

基于 MPEG-AAC 编码器的压缩域音频增强方法

邓 峰, 鲍 枫, 鲍长春

(北京工业大学电子信息与控制工程学院, 北京 100124)

摘 要: 本文基于 MPEG-AAC 音频编解码器, 提出了一种压缩域的音频增强方法. 首先, 对含噪音频信号的比特流进行解码, 得到含噪音频信号的 MDCT 系数; 然后, 利用修正的加权递归平均 (Modified Weighted Recursive Averaging, MWRA) 方法估计噪声功率; 再者, 利用基于听觉掩蔽原理的自适应 β -阶双曲余弦 (COSH) 统计模型, 对含噪音频的 MDCT 系数进行增强处理; 最后, 将增强后的 MDCT 系数重新量化编码, 得到用于解码的增强比特流. 实验结果表明, 本文提出的方法能有效去除 AAC 解码音频信号中的多种背景噪声, 其性能明显优于参考方法.

关键词: 音频增强; AAC 比特流; 压缩域; MWRA 噪声估计; COSH 估计器

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2014)07-1410-09

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.07.026

Audio Enhancement in Compressed Domain Based on MPEG-AAC Codec

DENG Feng, BAO Feng, BAO Chang-chun

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

Abstract: An audio enhancement method based on MPEG-AAC codec in compressed domain is proposed. First, the bit-stream derived from noisy audio signal is decoded to obtain MDCT coefficients of noisy audio signals. Then, the noise power is estimated by modified weighted recursive averaging (MWRA). Next, the adaptive β -order COSH statistic modal method is employed to enhance MDCT coefficients. Finally, the enhanced MDCT coefficients are re-quantized to obtain an enhanced bit-stream which is used to get the enhanced audio signals by AAC decoder. The test results indicate that the proposed algorithm can effectively remove the noises derived from AAC bit-stream of audio signals and obviously outperforms the reference noise reduction methods.

Key words: audio enhancement; AAC bit-stream; compressed domain; MWRA noise estimation; COSH estimator

1 引言

在用 MPEG 的 MP3 或 AAC 进行现场录音时, 常常会混入各种噪声, 如常见的多人说话 (Babble) 噪声、街道 (Street) 噪声等, 这严重影响了音频信号的回放质量. 因此, 如何消除隐含在这些录制音频中的噪声是一个十分有应用价值的课题. 目前有关语音增强的方法很多, 如谱减法^[1~3]、统计模型算法^[4~6]和变换域增强法等^[7~13]. 这些方法对语音信号而言, 性能颇佳, 但对音频信号的增强结果如何尚无报道.

因为由 MP3 或 AAC 录制的音频信号是以比特流形式存储的, 所以现有的工作在线性域的语音增强方法不能满足音频去噪的目的. 因此, 本文基于 MPEG-AAC 音频编解码器^[14], 提出了一种压缩域的音频增强方法, 即

首先对含噪音频信号的比特流进行解码, 得到含噪音频信号的 MDCT 系数, 其次, 对含噪音频的 MDCT 系数进行增强处理, 最后, 将增强后的 MDCT 系数重新量化编码, 得到用于解码的增强比特流.

对含噪音频的 MDCT 系数进行增强处理是音频增强的关键, 为此基于听觉掩蔽原理, 提出一种自适应 β -阶双曲余弦 (COSH) 统计模型方法, 用于修正含噪音频信号的 MDCT 系数, 噪声估计采用了修正的加权递归平均 (MWRA) 法. 实验结果验证了该方法的有效性.

2 AAC 编码器简介

AAC 音频编码器^[14]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 编码器定义了三种框架(Profile),分别是低复杂度框架(Low Complexity Profile, LC),主框架(Main Profile)和可分级采样率框架(Scalable Sampling Rate Profile, SSR).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.

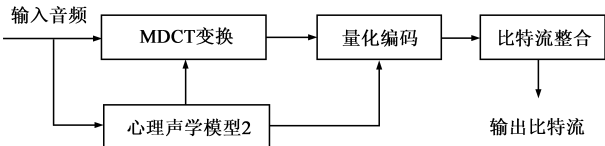


图1 AAC编码器结构框图

AAC 编码器的编码过程分为如下三步:

(1)首先对输入音频进行分帧,并利用 2048 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:一帧音频信号的变换块类型,即长变换块或短变换块.为了能有效地对平稳信号和暂态信号进行编码,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.

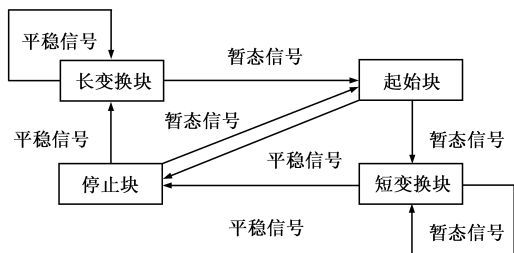


图2 AAC编码器变换块切换示意图

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(Modified Discrete Cosine Transformation, MDCT).对于平稳信号使用长变换块,即对 2048 个样点加窗并进行 MDCT 变换,得到 $N = 1024$ 个 MDCT 系数;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;对于起始和停止的过渡块,处理方式同长变换块,得到 $N = 1024$ 个 MDCT 系数.

(3)将上步骤得到的 MDCT 系数、变换块的类型以及其他相关的控制参数进行量化,并进行霍夫曼(Huffman)编码和比特流整合处理,得到一帧音频信号的

AAC 比特流.

3 基于 MPEG-AAC 的压缩域音频增强方法

结合 MPEG-AAC 编码器,本文提出的压缩域音频增强原理框图如图 3 所示.

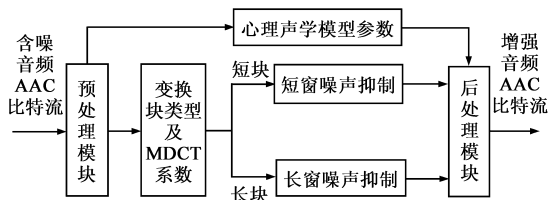


图3 基于MPEG-AAC的压缩域音频增强方法原理框图

从图 3 可知,首先是对含噪音频信号的 AAC 比特流进行预处理,即利用 AAC 解码器对其进行解码,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.如果当前块为短变换块,则采用短窗噪声抑制方法;如果是长变换块,则采用长窗噪声消除方法;如果是过渡块,其增强方法同长变换块(由于过渡块与长变换块的长度相同);最后,根据变换块类型和心理声学模型 2 相关参数,将上步骤得到的增强音频 MDCT 系数进行再量化编码处理,得到增强音频的 AAC 比特流.其中,长窗和短窗噪声抑制模块是本文算法的关键模块,详细介绍如下.

3.1 长窗噪声抑制

对于长窗噪声抑制,将人耳听觉掩蔽效应与统计模型方法相结合,提出一种基于听觉掩蔽的自适应 β -阶双曲余弦(COSH)统计模型方法.该方法的基本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,这样就可以在消除噪声的同时减少音频信号的失真.

假设 AAC 编码前的纯净音频 $x(n)$ 被加性噪声 $d(n)$ 所污染,这样可以得到长窗含噪音频信号 $y(n) = x(n) + d(n)$,则其在 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 系数的序号, $Y(k)$ 、 $X(k)$ 和 $D(k)$ 分别为长窗含噪音频、纯净音频和噪声信号的第 k 个 MDCT 系数.

根据解析 AAC 比特流得到长窗含噪音频的 MDCT 系数,可以得到长窗噪声抑制方法原理如图 4 所示.

首先,利用含噪音频的 MDCT 系数构建修正离散傅立叶变换(Modified Discrete Fourier Transform, MDFT)系数^[15];其次,利用修正的加权递归平均方法(Modified Weighted Recursive Averaging, MWRA)估计噪声功率,同时采用最小均方误差(Minimum Mean Square Error, MMSE)幅度谱估计器^[3]对构建的 MDFT 系数进行预增

强处理;然后,根据预增强的 MDFT 系数计算听觉掩蔽阈值和掩蔽参数;接下来,计算临界带子带信噪比,并利用其进行 β 参数的自适应计算;最后,结合掩蔽参数得到自适应 β -阶 COSH 幅度谱估计器,从而得到增强音频的 MDCT 系数,实现长窗噪声的抑制。

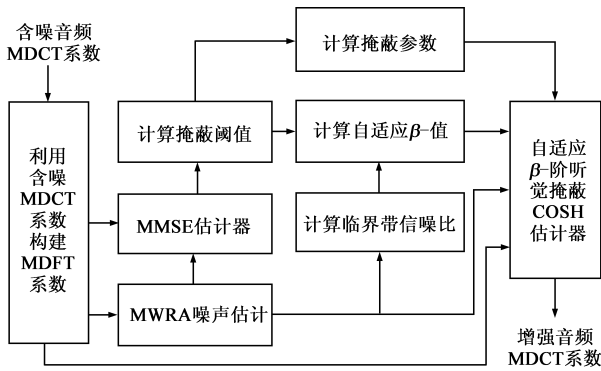


图4 长窗噪声抑制方法原理框图

3.1.1 MDFT 系数的构建

实验发现,基于 MDCT 系数得到的幅度谱估计器为类维纳滤波器形式,其增强性能较差;而基于傅立叶变换系数得到的统计模型增强方法能获得较好的增强效果.因此,这里利用 MDCT 系数来构建 MDFT 系数^[15]:首先根据 MDCT 系数 X_{mdct} 派生出修正离散正弦变换 (Modified Discrete Sine Transform, MDST) 系数 Y_{mdst} ,然后将其作为虚部,这样,可以得到 MDFT 系数 $Z_{\text{mdft}} = X_{\text{mdct}} + jY_{\text{mdst}}$.其具体构建过程如下^[15]:

(1)首先利用含噪音频的 MDCT 系数 $X_{\text{mdct}}(k)$

派生出 MDST 系数 $Y_{\text{mdst}}(k)$,表达式如下:

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

其中, k 表示 MDCT 系数的序号, N 为帧长, $m=5$ 为经验常数,用来控制 MDCT 系数到 MDST 系数的转换误差, $\theta = \pi/2N$ 表示转换相角, $idx(\cdot)$ 表示取 MDCT 系数序号的函数, $X_{\text{mdct}}^-(k)$ 和 $X_{\text{mdct}}^+(k)$ 计算式分别为:

$$X_{\text{mdct}}^-(k) = [X_{\text{mdct}}(\lambda+1, k) - X_{\text{mdct}}(\lambda-1, k)]/2 \quad (3)$$

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

其中, λ 为帧索引号.

(2)利用 $X_{\text{mdct}}(k)$ 和 $Y_{\text{mdst}}(k)$ 构建 MDFT 系数如下

$$Z_{\text{mdft}}(k) = \begin{cases} Z_-(0) - jZ_-(0), & k=0 \\ -Z_+(k-1) - jZ_-(k), & k=1, \dots, N-1 \\ -Z_+(N-1) - jZ_-(N-1), & k=N \end{cases} \quad (4)$$

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

根据文献[15]可知, MDFT 系数与 DFT 系数具有相同的特性,因此,可以将 MDFT 系数用于长窗噪声抑制方法中的噪声估计、掩蔽阈值 $T(k)$ 和 COSH 增益函数等的计算。

3.1.2 MWRA 噪声估计

音频信号中的弱信号成分(能量较小的成分)对音频质量具有非常重要的作用.然而,传统的语音增强方法通常把弱信号成分估计成噪声,这样就会造成弱信号的损失,从而导致增强音频的失真.为了减少弱信号成分的损失,本文提出一种修正的加权递归平均 (MWRA) 噪声估计方法。

对于频点 k 处的含噪音频,可以进行如下三种假设^[16]: H_0^k 、 H_1^k 和 H_2^k , 分别表示非音频信号、强音频信号和弱音频信号.根据这三种假设,得到 MWRA 噪声估计的更新准则:在 H_0^k 假设下,用一阶时域递归平均算法更新噪声功率谱;在 H_1^k 假设下,不更新噪声,直接利用前一帧的噪声功率谱代替当前帧的噪声估计;在 H_2^k 假设下,先对含噪音频功率进行加权,再用一阶递归平均方法来更新噪声功率谱,其数学表达式为^[16]:

$$\begin{cases} H_0^k: \hat{\sigma}_b^2(\lambda, k) = \alpha_d \hat{\sigma}_b^2(\lambda-1, k) + (1-\alpha_d) |Y(\lambda, k)|^2 \\ H_1^k: \hat{\sigma}_b^2(\lambda, k) = \hat{\sigma}_b^2(\lambda-1, k) \\ H_2^k: \hat{\sigma}_b^2(\lambda, k) = \alpha_d \hat{\sigma}_b^2(\lambda-1, k) + (1-\alpha_d) [W(\lambda, k) |Y(\lambda, k)|^2] \end{cases} \quad (5)$$

其中, λ 为帧索引号, k 为 MDFT 系数的序号, $\alpha_d = 0.95$ 为常数平滑因子, $Y(\lambda, k)$ 为含噪音频的 MDFT 系数, $W(\lambda, k)$ 为加权函数,其表达式如下:

$$W(\lambda, k) = \begin{cases} 1, & \gamma(\lambda, k) < \gamma_{\min} \\ \frac{\gamma_{\max} - \gamma(\lambda, k)}{\gamma_{\max} - \gamma_{\min}}, & \gamma_{\min} \leq \gamma(\lambda, k) < \gamma_{\max} \\ 0, & \gamma(\lambda, k) \geq \gamma_{\max} \end{cases} \quad (6)$$

其中, $\gamma_{\max} = 18$ 和 $\gamma_{\min} = 2$ 为经验常数, $\gamma(\lambda, k)$ 可以看作是后验信噪比,其计算式如下:

$$\gamma(\lambda, k) = \frac{S(\lambda, k)}{\hat{\sigma}_b^2(\lambda-1, k)} \quad (7)$$

其中, $S(\lambda, k)$ 为平滑的含噪音频功率谱.

根据式(5)可知,在 H_0^k 的假设下,可以看作加权函数 $W(\lambda, k) = 1$,这样,式(5)可变换为:

$$\begin{cases} H_1^k: & \hat{\sigma}_b^2(\lambda, k) = \hat{\sigma}_b^2(\lambda-1, k) \\ H_0^k, H_2^k: & \hat{\sigma}_b^2(\lambda, k) = \alpha_d \hat{\sigma}_b^2(\lambda-1, k) + (1-\alpha_d) [W(\lambda, k) |Y(\lambda, k)|^2] \end{cases} \quad (8)$$

将 $p(\lambda, k) = P(H_1^k | Y(\lambda, k))$ 表示当前帧为强音频信号的概率:若 $p(\lambda, k) = 1$,即当前帧为强音频信号,不需要进行噪声估计更新;若 $p(\lambda, k) = 0$,即当前帧可能是噪声或弱信号,需要进行噪声估计更新.因此,利用 $p(\lambda, k)$ 可简化式(8)得:

$$\begin{aligned} \hat{\sigma}_D^2(\lambda, k) = & p(\lambda, k) \hat{\sigma}_D^2(\lambda - 1, k) \\ & + [1 - p(\lambda, k)] \{ \alpha \hat{\sigma}_D^2(\lambda - 1, k) \\ & + (1 - \alpha_d) [W(\lambda, k) |Y(\lambda, k)|^2] \} \end{aligned} \quad (9)$$

进一步简化式(9),得到 MWRA 噪声估计表达式:

$$\begin{aligned} \hat{\sigma}_D^2(\lambda, k) = & \bar{\alpha}_d(\lambda, k) \hat{\sigma}_D^2(\lambda - 1, k) \\ & + (1 - \bar{\alpha}_d(\lambda, k)) [W(\lambda, k) |Y(\lambda, k)|^2] \end{aligned} \quad (10)$$

其中, $\bar{\alpha}_d(\lambda, k) = \alpha_d + (1 - \alpha_d)p(\lambda, k)$ 为平滑因子。

从上面的推导过程可以看出,在该噪声估计方法中,需要计算式(7)的 $S(\lambda, k)$ 和 $p(\lambda, k)$ 。

(1) $S(\lambda, k)$ 的计算

为了计算 $S(\lambda, k)$ 值,可以进行如下两个假设: H_0 表示当前帧不是弱信号帧和 H_1 表示当前帧为弱信号帧,这样,可得 $S(\lambda, k)$ 的计算表达式:

$$\begin{cases} H_0: S(\lambda, k) = S_f(\lambda, k) \\ H_1: S(\lambda, k) = \alpha_p S(\lambda - 1, k) + (1 - \alpha_p) S_f(\lambda, k) \end{cases} \quad (11)$$

其中, $\alpha_p = 0.2$ 为常数平滑因子, $S_f(\lambda, k)$ 为相邻频点通过加窗得到的平滑 MDFT 功率谱,其计算式为:

$$S_f(\lambda, k) = \sum_{i=-L_w}^{L_w} w(i) |Y(\lambda, k - i)|^2 \quad (12)$$

其中, $w(i)$ 为归一化的汉明窗,窗长为 $2L_w + 1$, 这里取 $L_w = 2$ 。

本文用 $\varphi(\lambda, k) = P(H_1 | Y(\lambda, k))$ 表示当前帧为弱信号帧的概率。若 $\varphi(\lambda, k) = 1$ 表示当前帧属于弱信号帧,若 $\varphi(\lambda, k) = 0$ 表示当前帧不为弱信号帧。这样,利用 $\varphi(\lambda, k)$ 对式(11)进行化简得:

$$S(\lambda, k) = \varphi(\lambda, k) \{ \alpha_p S(\lambda - 1, k) + (1 - \alpha_p) S_f(\lambda, k) \} + [1 - \varphi(\lambda, k)] S_f(\lambda, k) \quad (13)$$

整理式(13)可得:

$$S(\lambda, k) = \bar{\alpha}_p(\lambda, k) S(\lambda - 1, k) + [1 - \bar{\alpha}_p(\lambda, k)] S_f(\lambda, k) \quad (14)$$

其中, $\bar{\alpha}_p(\lambda, k) = \alpha_p \varphi(\lambda, k)$ 为平滑因子, $\varphi(\lambda, k)$ 的计算方法如下:

if $\varphi(\lambda - 1, k) = 0$

$$\varphi(\lambda, k) = \begin{cases} 1, & \text{if } S_f(\lambda, k) \leq S(\lambda - 1, k) \ \&\& \ \frac{S_f(\lambda - 1, k)}{\hat{\sigma}_D^2(\lambda - 1, k)} \\ & > \sigma_{\max} \ \&\& \ \frac{|Y(\lambda, k)|^2}{\hat{\sigma}_D^2(\lambda - 1, k)} > \sigma_{\min} \\ 0, & \text{其他} \end{cases}$$

if $\varphi(\lambda - 1, k) = 1$

$$\varphi(\lambda, k) = \begin{cases} 0, & \text{if } S_f(\lambda, k) \geq S(\lambda - 1, k) \ \parallel \ \frac{|Y(\lambda, k)|^2}{\hat{\sigma}_D^2(\lambda - 1, k)} \leq \sigma_{\min} \\ 1, & \text{其他} \end{cases} \quad (15)$$

其中, $\sigma_{\max} = 12$ 表示强音频信号阈值, $\sigma_{\min} = 1.2$ 表示噪声阈值。

(2) $p(\lambda, k)$ 的计算

根据相邻帧之间的相关性可知,如果前两帧为强音频信号,则当前帧为强音频信号的概率要大于当前帧直接被判为强音频信号的概率,即:

$$\begin{aligned} P(H(\lambda, k) = H_1^k | H(\lambda - 1, k) = H_1^k, H(\lambda - 2, k) = H_1^k) \\ > p(H(\lambda, k) = H_1^k) \end{aligned} \quad (16)$$

因此,根据式(16)的概率关系,利用条件最大后验概率准则^[17] (Conditional Maximum a Posteriori Criterion, CMAP)来计算当前帧强音频信号的存在概率 $p(\lambda, k)$, 这样,可得如下的决策规则:

$$\frac{P(H(\lambda, k) = H_1^k | S_r(\lambda, k), H(\lambda - 1, k) = H_1^k, H(\lambda - 2, k) = H_1^k) < \alpha}{P(H(\lambda, k) = H_0^k | S_r(\lambda, k), H(\lambda - 1, k) = H_1^k, H(\lambda - 2, k) = H_1^k) > \alpha} \quad (17)$$

其中, α 为常数阈值, $S_r(\lambda, k)$ 为平滑功率比:

$$S_r(\lambda, k) = \frac{S(\lambda, k)}{S_{\min}(\lambda, k)} \quad (18)$$

其中, $S_{\min}(\lambda, k)$ 为 D 帧内 $S(\lambda, k)$ 功率谱的最小值。

根据 Bayes 公式,变换式(17)可得^[17]:

$$\frac{P(S_r(\lambda, k) | H(\lambda, k) = H_1^k, H(\lambda - 1, k) = H_1^k, H(\lambda - 2, k) = H_1^k) < \alpha'_{ij}}{P(S_r(\lambda, k) | H(\lambda, k) = H_0^k, H(\lambda - 1, k) = H_1^k, H(\lambda - 2, k) = H_1^k) > \alpha'_{ij}} \quad (19)$$

其中, α'_{ij} 为检测音频是否存在的阈值,计算式如下:

$$\begin{aligned} \alpha'_{ij} = & \frac{P(H(\lambda, k) = H_0^k | H(\lambda - 1, k) = H_1^k, H(\lambda - 2, k) = H_1^k)}{\alpha \frac{P(H(\lambda, k) = H_0^k | H(\lambda - 1, k) = H_1^k, H(\lambda - 2, k) = H_1^k)}{P(H(\lambda, k) = H_1^k | H(\lambda - 1, k) = H_1^k, H(\lambda - 2, k) = H_1^k)}} \end{aligned} \quad (20)$$

其中, α'_{00} 表示前两帧都不存在音频, α'_{01} 表示前一帧不存在音频而前第二帧存在音频,同理可知 α'_{10} 、 α'_{11} 。

简化式(19)可得^[17]:

$$\frac{P(S_r(\lambda, k) | H(\lambda, k) = H_1^k)}{P(S_r(\lambda, k) | H(\lambda, k) = H_0^k)} < \alpha'_{ij}, i, j \in [0, 1] \quad (21)$$

最后根据式(21)可得到如下软决策规则^[17]:

$$S_r(\lambda, k) \stackrel{H_1^k}{<} \eta \stackrel{H_0^k}{>} \quad (22)$$

其中, η 为决策阈值,其计算表达式如下:

$$\begin{aligned} \eta = & \{ p(\lambda - 1, k) \cdot p(\lambda - 2, k) \cdot \alpha'_{00} \} \\ & + \{ p(\lambda - 1, k) \cdot [1 - p(\lambda - 2, k)] \cdot \alpha'_{10} \} \\ & + \{ [1 - p(\lambda - 1, k)] \cdot p(\lambda - 2, k) \cdot \alpha'_{01} \} \\ & + \{ [1 - p(\lambda - 1, k)] \cdot [1 - p(\lambda - 2, k)] \cdot \alpha'_{11} \} \end{aligned} \quad (23)$$

这里取 $\alpha'_{00} = 5.2, \alpha'_{01} = 1.69, \alpha'_{10} = 0.018, \alpha'_{11} = 0.002$.

这样,根据式(22)和式(23)可得强音频信号存在概率 $p(\lambda, k)$ 的计算表达式为:

$$p(\lambda, k) = \begin{cases} 1, & S_r(\lambda, k) \geq \eta \\ 0, & S_r(\lambda, k) < \eta \end{cases} \quad (24)$$

最后对 $p(\lambda, k)$ 进行一阶时域平均处理得到平滑概率 $\hat{p}(\lambda, k)$, 并将其替代式(10)中强音频信号存在概率 $p(\lambda, k)$, 用于计算平滑因子 $\bar{\alpha}_d(\lambda, k)$.

$$\hat{p}(\lambda, k) = 0.2 \cdot \hat{p}(\lambda - 1, k) + 0.8 \cdot p(\lambda, k) \quad (25)$$

其中, $\hat{p}(\lambda - 1, k)$ 为前一帧的强音频信号存在概率.

3.1.3 基于听觉掩蔽的自适应 β -COSH 增益函数

假设 MDFT 系数服从复高斯分布, 并利用加权双曲余弦(COSH)代价函数^[4]和听觉掩蔽效应得到自适应 β -阶 COSH 估计器.

加权双曲余弦(COSH)代价函数定义为^[4]:

$$d_{\beta\text{cosh}}(X_{\text{mdft-}k}, \hat{X}_{\text{mdft-}k}) = X_{\text{mdft-}k}^{\beta} \left(\frac{X_{\text{mdft-}k}}{\hat{X}_{\text{mdft-}k}} + \frac{\hat{X}_{\text{mdft-}k}}{X_{\text{mdft-}k}} - 1 \right) \quad (26)$$

其中, $X_{\text{mdft-}k}$ 表示纯净音频第 k 个频点的 MDFT 系数幅度谱; $\hat{X}_{\text{mdft-}k}$ 表示增强音频第 k 个频点的 MDFT 系数幅度谱; β 为功率指数.

根据式(26)的代价函数, 定义贝叶斯风险函数:

$$R = \int_0^{\infty} \left\{ X_{\text{mdft-}k}^{\beta} \left[\frac{X_{\text{mdft-}k}}{\hat{X}_{\text{mdft-}k}} + \frac{\hat{X}_{\text{mdft-}k}}{X_{\text{mdft-}k}} - 1 \right] \right\} p(X_{\text{mdft-}k} | Y_{\text{mdft}}(k)) dX_{\text{mdft-}k} \quad (27)$$

其中, $Y_{\text{mdft}}(k)$ 表示含噪音频的 MDFT 系数.

为了使贝叶斯风险函数 R 最小, 将式(27)对 $\hat{X}_{\text{mdft-}k}$ 求偏导, 并令其为 0 得:

$$\frac{\partial R}{\partial \hat{X}_{\text{mdft-}k}} = \int_0^{\infty} X_{\text{mdft-}k}^{\beta} \left[\frac{-X_{\text{mdft-}k}}{\hat{X}_{\text{mdft-}k}^2} + \frac{1}{X_{\text{mdft-}k}} \right] p(X_{\text{mdft-}k} | Y_{\text{mdft}}(k)) dX_{\text{mdft-}k} = 0 \quad (28)$$

求解上式可以得到 $\hat{X}_{\text{mdft-}k}$ 的估计为:

$$\hat{X}_{\text{mdft-}k} = \sqrt{\frac{\int_0^{\infty} X_{\text{mdft-}k}^{\beta+1} p(X_{\text{mdft-}k} | Y_{\text{mdft}}(k)) dX_{\text{mdft-}k}}{\int_0^{\infty} X_{\text{mdft-}k}^{\beta-1} p(X_{\text{mdft-}k} | Y_{\text{mdft}}(k)) dX_{\text{mdft-}k}}} \quad (29)$$

根据高斯统计模型, 得到 β -阶 COSH 估计器的增益函数 $G_{\beta\text{cosh}}(\xi_k, \gamma_k, \beta)$ 表达式为^[4]:

$$G_{\beta\text{cosh}}(\xi_k, \gamma_k, \beta) = \frac{\hat{X}_{\text{mdft-}k}}{Y_{\text{mdft-}k}} = \frac{1}{\gamma_k} \sqrt{\frac{v_k \Gamma\left(\frac{\beta}{2} + 1\right) \Phi\left(-\frac{\beta+1}{2}, 1, -\nu_k\right)}{\Gamma\left(\frac{\beta+1}{2}\right) \Phi\left(-\frac{\beta-1}{2}, 1, -\nu_k\right)}} \quad (30)$$

其中, $\Gamma(x)$ 为 Gamma 函数; $\Phi(x, y, z)$ 为融合超几何函数; $\hat{X}_{\text{mdft-}k}$ 表示增强音频的 MDFT 系数幅度谱 $Y_{\text{mdft-}k}$ 表示

含噪音频的 MDFT 系数幅度谱; ξ_k 为先验信噪比, γ_k 为后验信噪比, ν_k 是先验后验信噪比的函数, 其计算表达式如下:

$$\xi_k = \frac{\hat{X}_{\text{mdft-}k}^2}{\hat{\sigma}_D^2(k)}, \gamma_k = \frac{Y_{\text{mdft-}k}^2}{\hat{\sigma}_D^2(k)}, \nu_k = \frac{\xi_k}{1 + \xi_k} \gamma_k \quad (31)$$

根据人耳听觉掩蔽效应^[15]可知, 低于掩蔽阈值的成分不能被人耳所感知. 而在音频增强中, 如果消除低于掩蔽阈值的噪声成分, 则势必会去除一些音频成分, 从而造成增强音频信号的损失; 如果对那些不被人耳感知的噪声成分采用听觉掩蔽而不是消除的方式, 就可以减少增强音频信号的损失. 因此, 本文结合人耳听觉掩蔽效应提出一种基于听觉掩蔽的自适应 β -阶 COSH 幅度谱估计器. 为了避免增强音频帧间出现能量突变, 引入谱地板系数来保留一些舒适背景噪声, 即:

$$\hat{G}_{\beta\text{cosh}}(\xi_k, \gamma_k, \beta) = \begin{cases} \frac{1}{\gamma_k} \sqrt{\frac{\nu_k \Gamma\left(\frac{\beta}{2} + 1\right) \Phi\left(-\frac{\beta+1}{2}, 1, -\nu_k\right)}{\Gamma\left(\frac{\beta+1}{2}\right) \Phi\left(-\frac{\beta-1}{2}, 1, -\nu_k\right)}}, & \text{if } \gamma_k \geq [\rho_1(k) + \rho_2(k)] \\ \sqrt{\frac{\rho_2(k)}{\gamma_k}}, & \text{其他} \end{cases} \quad (32)$$

其中, $\rho_1(k)$ 和 $\rho_2(k)$ 为掩蔽参数, 其计算式如下:

$$\begin{bmatrix} \rho_1(k) \\ \rho_2(k) \end{bmatrix} = \Theta_{\mathcal{T}_h}(k) \begin{bmatrix} \rho_{1\max} - \rho_{1\min} \\ \rho_{2\max} - \rho_{2\min} \end{bmatrix} + \begin{bmatrix} \rho_{1\min} \\ \rho_{2\min} \end{bmatrix} \quad (33)$$

其中, $\rho_{1\max} = 10$ 和 $\rho_{2\max} = 0.015$ 表示两个掩蔽参数的最大值; $\rho_{1\min} = 1$ 和 $\rho_{2\min} = 0.008$ 表示两个掩蔽参数的最小值. $\Theta_{\mathcal{T}_h}(k)$ 表示听觉感知参数, 其计算式为:

$$\Theta_{\mathcal{T}_h}(k) = \frac{T(k) - \arg[\min_{0 \leq k < N} T(k)]}{\arg[\max_{0 \leq k < N} T(k)] - \arg[\min_{0 \leq k < N} T(k)]} \quad (34)$$

其中, $T(k)$ 为掩蔽阈值. 图 5 中给出了在先验信噪比 $\xi_k = 5\text{dB}$ 的条件下不同 β 值对 $G_{\beta\text{cosh}}(\xi_k, \gamma_k, \beta)$ 增益函数的影响, 并将其与 MMSE 增益函数进行比较.

从图 5 可以看出, 对于确定的 ξ_k 和 γ_k , MMSE 增益函数为一定值, 并且 MMSE 的增益函数值衰减, 会残留较多的背景噪声. 然而, 在确定的 ξ_k 和 γ_k , β -COSH 增益函数根据不同的 β 值自适应地选择合适的增益函数, 从而对噪声进行不同程度的衰减, 这样相对于固定增益函数(如 MMSE 估计器等)有更好的增强效果.

因此, 对于长窗含噪音频 MDCT 系数 $Y(k)$, 利用式(32)的增益函数对其进行噪声抑制, 得到长窗增强音频的 MDCT 系数 $\hat{X}(k)$:

$$\hat{X}(k) = \hat{G}_{\beta\text{cosh}}(\xi_k, \gamma_k, \beta) \cdot Y(k), k \in [0, 1, \dots, 1024] \quad (35)$$

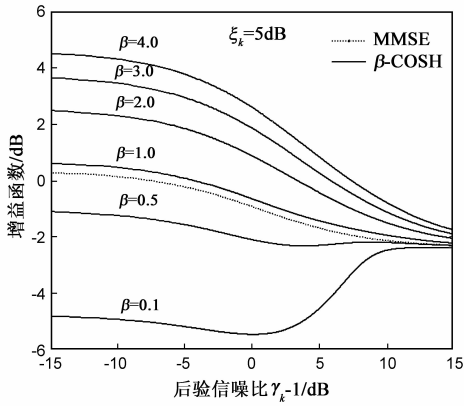


图5 不同 β -值对 $G_{\beta\text{-cosh}}(\xi_k, \gamma_k, \beta)$ 增益函数的影响以及与MMSE增益函数曲线对比图

3.1.4 β 参数的自适应计算

对于式(32)增益函数中 β 参数的计算,考虑其依赖于临界带子带信噪比 $\Xi(b)$;另外,为了在噪声消除和音频失真之间进行折中,本文还考虑 β 值还依赖于掩蔽阈值 $T(k)$.因此,将 β 值看作是临界带子带信噪比 $\Xi(b)$ 和听觉感知参数 $\Theta_m(k)$ 的函数 $F(\cdot, \cdot)$,该函数为两个变量的多项式^[5],即:

$$\begin{aligned} \beta(b, k) &= F(\Xi(b), \Theta_m(k)) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{ij} \Xi(b)^i \Theta_m(k)^j \\ &= c_{00} + c_{10} \Xi(b) + c_{01} \Theta_m(k) \\ &\quad + c_{11} \Xi(b) \Theta_m(k) + \Theta(O_h) \end{aligned} \quad (36)$$

其中, $b(0 \leq b < 25)$ 为临界带索引号; k 表示频点索引; $\Theta(O_h)$ 表示高阶多项式项; c_{ij} 表示多项式系数.

针对上式,省略其高阶多项式 $\Theta(O_h)$,并用 $\tau_i (i = 0, 1, 2, 3)$ 代替 c_{ij} ,可改写式(36)为:

$$\beta(b, k) = \tau_0 + \tau_1 \Xi(b) + \tau_2 \Theta_m(k) + \tau_3 \Xi(b) \Theta_m(k) \quad (37)$$

其中,前四项多项式系数 $\tau_i (i = 0, 1, 2, 3)$ 为经验常数,通过大量实验发现它们分别取0.542, 0.121, 0.581和0.147时能获得较好的语音增强效果.临界带子带信噪比 $\Xi(b)$ 的计算式如下:

$$\Xi(b) = 10 \log_{10} \frac{\sum_{k=B_{\text{low}}(b)}^{B_{\text{up}}(b)} |Y_{\text{mdft-}k} - \hat{\sigma}_D(k)|^2}{\sum_{k=B_{\text{low}}(b)}^{B_{\text{up}}(b)} \hat{\sigma}_D^2(k)} \quad (38)$$

其中, $B_{\text{up}}(b)$ 和 $B_{\text{low}}(b)$ 分别表示临界子带 b 所对应的频率上界和下界; $\hat{\sigma}_D^2(k)$ 为估计的噪声功率; $Y_{\text{mdft-}k}$ 为含噪声音频的MDFT系数幅度谱.

为了防止 β 值的过大和过小导致过多的噪声残留和音频损失,从而需要对 β 值的取值范围进行限制.实验发现 β 值在0.1~4范围内增强效果较好:

$$\beta(b, k) = \max\{\min[\beta(b, k), 4], 0.1\} \quad (39)$$

最后,为了保证增强音频在相邻临界带之间的自然过渡,我们在相邻临界带之间对 $\beta(b, k)$ 进行一阶平滑处理得:

$$\hat{\beta}(b, k) = \zeta \cdot \hat{\beta}(b-1, k) + (1 - \zeta) \cdot \beta(b, k) \quad (40)$$

其中, $\zeta = 0.4$ 是平滑常数.

3.2 短窗噪声抑制

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:

$$Y_s(k) = X_s(k) + D_s(k), k \in [0, 1, \dots, 128] \quad (41)$$

其中, k 表示短窗子帧MDCT系数频点索引, $Y_s(k)$ 、 $X_s(k)$ 和 $D_s(k)$ 分别为子帧第 k 个含噪音频、纯净音频和噪声的MDCT系数.

根据解析短窗AAC比特流得到的8个子帧含噪音频MDCT系数,得到短窗噪声抑制流程如图6所示.

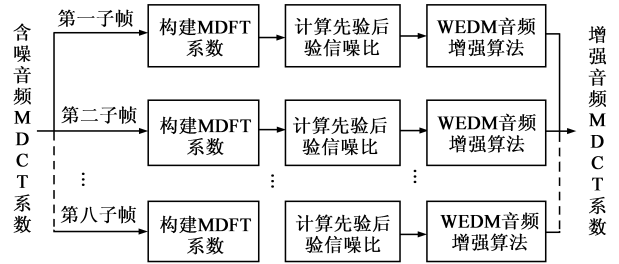


图6 短窗噪声抑制流程图

根据图6可知,对于每个子帧,首先根据长窗的噪声估计得到短窗各子帧的噪声估计;其次是利用每子帧的MDCT系数构建MDFT系数,并根据MDFT系数计算子帧含噪音频功率谱;然后根据子帧的噪声估计和含噪音频功率谱计算先验信噪比和后验信噪比;最后采用加权欧式距离测度(Weighted Euclidean Distortion, WEDM)估计器^[4]对各个子帧进行增强处理,得到各个子帧增强音频的MDCT系数.短窗噪声抑制的具体步骤如下:

(1)首先是利用长窗噪声估计计算短窗各子帧的噪声估计.由于长窗帧长 $N = 1024$,短窗子帧长 $N_s = 128$,所以需要长窗 N 个频点的噪声估计映射成短窗 N_s 个频点的噪声估计.本文采用加一个窗长为 $L = 8$ 的归一化汉明窗的方式来完成长窗到短窗噪声估计 $\lambda_{\text{sub-frame}}(k)$ 的映射,映射表达式如下:

$$\lambda_{\text{sub-frame}}(k) = \sum_{i=0}^{L-1} w(i) \hat{\sigma}_D^2(k \cdot M + i), k \in [0, 1, \dots, N_s] \quad (42)$$

其中, $\hat{\sigma}_D^2(k)$ 为长窗的噪声估计值; $w(i)$ 为归一化汉明窗函数。

(2) 利用子帧 λ_s 的 MDCT 系数构建 MDFT 系数 $Y_{\text{mdft-s}}(\lambda_s, k)$, 并根据其计算子帧含噪音频功率谱。

(3) 利用直接判决^[4] (Decision-Directed, DD) 先验信噪比方法计算先验信噪比 $\xi(\lambda_s, k)$, 并利用子帧噪声估计和含噪音频功率谱计算后验信噪比 $\gamma(\lambda_s, k)$ 如下:

$$\xi(\lambda_s, k) = \alpha \frac{|\hat{X}_{\text{mdft}}(\lambda_s - 1, k)|^2}{\lambda_{\text{sub-frame}}(k)} + (1 - \alpha) \max\{\gamma(\lambda_s, k) - 1, \xi_{\min}\} \quad (43)$$

$$\gamma(\lambda_s, k) = \frac{|Y_{\text{mdft-s}}(\lambda_s, k)|^2}{\lambda_{\text{sub-frame}}(k)} \quad (44)$$

其中, k 为子帧频点索引; $\alpha = 0.98$ 为常数平滑因子; $\xi_{\min} = -15\text{dB}$ 为先验信噪比的最小值。

(1) 根据 WEDM 估计器^[4] $G_{\text{wedm}}(\xi_k, \gamma_k)$ 对各个子帧的含噪音频 MDCT 系数 $Y_s(k)$ 分别进行增强处理, 得到各个子帧增强音频的 MDCT 系数:

$$\hat{X}_s(k) = G_{\text{wedm}}(\xi_k, \gamma_k) \cdot Y_s(k), k \in [0, 1, \dots, 128) \quad (45)$$

(2) 整合 8 个子帧的增强音频 MDCT 系数, 得到短窗增强音频 1024 个 MDCT 系数, 从而实现短窗噪声的抑制。

4 性能评价

为了验证本文方法的有效性, 将分别评价噪声估计和音频增强性能。

本文所用的测试数据取自 3GPP 和 MPEG 标准测试序列, 采样率为 44.1kHz, 共计 32 段不同风格的音频数据, 包括大提琴、鼓、铜钹、吉他、交响乐和摇滚乐等音频类别。对这些纯净测试音频序列首先采用 ITU-T 的 P.56 标准添加各种噪声, 噪声类型分别为 Babble 噪声、Office 噪声、Street 噪声和 Volvo 噪声, 信噪比分别为 6dB、12dB 和 18dB; 然后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 96kbps 比特率为例进行测试; 接下来利用本文算法对 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, 并将其与参考算法的增强音频进行比较。

4.1 噪声估计方法的性能评价

将文献[18]提出的最小值控制递归平均 (Minima Controlled Recursive Averaging, MCRA) 噪声估计算法^[18]作为参考算法, 与本文算法进行比较, 并采用均方误差 (Mean Square Error, MSE) 准则来衡量噪声估计算法的性能, 其计算式如下:

$$\epsilon_{\text{MSE}} = \frac{1}{L} \sum_{\lambda=0}^{L-1} \frac{\sum_{k=0}^{N-1} [\hat{\sigma}_D^2(\lambda, k) - \sigma_D^2(\lambda, k)]^2}{\sum_{k=0}^{N-1} [\sigma_D^2(\lambda, k)]^2} \quad (46)$$

其中, k 为频点索引号, λ 为帧索引号, L 为测试音频信号的总帧数, N 为帧长, $\sigma_D^2(\lambda, k)$ 和 $\hat{\sigma}_D^2(\lambda, k)$ 分别表示原始噪声功率谱与估计的噪声功率谱。

从式(46)可以看出, 若 ϵ_{MSE} 越大, 则说明噪声估计越不准确; 反之, ϵ_{MSE} 越小, 则说明噪声估计算法性能越好。表 1 给出了本文噪声估计方法与参考算法的 MSE 比较结果。

表 1 噪声估计算法性能比较

噪声类型	信噪比	WCRA 方法	本文方法
Babble 噪声	6dB	1.7127	1.3796
	12dB	2.2609	1.8652
	18dB	3.0609	2.4975
Office 噪声	6dB	1.1806	0.9206
	12dB	2.1345	1.4394
	18dB	2.6354	1.7485
Street 噪声	6dB	1.3555	1.0588
	12dB	2.268	1.850
	18dB	2.946	2.352
Volvo 噪声	6dB	1.3909	1.0761
	12dB	2.1597	1.8465
	18dB	2.5458	2.3746

从表 1 可以看出, 在不同噪声类型和不同信噪比情况下, 本文噪声估计算法的 MSE 值都小于参考算法, 从而可以说明本文噪声估计算法性能优于参考噪声估计算法。

4.2 压缩域增强方法的性能评价

对于本文压缩域音频增强方法的性能评价, 将文献[4]提出的 WEDM 算法和文献[6]提出的改进的 β 阶 WEDM 算法 (简称为 I- β -WEDM) 作为参考算法, 并分别从音频频谱图分析、平均分信噪比 segSNR 和 ITU-R BS.1387-1 标准中的音频质量感知评估^[19] (Perceptual Evaluation of Audio Quality, PEAQ) 三个方面与参考算法进行比较。

4.2.1 音频频谱图分析

图 7 给出了原始音频, 含 12dB 街道 (Street) 噪声的含噪音频以及三种增强音频的频谱。

从图 7 可以看出, WEDM 算法和 I- β -WEDM 算法的增强音频损失较大, 增强音频效果较差; 本文算法不仅有效的消除了噪声, 而且音频损失较小, 听觉效果得到明显改善。

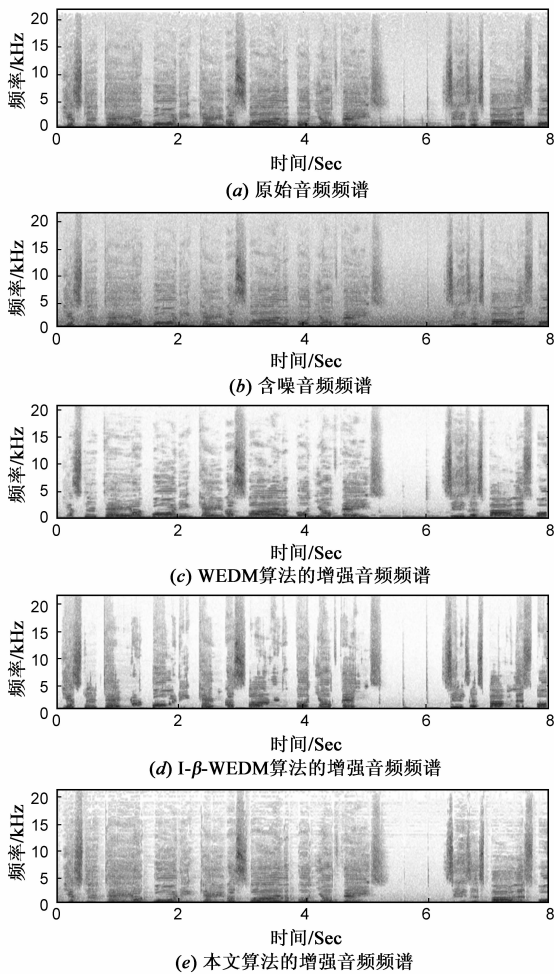


图7 原始音频、含噪音频和三种算法增强音频频谱对比图

4.2.2 平均分信噪比 segSNR 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:

$$SNR_{seg} = \frac{1}{L} \sum_{l=0}^{L-1} 10 \cdot \log_{10} \left(\frac{\sum_{n=M}^{M+N-1} x^2(n)}{\sum_{n=M}^{M+N-1} [x(n) - y(n)]^2} \right) \quad (47)$$

其中, 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, $y(n)$ 表示增强的时域音频信号.表 2 给出了各增强方法分段信噪比提高的结果.

从表 2 可以看出,本文算法的分段信噪比提高程度明显优于其他两种参考算法,说明本文算法比参考算法具有更好的噪声抑制性能.

4.2.3 PEAQ 测试

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^[19] (Objective Difference Grade, ODG),它是一种模拟人耳主观听觉的客观

测试方法.当 ODG 分增加 0.1 时,说明音频质量得到显著改善.采用 ODG 分的提高 (ODG Improvement, ODGI) 来衡量增强音频的质量改善情况,其测试结果如表 3 所示.

从表 3 可以看出,本文算法得到增强音频 ODG 提高程度明显好于其他两种增强算法,从而可以说明本文算法的增强性能优于参考算法.

表 2 分段信噪比提高结果/dB

噪声类型	信噪比	WEDM 方法	$1-\beta$ -WEDM 方法	本文方法
Babble 噪声	6dB	2.023	2.326	5.627
	12dB	2.145	2.421	5.903
	18dB	2.934	3.043	6.653
Office 噪声	6dB	2.563	2.734	6.291
	12dB	2.616	2.798	6.320
	18dB	3.267	3.361	6.817
Street 噪声	6dB	2.470	3.016	6.634
	12dB	2.811	3.033	6.766
	18dB	3.267	3.362	6.817
Volvo 噪声	6dB	2.375	2.561	6.476
	12dB	2.669	2.748	6.652
	18dB	3.262	3.351	7.038

表 3 客观音频质量测试 ODGI 得分

噪声类型	信噪比	WEDM 方法	$1-\beta$ -WEDM 方法	本文方法
Babble 噪声	6dB	0.13	0.18	0.32
	12dB	0.11	0.15	0.29
	18dB	0.09	0.12	0.24
Office 噪声	6dB	0.12	0.16	0.34
	12dB	0.10	0.14	0.31
	18dB	0.07	0.11	0.27
Street 噪声	6dB	0.12	0.17	0.33
	12dB	0.11	0.15	0.31
	18dB	0.08	0.11	0.26
Volvo 噪声	6dB	0.12	0.16	0.35
	12dB	0.11	0.13	0.30
	18dB	0.09	0.10	0.25

5 总结

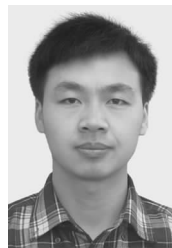
本文提出了一种基于 MPEG-AAC 编码器的压缩域音频增强方法.该方法的输入输出均为音频信号的 AAC 比特流.在压缩域,针对 AAC 的长变换块类型,首先提出了一种改进的加权递归平均方法 (MWRA) 来估计噪声功率;然后根据人耳听觉掩蔽效应,提出了一种

基于听觉掩蔽的自适应 β -阶 COSH 统计模型增强方法,消除长窗 AAC 比特流中的各种噪声.另外,针对 AAC 的短变换块类型,本文利用 WEDM 增强算法消除短窗 AAC 比特流中的各种噪声.测试结果表明,在 96kbps 码率下,本文所提算法能有效去除音频信号 AAC 比特流中的噪声,并且噪声抑制性能和增强音频的听觉质量都优于参考算法,进而说明了本文方法的有效性.由于本文方法是根据解码 AAC 比特流得到的 MDCT 系数消除各种噪声的,所以从理论上讲,AAC 比特率的大小对本文方法的性能影响不大,即对各种 AAC 比特率均适用.接下来的工作需要进一步验证其他比特率下的噪声消除能力.

参考文献

- [1] S F Boll. Suppression of Acoustic Noise in Speech Using Spectral Subtraction[J]. IEEE Transactions on Acoustics, Speech, & Signal Processing, 1979, 27(2): 113 – 120.
- [2] J F Li, S Sakamoto, S Hongo, et al. Adaptive β -order Generalized Spectral Subtraction for Speech Enhancement[J]. Signal Process, 2008, 88(11): 2764 – 2776.
- [3] R Miyazaki, H Saruwatari, T Inoue, et al. Musical-noise-free speech enhancement based on optimized iterative spectral subtraction[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(7): 2080 – 2094.
- [4] P Loizou. Speechenhancement based on perceptually motivated bayesian estimators of the magnitude spectrum[J]. IEEE Transactions on Speech Audio Process, 2005, 13(5): 857 – 869.
- [5] C H You, et al. Masking-based β -order MMSE speech enhancement[J]. Speech Communication, 2006, 48(1): 57 – 70.
- [6] N Li, C C Bao, et al. Adaptive β -order WEDM spectral amplitude estimator for speech enhancement[A]. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)[C]. Bilbao, Spain: IEEE, 2011. 155 – 160.
- [7] D L Donoho. De-noising by Soft-thresholding [J]. IEEE Transactions on Information Theory, 1995, 41(3): 613 – 627.
- [8] Y T Lin, J L Cai. A new threshold function for signal denoising based on wavelet transform[A]. International Conference on Measuring Technology and Mechatronics Automation (ICMT-MA)[C]. Changsha City: IEEE, 2010. 200 – 203.
- [9] Y Ghanbari, M R Karami-Mollaei. A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets[J]. Speech Communication, 2006, 48(8): 927 – 940.
- [10] I Y Soon, et al. Noisy speech enhancement using discrete cosine transform[J]. Speech Communication, 1998, 24(3): 249 – 257.
- [11] P C Loizou, B Chen. A Laplacian-based MMSE estimator for speech enhancement [J]. Speech Communication, 2007, 49(2): 134 – 143.
- [12] S Salahuddin, S Z Al Islam, Md K Hassan, et al. DCT speech enhancement with hard and soft thresholding criteria[J]. Electronics Letters, 2002, 38(13): 669 – 670.
- [13] P N Le, et al. An improved soft threshold method for DCT speech enhancement[A]. Second International Conference on Communications and Electronics[C]. Hoi an: IEEE, 2008. 268 – 271.
- [14] ISO/IEC 13818 – 7. Information Technology-Generic Coding of Moving Pictures and Associated Audio Information-part 7: Advanced Audio Coding[S]. 1997.
- [15] F Deng, 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 (WCSP)[C]. Nanjing, China, 2011. 1 – 5.
- [16] 冯炎. 基于噪声估计和先验信噪比估计的语音增强算法研究[D]. 成都:西南交通大学, 2010. 11 – 37.
Y Feng. Speech Enhancement method based on noise estimation and priori SNR estimation[D]. Chengdu: Southwest Jiaotong University, 2010. 11 – 37. (in Chinese)
- [17] J M Kum, J H Chang. Speechenhancement based on minima controlled recursive averaging incorporating second-order conditional MAP criterion [J]. IEEE Signal Processing Letters, 2009, 16(7): 624 – 627.
- [18] I Cohen, B Berdugo. Noiseestimation by minima controlled recursive averaging for robust speech enhancement [J]. IEEE Signal Processing Letters, 2002, 9(1): 12 – 15.
- [19] ITU-R Rec. BS. 1387 – 1. Method for Objective Measurements of Perceived Audio Quality[S]. 2001.

作者简介



邓峰 男. 1983 年 12 月出生于湖南郴州. 2012 年进入北京工业大学攻读博士学位, 主要研究方向为语音和音频增强.
E-mail: dengfeng@emails.bjut.edu.cn

鲍枫 男. 1990 年 6 月出生于吉林长春. 2012 年进入北京工业大学攻读硕士学位, 主要研究方向为语音和音频增强.
E-mail: baofeng@emails.bjut.edu.cn

鲍长春 男. 1965 年 6 月出生于内蒙古赤峰. 博士, 北京工业大学教授、博士生导师, IEEE 高级会员, 国际语音通信学会 (ISCA) 会员, 亚太信号与信息处理学会 (APSIPA) 会员, 中国电子学会理事, 中国声学学会理事, 中国电子学会信号处理分会委员. 主要研究方向为语音与音频信号处理.
E-mail: baochch@bjut.edu.cn