变换块长度为 256 个样点，并分别进行 MDCT 变换得到各短变换块  $N_s=128$  个 MDCT 系数谱；对于起始和停止的过渡块，处理方式同长变换块，得到  $N=1 024$  个 MDCT 系数谱。

3) 将上述步骤得到的 MDCT 系数谱、变换块的类型以及其他相关的控制参数进行量化和霍夫曼 (Huffman) 编码处理，并进行比特流整合得到一帧音频信号的 AAC 比特流信息。

### 3 基于 AAC 比特流的 Hiss 噪声抑制方法

结合 AAC 音频编码器，本文提出了一种基于 AAC 比特流的 Hiss 噪声抑制方法，其原理框图如图 3 所示。

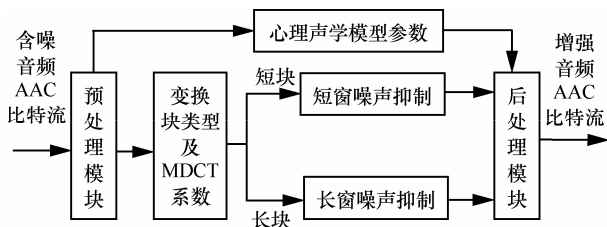


图 3 基于 AAC 比特流的 Hiss 噪声抑制原理框图

首先是对含噪音频信号的 AAC 比特流进行预处理，即利用 AAC 解码算法对其进行解析，得到含噪音频信号的 MDCT 系数、变换块类型以及心理声学模型 2 相关参数；然后，根据变换块的类型对 MDCT 系数分别进行增强处理：如果当前块为短块，则采用短窗 Hiss 噪声抑制方法；如果当前块为长块则采用长窗 Hiss 噪声抑制方法；如果是过渡块，由于过渡块和长变换块的 MDCT 系数数目都为 1 024，所以本文将过渡块当作长变换块，采用长块一样的处理方法，得到增强音频的 MDCT 系数；最后，将上述步骤得到的增强音频 MDCT 系数、变换块类型并结合心理声学模型 2 相关参数进行再量化

编码处理，得到增强音频的 AAC 比特流，并用 AAC 解码器对其进行解码，得到最终的增强音频信号。其中，长窗和短窗 Hiss 噪声抑制模块是本方法的关键模块，下面对这 2 个模块进行详细介绍。

#### 3.1 长窗 Hiss 噪声抑制

传统的噪声抑制算法都是在噪声衰减和信号失真之间进行折中，增强效果不理想。因此，本文引入人耳听觉掩蔽效应，提出了一种基于听觉掩蔽的参数阈值法来抑制长窗 Hiss 噪声。该算法的主要思想是尽量将残差噪声的能量抑制在听觉掩蔽阈值以下，使其不被人耳所感知，而不是完全消除残差噪声，这样就能达到抑制噪声的同时减少音频失真的目的。

假设 AAC 编码前的纯净音频  $x(n)$  被加性 Hiss 噪声  $d(n)$  所污染，这样就可以得到长窗含噪音频  $y(n) = x(n) + d(n)$ ，其在 MDCT 域的表达式为

$$Y(k) = X(k) + D(k) \quad k \in [0, 1, \dots, 1024] \quad (1)$$

其中， $k$  表示 MDCT 系数频点， $Y(k)$ 、 $X(k)$  和  $D(k)$  分别为长窗第  $k$  个含噪音频、纯净音频和噪声信号的 MDCT 系数谱。

根据解析 AAC 比特流得到长变换块含噪音频 MDCT 系数谱，可以得到长窗 Hiss 噪声抑制算法原理框图如图 4 所示。首先对含噪 MDCT 系数采用修正的中值绝对偏差 (MMAD, modified median absolute deviation) 方法估计 Hiss 噪声<sup>[1]</sup>，同时采用基本阈值函数对其进行预增强处理，得到预增强的 MDCT 系数谱；然后利用预增强的 MDCT 系数谱构建修正的离散傅立叶变换 (MDFT, modified discrete Fourier transform) 系数谱，并根据 MDFT 谱系数计算人耳听觉掩蔽阈值和掩蔽参数；最后采用参数阈值函数对含噪 MDCT 系数谱进行增强处理，得到增强音频的 MDCT 系数谱，从而实现长窗 Hiss 噪声的抑制。下面详细介绍其中的关键模块。

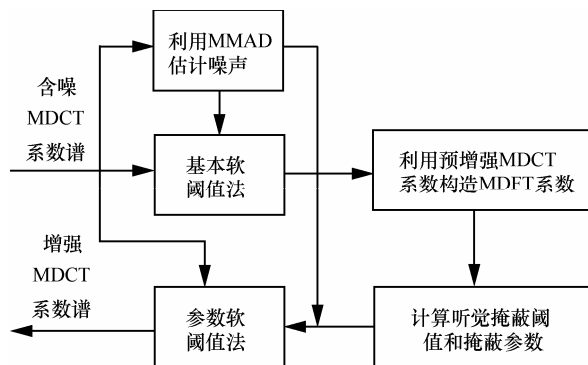


图 4 长窗 Hiss 噪声抑制算法原理框图

### 3.1.1 Hiss 噪声的估计

实验发现 MDCT 谱的高阶系数 ( $k > N/2$ ) 主要是噪声成分, 根据这个特性首先采用中值绝对偏差 (MAD, median absolute deviation) 来进行 Hiss 噪声阈值的粗估计。另外, 考虑到高阶 MDCT 系数还可能存在少量的音频成分, 如果直接把这些音频成分当做噪声的话, 这样会造成噪声的过估现象, 所以为了避免噪声过估问题, 利用 MMAD 方法结合粗估计值进行 Hiss 噪声的细估计。

Hiss 噪声粗估计: 利用 MAD 方法进行 Hiss 噪声的粗估计  $\sigma_v$ , 估计表达式如下<sup>[11]</sup>

$$\sigma_v = V_{\text{MAD}}(\mathbf{Y})/0.6745 \quad (2)$$

其中, MAD 是一个顽健的统计量, 0.6745 是根据统计分布得到的经验值。  $V_{\text{MAD}}(\mathbf{Y})$  表示根据 MDCT 高阶系数向量  $\mathbf{Y}$  计算 MAD 的值, 其计算表达式为

$$V_{\text{MAD}}(\mathbf{Y}) = \text{median}(|\mathbf{Y} - \text{median}(\mathbf{Y})|) \quad (3)$$

其中,  $\text{median}(\mathbf{Y})$  表示取向量  $\mathbf{Y}$  的中值。

Hiss 噪声细估计: 利用 MMAD 方法进行 Hiss 噪声的细估计。根据 MDCT 高阶系数是否存在音频成分可以进行如下 2 个假设<sup>[12]</sup>。

$H_0$ : MDCT 高阶系数不存在音频, 即  $Y(k) = D(k)$ ;

$H_1$ : MDCT 高阶系数存在音频, 即  $Y(k) = X(k) + D(k)$ 。

首先, 根据以上假设, 利用贝叶斯准则, 可以得到条件概率  $p(H_1 | Y(k))$  如下

$$p(H_1 | Y(k)) = \frac{p(H_1)p(Y(k) | H_1)}{p(H_0)p(Y(k) | H_0) + p(H_1)p(Y(k) | H_1)} \quad (4)$$

假设 MDCT 高阶系数存在音频和噪声的概率相等, 即  $p(H_1) = p(H_0) = 0.5$ , 则式(4)可以简化为

$$p(H_1 | Y(k)) = \frac{1}{1 + p(Y(k) | H_0) / p(Y(k) | H_1)} \quad (5)$$

假设 MDCT 高阶系数  $Y(k)$  服从高斯分布, 可得先验条件概率  $p(Y(k) | H_0)$  和  $p(Y(k) | H_1)$  的表达式

$$p(Y(k) | H_0) = \frac{1}{\sqrt{2\pi\sigma_{d,k}^2}} \exp\left(-\frac{|Y(k)|^2}{2\sigma_{d,k}^2}\right) \quad (6)$$

$$p(Y(k) | H_1) = \frac{1}{\sqrt{2\pi(\sigma_{d,k}^2 + \sigma_{x,k}^2)}} \exp\left(-\frac{|Y(k)|^2}{2(\sigma_{d,k}^2 + \sigma_{x,k}^2)}\right) \quad (7)$$

其中,  $\sigma_{d,k}^2$  和  $\sigma_{x,k}^2$  分别表示噪声和音频信号第  $k$  个 MDCT 系数的方差。

定义后验信噪比为  $\gamma_k = Y^2(k) / \sigma_v^2$ , 先验信噪比为  $\zeta_k = \gamma_k - 1$ , 这样, 将式(6)、式(7)代入式(5), 可以得到  $p(H_1 | Y(k))$  的计算式

$$p(H_1 | Y(k)) = \frac{1}{(1 + \sqrt{1 + \zeta_k} \exp[-\gamma_k \zeta_k / 2(1 + \zeta_k)])} \quad (8)$$

其次, 将条件概率  $p(H_1 | Y(k))$  与阈值  $\nu$  进行比较, 如果大于  $\nu$  说明该 MDCT 高阶系数为音频成分, 则将其去掉, 否则保留该 MDCT 高阶系数用于估计 Hiss 噪声, 即保留 MDCT 高阶系数向量  $\mathbf{Y}$  中满足  $p(H_1 | Y(k)) \leq \nu$  条件的 MDCT 系数, 将其用于计算当前帧  $\lambda$  的 MAD 值:

$$\hat{V}_{\text{MAD}}(\lambda) = V_{\text{MAD}}(\mathbf{Y} | [p(H_1 | Y(k)) \leq \nu]) \quad (9)$$

其中,  $\lambda$  为帧索引,  $\nu$  为条件概率阈值, 它是信噪比  $r(k)$  的函数, 其表达式如下

$$\nu = \begin{cases} 1.0, & r(k) < 6 \\ -0.06r(k) + 1.35, & 6 \leq r(k) < 10 \\ -0.025r(k) + 1.0, & 10 \leq r(k) < 20 \\ 0.5, & r(k) \geq 20 \end{cases} \quad (10)$$

其中,  $r(k)$  表示信噪比, 是式(2)粗估计的噪声阈值  $\sigma_v$  和含噪信号功率  $Y^2(k)$  的函数, 其计算式如下

$$r(k) = 10 \lg \left[ \frac{(Y^2(k) - \sigma_v^2)}{\sigma_v^2} \right] \quad (11)$$

再次, 为了避免因噪声估计误差导致的帧与帧之间残留噪声的突变, 本文利用相邻帧之间的相关性, 对  $\hat{V}_{\text{MAD}}(\lambda)$  进行帧间一阶递归平滑处理得到平滑的  $\bar{V}_{\text{MAD}}(\lambda)$

$$\bar{V}_{\text{MAD}}(\lambda) = \alpha \bar{V}_{\text{MAD}}(\lambda - 1) + (1 - \alpha) \hat{V}_{\text{MAD}}(\lambda) \quad (12)$$

其中,  $\lambda$  为帧索引,  $\alpha = 0.95$  为常数平滑因子,  $\bar{V}_{\text{MAD}}(\lambda - 1)$  为前一帧的估计值。

最后, 利用平滑的  $\bar{V}_{\text{MAD}}(\lambda)$  来估计 Hiss 噪声。考虑到噪声估计值与实际噪声存在偏差, 本文采用偏差因子  $\eta$  对其进行补偿, 从而得到 Hiss 噪声的细估计表达式如下

$$\hat{\sigma}_v = \eta \frac{\bar{V}_{\text{MAD}}(\lambda)}{0.6745} \quad (13)$$

其中,  $\lambda$  为帧索引号,  $\eta$  为常数偏差因子, 实验中取 1.3, 用于补偿噪声估计与实际噪声的偏差。

### 3.1.2 MDFT 谱系数的构建

通常，人耳听觉掩蔽阈值是根据 FFT 系数计算得到的，而不能直接利用 MDCT 系数来计算掩蔽阈值。为了能够利用 MDCT 系数来计算掩蔽阈值，本文首先根据基本软阈值法得到的预增强 MDCT 系数构建 MDFT 系数<sup>[13]</sup>，然后利用 MDFT 系数计算掩蔽阈值。MDFT 系数的具体构建过程如下。

首先利用预增强 MDCT 系数  $X_{mdct}(k)$  派生出修正的离散正弦变换 (MDST, modified discrete sine transform) 系数  $Y_{mdst}(k)$ ，表达式如下

$$Y_{mdst}(k) = (-1)^k X_{mdct}^-(k) + (-1)^k \sum_{l=-m}^{m-1} \frac{X_{mdct}^+(idx(k-2l-1))}{(-1)^l N \sin[\theta(2l+1)]} \quad (14)$$

其中， $k$  为 MDCT 系数频点索引， $N$  为帧长， $m$  为经验常数值，在实验中取  $m=5$  用来控制 MDCT 系数到 MDST 系数的转换误差， $\theta$  表示转换相角，这里取  $\theta=\pi/2N$ ， $idx(\cdot)$  表示取 MDCT 系数索引的函数， $X_{mdct}^-(k)$  和  $X_{mdct}^+(k)$  计算式如下

$$X_{mdct}^-(k) = [X_{mdct}(\lambda+1, k) - X_{mdct}(\lambda-1, k)]/2 \quad (15)$$

$$X_{mdct}^+(k) = [X_{mdct}(\lambda+1, k) + X_{mdct}(\lambda-1, k)]/2 \quad (16)$$

其中， $\lambda$  为帧索引号。

然后利用  $X_{mdct}(k)$  和  $Y_{mdst}(k)$  根据式(17)构建 MDFT 系数  $Z_{mdft}(k)$ ：

$$Z_{mdft}(k) = \begin{cases} Z_-(0) - jZ_-(0), & k=0 \\ -Z_+(k-1) - jZ_-(k), & k=1, \dots, N-1 \\ -Z_+(N-1) - jZ_-(N-1), & k=N \end{cases} \quad (17)$$

其中， $Z_-(k) = Y_{mdst}(k) - X_{mdct}(k)$ ， $Z_+(k) = Y_{mdst}(k) + X_{mdct}(k)$ 。

根据文献[13]可知，式(17)可以转换成

$$\begin{cases} Z_{mdft}(k) = e^{-j(k)} \sum_{n=0}^{2N-1} x(n) e^{-j\frac{\pi}{N}nk} \\ \mathbf{j}(k) = \frac{\pi k(1+N)}{2N} + \frac{\pi}{4} \end{cases} \quad (18)$$

从式(18)可知，MDFT 系数与 DFT 系数具有相同的幅度，只是存在相位差  $\varphi(k)$ ，这说明 MDFT 系数与 DFT 系数具有相同的时频特性。这里只考虑掩蔽特性，下面通过实验来验证 MDFT 系数能用于掩蔽阈值的计算。

由于听觉掩蔽阈值是根据 FFT 功率谱计算的，所以首先验证基于 MDFT 和 FFT 系数的功率谱是

否一致，图 5 为一段音频信号的 MDFT 和 FFT 功率谱对比，从图 5 可以看出 MDFT 和 FFT 功率谱基本是一致的；然后验证基于 MDFT 系数和 FFT 系数的掩蔽阈值差异，从图 6 可以看出基于 MDFT 与 FFT 计算的掩蔽阈值近似相等，从而可以说明 MDFT 系数能用于掩蔽阈值计算。

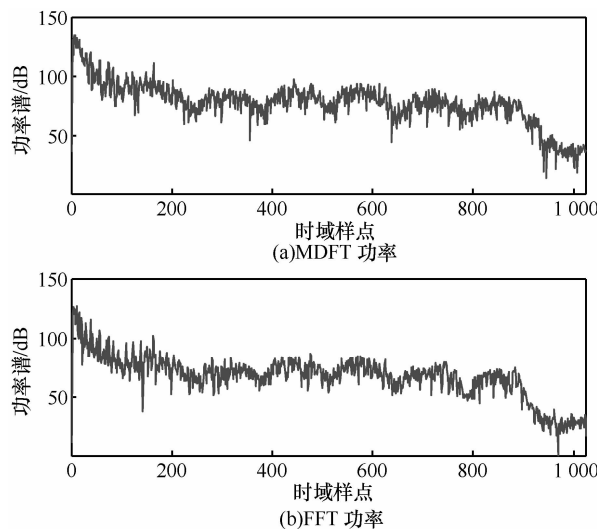


图 5 基于 MDFT 与 FFT 系数的功率谱对比

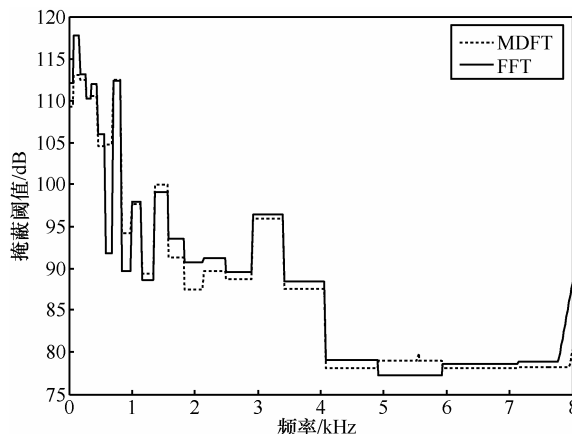


图 6 基于 MDFT 与 FFT 系数的掩蔽阈值对比

### 3.1.3 掩蔽阈值的计算

掩蔽效应是指人耳不能区分 2 个几乎同时出现的声音。本文只考虑频域掩蔽，它是指弱音频会被同时出现的强音频掩蔽掉，从而使其不能被人耳所感知。

根据 MDFT 系数谱，掩蔽阈值计算步骤如下<sup>[11]</sup>。

1) 临界带的划分：在整个频率范围内划分 25 个临界带，一个临界带表示一个巴克 (Bark) 带，这样可以得到 MDFT 系数频点  $k$  与 Bark 索引号  $z$  的映射关系如表 1 所示。

表 1 MDFT 系数与临界带的映射

Bark(z)	频点(k)	频率间隔/Hz
1	1~4	0~100
2	5~9	100~200
3	10~14	200~300
4	15~18	300~400
5	19~23	400~510
6	24~29	510~630
7	30~35	630~770
8	36~42	770~920
9	43~50	920~1 080
10	51~58	1 080~1 270
11	59~68	1 270~1 480
12	69~79	1 480~1 720
13	80~92	1 720~2 000
14	93~107	2 000~2 320
15	108~125	2 320~2 700
16	126~146	2 700~3 150
17	147~171	3 150~3 700
18	172~204	3 700~4 400
19	205~246	4 400~5 300
20	247~297	5 300~6 400
21	298~357	6 400~7 700
22	358~441	7 700~9 500
23	442~557	9 500~12 000
24	558~719	12 000~15 500
25	720~1 024	15 500

2) Bark 功率谱的计算：根据临界带的划分，先计算临界带中各个 MDFT 系数频点的功率谱，然后将临界带中各频点功率谱求和得到每个临界带的 Bark 功率谱。

3) Bark 功率谱的扩展：由于相临的临界带之间也存在掩蔽效应，这样需要将步骤 2)得到的 Bark 功率谱与扩展函数  $S_F(z)$ 进行卷积，得到扩展的 Bark 功率谱。

4) 掩蔽阈值的计算：由于纯音和噪声具有不同的掩蔽特性，因此需要利用谱平坦测度来判断信号是否具有纯音特征或噪声特征，并由此计算偏差因子，然后将偏差因子从扩展 Bark 功率谱中减去，得到未修正的掩蔽阈值。

5) 掩蔽阈值的修正：为使各频点的掩蔽阈值均不小于绝对听觉阈值，将步骤 4)得到的掩蔽阈值与绝对听觉阈值相比，取其大者作为最终的 Bark 掩蔽阈值  $T(z)$ ，并将其转换回线性频率刻度得到各频点的掩蔽阈值  $T(k)$ 。

### 3.1.4 阈值函数及掩蔽参数的计算

本文引入人耳听觉掩蔽效应到音频增强中来，

其目的是尽量将残留噪声能量抑制在纯净音频的掩蔽阈值以下，使其不能被人耳感知。因此，首先对含噪音频进行预增强处理，将得到的预增强音频计算听觉掩蔽阈值；最后利用基于听觉掩蔽的参数阈值函数进行增强处理，得到增强音频的 MDCT 系数<sup>[11]</sup>。

1) 预增强处理：本文采用基本软阈值法对含噪音频信号进行预增强处理，得到预增强音频的 MDCT 系数  $\hat{X}(k)$  如下

$$\hat{X}(k) = \begin{cases} Y(k)[1 - \hat{\sigma}_v / |Y(k)|]^p, & [\hat{\sigma}_v / |Y(k)|] < 1 \\ 0 & , \text{其他} \end{cases} \quad (19)$$

其中， $p=0.4$  为常数因子，保证阈值函数的光滑连续性，并控制噪声的消除效果。 $\hat{\sigma}_v$  为式(13)估计的 Hiss 噪声阈值， $Y(k)$ 为含噪信号的 MDCT 系数。

2) 增强处理：设存在增益因子  $G(k)$ ，则有增强音频  $\hat{X}(k) = G(k)Y(k)$ ，这样得到增强音频  $\hat{X}(k)$  与纯净音频  $X(k)$ 的估计误差  $E(k)$ 为

$$E(k) = \hat{X}(k) - X(k) = E_x(k) + E_d(k) \quad (20)$$

其中， $E_x(k)$ 和  $E_d(k)$ 分别表示音频信号的失真误差和残留噪声，计算式分别如下

$$E_x(k) = [G(k) - 1]X(k) \quad (21)$$

$$E_d(k) = G(k)D(k) \quad (22)$$

假设音频信号与噪声信号不相关，则引入听觉掩蔽阈值  $T(k)$ 来得到较优的增益因子  $G(k)$ ，使其最小化音频失真且残留噪声在掩蔽阈值以下，即

$$\min_{G(k)} \{E_x^2(k)\} \quad (23)$$

约束条件为  $E_d^2(k) \leq T(k)$  (24)

因此，为了得到增益因子  $G(k)$ ，根据音频失真和残留噪声，构造代价函数  $J(k)$ 如下<sup>[14]</sup>

$$J(k) = E_x^2(k) + \mu(k)[E_d^2(k) - T(k)] \quad (25)$$

其中， $\mu(k)$ 为拉格朗日乘子。

将式(21)、式(22)代入式(25)，并使  $J(k)$ 最小，可以得到增益因子  $G(k)$ <sup>[14]</sup>：

$$G(k) = \frac{1}{\max \left\{ \sqrt{\frac{|D(k)|^2}{T(k)}}, 1 \right\}} \quad (26)$$

从式(26)可以看出，如果掩蔽阈值大于噪声功率时， $G(k)$ 较大，即对含噪音频进行较小的衰减；

如果掩蔽阈值小于噪声功率时,  $G(k)$ 较小, 即对含噪声频进行较大的衰减, 因此, 根据这个准则, 引入听觉掩蔽参数  $\alpha(k)$ 并将式(26)进行变换, 从而得到基于听觉掩蔽的参数阈值函数如下

$$\hat{X}(k) = \begin{cases} Y(k) \sqrt{\left\{1 - \beta \cdot \alpha(k) \left[ \frac{\hat{\sigma}_v}{|Y(k)|} \right]^2 \right\}}, & |Y(k)|^2 \geq \beta \cdot \alpha(k) \cdot \hat{\sigma}_v^2 \\ 0, & \text{其他} \end{cases} \quad (27)$$

其中,  $\hat{X}(k)$ 为增强音频的 MDCT 系数, 掩蔽参数  $\alpha(k)$ 是关键参数, 用来控制 Hiss 噪声的衰减程度。当掩蔽阈值  $T(k)$ 较大时, 说明音频信号的掩蔽能力较强, 可以较好地掩蔽噪声, 此时  $\alpha(k)$ 取一个较小值  $\alpha_{\min}$ , 即使式(27)对含噪信号进行较小的衰减以减小音频信号的失真; 当掩蔽阈值  $T(k)$ 较小时, 说明音频信号的掩蔽能力较弱, 其不能很好地掩蔽噪声, 此时  $\alpha(k)$ 取一个较大值  $\alpha_{\max}$ , 即使式(27)对含噪信号进行较大的衰减以尽可能地消除噪声; 对于其他  $\alpha(k)$ 参数值, 将其看作是掩蔽阈值  $T(k)$ 的函数, 采用插值方法得到, 这样, 得到掩蔽参数  $\alpha(k)$ 的计算表达式如下<sup>[11]</sup>

$$\alpha(k) = \begin{cases} \alpha_{\min}, & T(k) \geq T_{\max} \\ \alpha_{\max}, & T(k) \leq T_{\min} \\ \frac{[\alpha_{\max}(T_{\max} - T(k)) + \alpha_{\min}(T(k) - T_{\min})]}{T_{\max} - T_{\min}}, & \text{其他} \end{cases} \quad (28)$$

其中,  $\alpha_{\max}=7.0$  和  $\alpha_{\min}=1.0$  分别为掩蔽参数的最大值和最小值;  $T_{\max} = \arg \max_{0 \leq k < N} T(k)$  和  $T_{\min} = \arg \min_{0 \leq k < N} T(k)$  分别为当前帧掩蔽阈值的最大值和最小值。

对于含噪 AAC 比特流解码得到的 MDCT 系数, 由于其经过了量化处理, 所以其高阶系数会存在很多零值, 这样会影响 Hiss 噪声估计的准确性, 从而影响增强效果。为解决这个问题, 在式(27)的参数阈值函数中, 采用参数  $\beta$  对不同频段的 MDCT 系数进行不同程度的衰减,  $\beta$  的表达式如下

$$\beta = \begin{cases} 3.0, & 0 \leq k \leq k_{LF} \\ 4.0, & k_{LF} \leq k \leq k_{MF} \\ 5.0, & k_{MF} \leq k \leq N - 1 \end{cases} \quad (29)$$

其中,  $N=1024$  表示变换块的长度,  $k$  为 MDCT 系

数频点索引号,  $k_{LF}=N/5$  表示低频段的上界,  $k_{MF}=N/2$  表示中频段的上界。

### 3.2 短窗 Hiss 噪声抑制

在 AAC 编码器中, 一帧短窗音频被划分为 8 个子帧, 且各个子帧分别进行 MDCT 变换, 然后分别对每个子帧进行编码处理。因此, 对于短窗 Hiss 噪声的抑制, 也需要对每个子帧分别进行增强处理。假设 AAC 编码前的短窗各子帧纯净音频  $x_s(n)$  被加性 Hiss 噪声  $d_s(n)$  所污染, 则含噪信号  $y_s(n) = x_s(n) + d_s(n)$  变换到 MDCT 域的表达式为

$$Y_s(k) = X_s(k) + D_s(k), \quad k \in [0, 1, \dots, 128] \quad (30)$$

其中,  $k$  表示短窗子帧 MDCT 系数频点,  $Y_s(k)$ 、 $X_s(k)$  和  $D_s(k)$  分别为子帧第  $k$  个含噪音频、纯净音频和噪声信号的 MDCT 系数谱。

根据解析 AAC 比特流得到的 8 个子帧含噪音频 MDCT 系数, 得到短窗 Hiss 噪声抑制流程如图 7 所示。

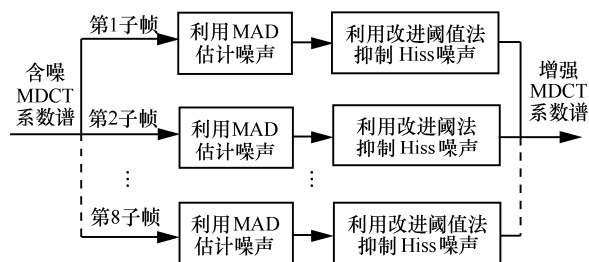


图 7 短窗 Hiss 噪声抑制流程图

首先根据长窗帧的噪声估计作为短窗各子帧的噪声估计初值; 然后利用 MAD 方法对各子帧进行噪声估计, 并结合噪声初值进行一阶平滑处理得到各子帧的噪声估计; 最后利用一种新的改进软阈值法对各子帧 MDCT 系数进行 Hiss 噪声的抑制, 得到各子帧增强音频的 MDCT 系数。

#### 3.2.1 短窗 Hiss 噪声的估计

由于短窗子帧帧长  $N_s=128$  较短, 并且其 MDCT 高阶系数存在很多零值, 从而使其高阶系数存在音频成分对噪声估计的影响不是很大, 所以这里不考虑其高阶系数存在音频成分的情况, 只采用 MAD 方法来估计各子帧的 Hiss 噪声, 其计算步骤如下。

- 1) 将该短窗之前的长窗噪声估计作为短窗子帧噪声估计的初值, 即作为子帧噪声估计的历史值  $\bar{v}_{MAD}$ 。
- 2) 根据  $\lambda_s$  子帧的 MDCT 系数, 利用式(2)计算

当前子帧的 MAD 值  $V_{\text{MAD}}(\lambda_s)$ 。

3) 为了避免由于噪声估计差异而引起的增强音频在长短窗切换处的听觉不连续性, 采用一阶递归平滑来处理得到各子帧的噪声估计:

$$\bar{V}_{\text{MAD}}(\lambda_s) = \alpha_s \bar{V}_{\text{MAD}} + (1 - \alpha_s) V_{\text{MAD}}(\lambda_s) \quad (31)$$

其中,  $\lambda_s$  为子帧索引号,  $\alpha_s=0.95$  为常数平滑因子。

由于短窗音频为暂态信号, 则各子帧的能量较大, 容易出现噪声过估现象, 所以根据式(32)对子帧噪声估计进行修正, 得到各子帧最终的噪声估计值  $\hat{\sigma}_{\text{vs}}$ :

$$\hat{\sigma}_{\text{vs}} = \eta_s \cdot \frac{\bar{V}_{\text{MAD}}(\lambda_s)}{0.6745} \quad (32)$$

其中,  $\eta_s$  为常数修正因子, 实验中取 0.25。

### 3.2.2 短窗 Hiss 噪声的抑制

本文提出了一种新的改进软阈值去噪方法来消除短窗各子帧中的 Hiss 噪声。

首先将子帧  $N_s=128$  个频点划分为若干子带, 每个子带的长度为  $L$ ; 然后计算每个子带的平均功率, 并将各个子带的平均功率与噪声功率进行比较, 如果子带平均功率大于噪声功率说明该子带为音频信号子带, 否则为噪声子带; 最后针对音频信号子带和噪声子带分别采取不同的噪声衰减处理办法。其阈值函数如下:

$$\hat{X}_s(k) = \begin{cases} \text{sign}(Y_s(k)) \{ \max[0, (|Y_s(k)| - m_y)] \}, \\ \frac{1}{L} \sum_{k=1}^L |Y_s(k)|^2 \geq \zeta \hat{\sigma}_{\text{vs}}^2 \\ \text{sign}(Y_s(k)) \{ \max[0, (|Y_s(k)| - m_d)] \}, \text{其他} \end{cases} \quad (33)$$

其中,  $\hat{X}_s(k)$  为子帧增强音频的 MDCT 系数谱;  $Y_s(k)$  为子帧含噪音频的 MDCT 系数谱;  $\text{sign}(x)$  表示取变量  $x$  的符号;  $\hat{\sigma}_{\text{vs}}$  为式(32)估计的子帧噪声阈值;  $\zeta$  为常数, 实验中取 20, 用来控制区分音频信号子带和噪声子带。

$m_y$  和  $m_d$  分别为音频信号子带因子和噪声子带因子, 其计算式如下<sup>[9]</sup>:

$$m_y = \frac{\rho_y \hat{\sigma}_{\text{vs}}}{\frac{1}{L} \sum_{j=1}^L j^2} \quad m_d = \frac{\rho_d \hat{\sigma}_{\text{vs}}}{\frac{1}{L} \sum_{j=1}^L j^2} \quad (34)$$

其中,  $j$  为子带长度的索引号,  $\rho_y$ 、 $\rho_d$  分别为小于 1 的常数, 并满足  $\rho_y=0.3\rho_d$  的关系, 实验中  $\rho_d$  取 0.8。

## 4 实验结果及分析

由于目前关于 Hiss 噪声消除的文献较少, 为了验证本文所提方法对 AAC 比特流中 Hiss 噪声抑制的有效性, 将现有的 3 种 Hiss 噪声抑制算法与本文算法进行比较, 包括文献[1]提出的结合最小值控制递归平均噪声估计<sup>[15]</sup>(MCRA, minima controlled recursive averaging algorithms)的谱减法、文献[6]提出的基于小波变换的自适应多阈值法和文献[9]提出的改进软阈值法。为了简化表示, 这 4 种算法分别表示为 Our、SS、Softthr 和 Wavelet。对于各 Hiss 噪声抑制算法性能的衡量, 采用平均分段信噪比  $SNR_{\text{seg}}$  和 ITU-R BS.1387-1 标准中的音频质量感知评估<sup>[16]</sup> (PEAQ, perceptual evaluation of audio quality) 作为客观质量评价方法, 并采用主观平均意见得分 (MOS, mean opinion score) 和 A/B 主观偏爱听觉测试进行主观质量评测。

在测试实验中, Hiss 噪声来自剑桥大学网站<sup>[17]</sup>, 测试数据库为 3GPP 和 MPEG 组织 44.1kHz 采样的音频数据各 16 段, 每段为一类音频, 主要包括小提琴、鼓、铜钹、吉他、交响乐和摇滚乐等音频类别。先采用 ITU 的加噪工具对纯净测试音频添加 Hiss 噪声, 得到信噪比分别为 6dB、12dB 和 18dB 的含噪音频, 并将这些含噪音频进行 AAC 编码得到含噪音频的 AAC 比特流文件, 这里以 AAC 编码器 96kbit/s 比特率为例进行测试; 再利用本文算法对 AAC 比特流文件进行增强处理得到增强音频的 AAC 比特流; 最后将增强音频的 AAC 比特流通过 AAC 解码器得到时域增强音频, 将其与参考算法的增强音频进行比较。本文算法和几种参考算法的 Hiss 噪声增强音频频谱如图 8 所示。

通过对比图 8 中增强音频的频谱图可以发现, SS 算法得到的增强音频引入了“音乐噪声”, 并且低频存在较大的损失; Softthr 算法虽然能有效地去除 Hiss 噪声, 但是增强信号的高频信息严重缺失; Wavelet 算法还残留不少 Hiss 噪声, 增强效果较差; Our 算法不仅有效地消除了 Hiss 噪声, 而且很好地保留了高频信息, 听觉效果明显得到改善。

本文采用 PEAQ 来衡量增强音频的客观质量。PEAQ 的评价标准为客观差异得分<sup>[16]</sup> (ODG, objective difference grade), 它是一种模拟人耳主观听觉的客观测试方法, 其得分范围为[-4,0], 得分为 0 表示差异最小, ODG 得分越小表示差异越大。当 ODG 增加 0.1 时, 说明音频质量得到显著改善,

本文将 ODG 的提高简称为 ODGI。根据纯净音频和增强音频的 ODG 得分, 将其减去纯净音频和含噪音频的 ODG 得分得到 ODGI 得分。测试结果如表 2 所示, 可以看出本文算法得到增强音频 ODG 改善程度明显好于其他 3 种增强算法, 说明本文算法对 Hiss 噪声的抑制效果明显优于参考算法。

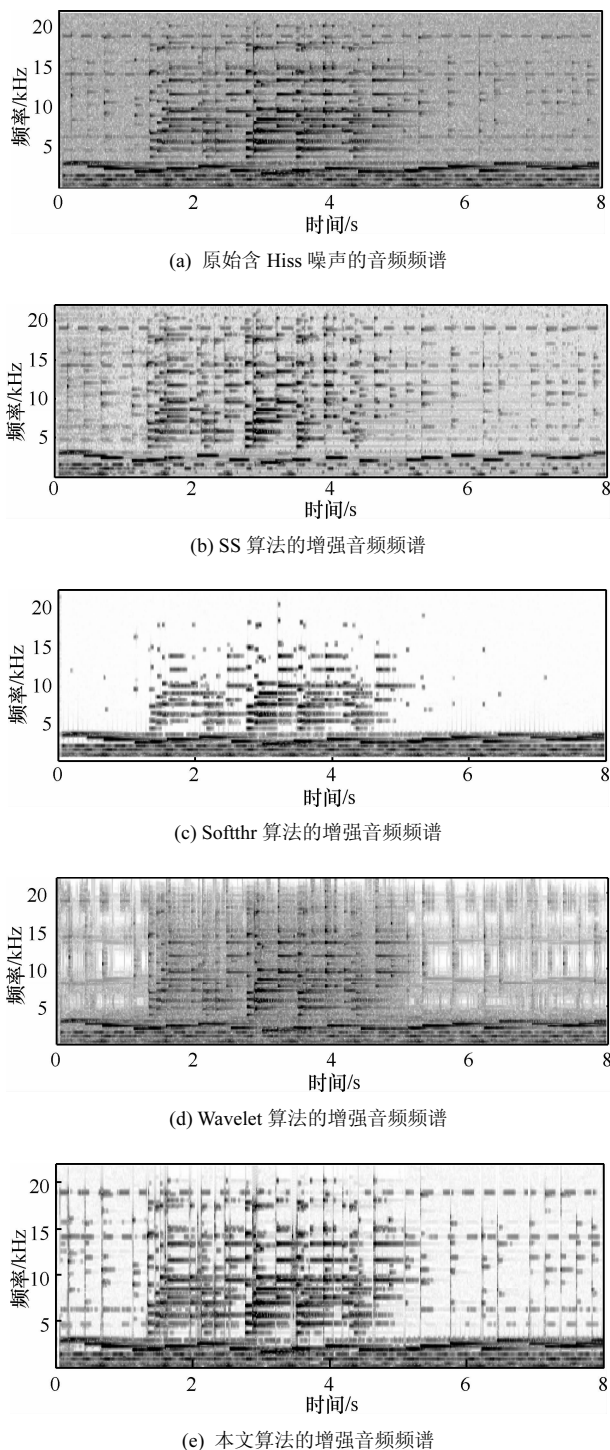


图 8 含噪音频和 4 种算法增强音频频谱对比

表 2 客观音频质量测试 ODGI 得分

算法	6dB	12dB	18dB
SS 算法	0.06	0.08	0.07
Softthr 算法	0.35	0.32	0.23
Wavelet 算法	0.12	0.14	0.09
本文算法	0.42	0.39	0.38

对于本算法的客观测试, 还采用分段信噪比  $SNR_{seg}$  提高来衡量不同增强算法的噪声抑制性能。 $SNR_{seg}$  为测试音频序列帧信噪比的几何平均, 式(35)和式(36)分别给出了原始含噪音频的分段信噪比  $SNR_{seg-in}$  和增强音频的分段信噪比  $SNR_{seg-out}$  计算表达式。

$$SNR_{seg-in} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \lg \left( \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} [x(n) - y(n)]^2} \right) \quad (35)$$

$$SNR_{seg-out} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \lg \left( \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} [x(n) - \hat{x}(n)]^2} \right) \quad (36)$$

其中,  $M$  为测试音频信号的总帧数,  $N$  为帧长,  $x(n)$  表示纯净时域音频信号,  $y(n)$  表示含噪时域音频信号,  $\hat{x}(n)$  表示增强时域音频信号。

根据式(35)、式(36)得到的分段信噪比  $SNR_{seg-in}$ 、 $SNR_{seg-out}$ , 可以得到增强算法的信噪比提高  $SNRI$ , 即  $SNRI = SNR_{seg-out} - SNR_{seg-in}$ ,  $SNRI$  测试结果如图 9 所示。

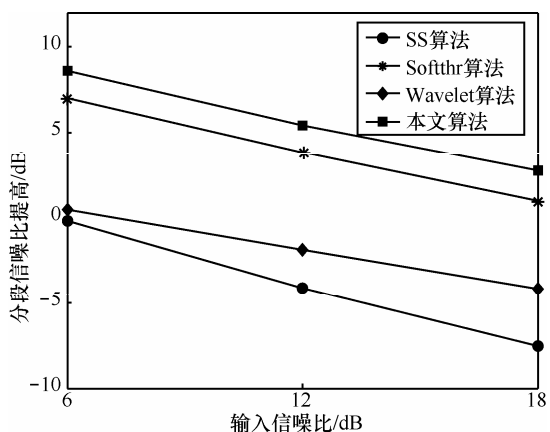


图 9 4 种算法分段信噪比提高测试结果

从图 9 可以看出, 本文算法的分段信噪比提高程度明显好于 3 种参考算法, 说明本文算法比参考算法具有更好的噪声抑制性能。

另外，本文采用平均 MOS 分和 A/B 听觉测试方法来衡量增强音频的主观质量。邀请 12 名测听者（6 男，6 女）分别对本文算法和 3 种参考算法的增强音频进行听觉测试。

对于主观平均 MOS 的评分，是根据测听者的感受将增强音频划分为 5 个等级，分值范围由低到高为 [1, 2, 3, 4, 5]，其中，“1”表示增强音频的质量极差，失真严重无法忍受；“5”表示增强音频与原始音频只有很小的差异，几乎感觉不到失真。在测听实验时，测听者分别对原始音频和增强音频进行主观听觉对比并选择合适的分数；然后分别统计 3 种信噪比下各增强算法得到的增强音频对应的 MOS 分；最后计算出 3 种信噪比下各增强算法的平均 MOS 分。平均 MOS 得分测试结果如图 10 所示。

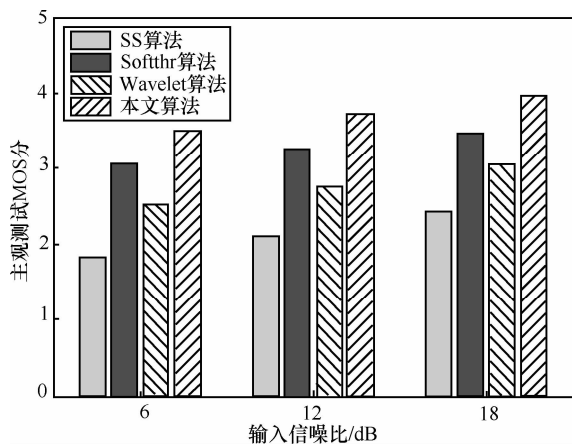


图 10 主观测试平均 MOS 分

从图 10 可以看出，本文方法在 3 种信噪比下的平均 MOS 分均高于几种参考算法的 MOS 分。

对于 A/B 听觉测试，针对 3 种信噪比分别测试。在任意一种信噪比情况下，将本文算法和其中一种参考算法的增强音频测试序列随机排列，测听者从中选择自己偏爱的一种，或者选择无偏爱，并统计该信噪比下的 A/B 测试结果；同理得到其他信噪比下的 A/B 偏爱测试结果；最后计算 3 种信噪比下总的平均 A/B 测试结果。因此，本文算法与 3 种参考算法的 A/B 测试结果如表 3~表 5 所示。

表 3 SS 算法和本文算法的 A/B 测试结果

信噪比	偏爱 SS 算法	偏爱本文算法	无偏爱
6dB	8.9%	69.7%	21.4%
12dB	12.4%	60.4%	27.2%
18dB	16.7%	46.8%	36.5%
平均	12.7%	58.9%	28.4%

表 4 Softthr 算法和本文算法的 A/B 测试结果

信噪比	偏爱 Softthr 算法	偏爱本文算法	无偏爱
6dB	30.2%	43.1%	26.7%
12dB	31.3%	37.8%	30.9%
18dB	33.9%	34.6%	31.5%
平均	31.8%	38.5%	29.7%

表 5 Wavelet 算法和本文算法的 A/B 测试结果

信噪比	偏爱 Wavelet 算法	偏爱本文算法	无偏爱
6dB	20.3%	51.2%	28.5%
12dB	22.7%	45.4%	31.9%
18dB	26.3%	37.2%	36.5%
平均	23.1%	44.6%	32.3%

从表 3~表 5 可以看出，在任一种信噪比以及总平均情况下，测听者都明显偏爱本文算法的增强音频信号，从而可以说明本文算法增强音频的主观听觉质量明显优于 3 种参考算法。

## 5 结束语

本文提出了一种基于 AAC 比特流的 Hiss 噪声抑制算法，该算法的输入输出均为音频信号的 AAC 比特流。在压缩域，针对 AAC 的长、短变换块类型，分别提出了一种基于听觉掩蔽的参数软阈值方法和一种新的改进软阈值方法进行 Hiss 噪声的抑制处理。主观和客观测试结果表明，在 96kbit/s 码率下，本文所提算法能有效去除音频信号 AAC 比特流中的 Hiss 噪声，并且噪声抑制性能和增强音频的听觉质量都明显优于参考算法，从而说明本文方法的有效性。由于本文方法是根据解码 AAC 码流得到的 MDCT 系数消除 Hiss 噪声的，所以理论上 AAC 码率的大小对本文方法的性能影响不大，对所有 AAC 码率均适用。接下来的工作需要进一步实验验证本文方法在 AAC 其他码率下均能有效消除 Hiss 噪声。

## 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. 135-149.
- [2] DONOHO D L. De-noising by Soft-thresholding[J]. IEEE Trans Inf Theory, 1995, 41: 613-627.
- [3] PENG Y H. De-noising by modified soft-thresholding[A]. IEEE APCCAS[C].2000. 760-762.
- [4] LIN Y, CAI J. A new threshold function for signal denoising based on wavelet transform[A]. ICMTMA[C].2010. 200-203.
- [5] GHANBARI Y, KARAMI-MOLLAIEI M R. A new approach for speech enhancement based on the adaptive thresholding of the wavelet

- packets[J]. *Speech Communication*, 2006, 48(8): 927-940.
- [6] 李占辉. 基于小波变换的音频修复技术研究与应用[D]. 上海: 东华大学, 2010. 33-42.  
LI Z H. *Research on Audio Restoration based on Wavelet*[D]. Shanghai: Donghua University, 2010. 33-42.
- [7] HASAN M K, ZILANY M S A, KHAN M R. DCT speech enhancement with hard and soft thresholding criteria[J]. *Electronics Letters*, 2002, 38(13): 669-670.
- [8] SALAHUDDIN S, AISLAM S Z, HASAN M K, *et al.* Soft thresholding for DCT speech enhancement[J]. *Electronics Letters* 21st November, 2002, 38(24): 1605-1607.
- [9] LE P N, AMBIKAI RAJAH E, CHOI E. An improved soft threshold method for DCT speech enhancement[A]. ICCE[C]. 2008. 268-271.
- [10] ISO/IEC 13818-7, Information Technology-Generic Coding of Moving Pictures and Associated Audio Information-part 7: Advanced Audio Coding[S]. 1997.
- [11] DENG F, BAO C C, XIA B Y, *et al.* A novel hiss noise reduction method for audio signals based on MDCT[A]. The 2011 International Conference on Wireless Communications and Signal Processing (WCSP2011)[C]. Nanjing, China, 2011.
- [12] HASAN M K, SALAHUDDIN S, KHAN M R. Reducing signal-bias from MAD estimated noise level for DCT speech enhancement[J]. *Signal Processing*, 2004, 84: 151-162.
- [13] CHEN S X, XIONG N X, JONG H P, *et al.* Analysis and synthesis of spatial parameters using MDCT[A]. MUE 2009[C]. 2009. 18-21.
- [14] LU C T, TSENG K F. A gain factor adapted by masking property and SNR variation for speech enhancement in colored-noise corruptions[J]. *Computer Speech and Language*, 2010, 24(4): 632-647.
- [15] COHEN I, BERDUGO B. Noise estimation by minima controlled recursive averaging for robust speech enhancement[J]. *IEEE Signal Process Lett*, 2002, 9(1): 12-15.
- [16] ITU-R RECOMMENDATION BS.1387-1: Method for Objective Measurements of Perceived Audio Quality[S]. 1998-2001.
- [17] [EB/OL]. <http://www-sigproc.eng.cam.ac.uk/%7Eesjg/springer/index.html>.

#### 作者简介:



邓峰 (1983-), 男, 湖南郴州人, 北京工业大学博士生, 主要研究方向为语音、音频增强。



鲍长春 (1965-), 男, 内蒙古赤峰人, 博士, 北京工业大学教授、博士生导师, 主要研究方向为语音与音频信号处理。



鲍枫 (1990-), 男, 北京人, 北京工业大学研究生, 主要研究方向为语音、音频增强。