

文章编号: 1001 - 893X(2012)04 - 0534 - 05

MPEG - 4 AAC 音频编码中量化模块的改进*

刘红甫, 张 聪, 梁瑞凡

(武汉工业学院 数学与计算机学院, 武汉 430023)

摘 要: MPEG - 4 AAC 音频编码标准中的量化过程采用了 Brandenburg 提出的双循环模式, 取得了较好的编码质量和压缩比例。但在实际编码时, 该方法由于迭代次数过多, 会出现收敛速度较慢, 甚至死锁的情况。在深入分析 MPEG - 4 AAC 量化算法的基础上, 提出了一种快速计算每个比例因子频带的改进算法。该算法通过有效减少外循环计算过程, 并确保每个比例因子频带的量化失真低于允许值, 从而减少量化模块的计算量。实验结果表明, 在不影响音频编码质量的前提下, 该算法能有效地提高编码效率。

关键词: 音频编码; 量化模块; 比例因子; 编码效率

中图分类号: TN911 文献标志码: A doi: 10.3969/j.issn.1001-893x.2012.04.022

Improvement of Quantization Module in MPEG - 4 AAC Audio Encoding

LIU Hong-fu, ZHANG Cong, LIANG Rui-fan

(Mathematics and Computer College, Wuhan Polytechnic University, Wuhan 430023, China)

Abstract: The quantization process of MPEG - 4 AAC audio coding standard uses double circulation patterns proposed by Brandenburg and a better coding quality and compression ratio can be obtained. But in the process of actual coding, because of too many iteration times of the method, there will be slower convergence speed, and even deadlock situations. Based on the analysis of MPEG - 4 AAC quantization algorithm, this paper proposes a fast improved algorithm which can calculate frequency band of each scale factor. By reducing the process of the external circulation effectively and ensuring the quantization distortion of each scale factor band below the allowable value, the algorithm reduces the calculation of the quantization module. The experiment results show that the algorithm can improve the coding efficiency effectively without affecting the audio coding quality.

Key words: audio encoding; quantization module; scale factor; coding efficiency

1 引 言

音频编码是数字音频广播和互联网等多媒体应用的关键技术之一。作为有损变换域编码的一种,

感知音频编码器是目前主流编码器, 其利用人耳的听觉感知特性, 对编码噪声进行感知整形, 以较高的编码压缩比获得感知质量下较少的编码效果^[1]。其中 MPEG - 4 AAC^[2] 作为一种重要的主流感知音频

* 收稿日期: 2011 - 11 - 15; 修回日期: 2012 - 02 - 27

基金项目: 国家自然科学基金重点项目(60832002); 湖北省教育厅科研重点项目(D20101704); 武汉工业学院 2009 年研究生创新基金项目(09cx026)

Foundation Item: The National Natural Science Foundation of China(No.60832002); Key Project of Educational Commission of Hubei Province of China(No.20101704); 2009 Graduate Innovation Fund of Wuhan Polytechnic University(09cx026)

编码器,提供了非常高效的音频压缩算法,在移动通信、数字广播、互联网等领域得到了广泛应用。

MPEG-4 AAC 音频编码性能很大程度上取决于量化模块中有限比特的利用效率和运算复杂度,前者影响音频感知质量,后者决定编码的实时性。MPEG-4 AAC 的量化模块采用的是双循环搜索进行失真率控制,其中外循环控制量化失真,内循环控制比特率。该算法已被证实运算量大,并且比特分配非最优判决,尤其在低码率时,造成 MPEG-4 AAC 编码性能偏低^[3]。

MPEG-4 AAC 并未把量化模块的优化方案标准化,近年来产生了很多与此相关的新技术。比如文献[4]提出了一种基于子带量化噪声统计模型的量化技术,文献[5]提出了一种动态谱线分组减少量化噪声的技术,这些技术在一定程度上优化了量化模块的计算复杂度。

在上述研究的基础上,通过对量化模块原理的研究,发现在量化模块迭代循环过程中,外层循环通过不断地调整比例因子频带的比例因子,使最终计算的量化失真低于允许值。在外层循环开始计算之前,每个频带的比例因子都初始化为 0,这个值与最终的比例因子值相差比较大,这就造成了循环次数过多。因此,找到一个与最终值相差比较接近的值进行初始化,这样就可以减少循环迭代过程。本文通过对外层循环和量化失真计算公式的研究,提出了一种能减少迭代循环过程的算法。

2 MPEG-4 AAC 量化模块的计算

MPEG-4 AAC 的量化模块整体上依然沿用了 Brandenburg 与 1987 年在 OCF 中提出的双循环结构^[6]。内层循环控制编码时所需的比特数,当比特数超出了分配的限量,就统一地加大所有量化阶(即加大统一缩放因子),务求降低比特数,得到所需的最小量化阶;而外层循环控制量化误差,它对经过循环量化后的频谱进行量化误差检测,当某一个子带的量化误差能量超过掩蔽门限,就减少相应的量化阶(即加大子带缩放因子)且步长为 1,然后再进入内循环。

MPEG-4 AAC 音频编码的量化模块外层循环流程图如图 1 所示。

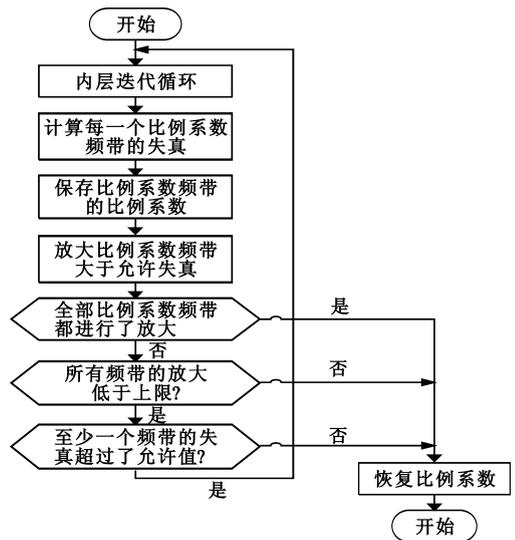


图 1 外层循环流程图

Fig.1 The flow chart of outer circulation

在 MPEG-4 AAC 音频编码过程中,音频数据经过 MDCT 变换实现由时域到频域的映射,得到音频的频谱;同时,音频数据也经过心理声学模块计算得到音频数据的掩蔽阈值,然后这些数据被传送到量化模块进行迭代计算。

在进入双层迭代循环之前,所有变量均需重置。首先计算全局增益的初始值,这样全部量化后的 MDCT 频谱数据能够经无噪声编码加入到比特流中,计算公式如下所示:

$$start_common_scalefac = ceiling\left(\frac{16}{3} \text{lb} \frac{max_mdct_line^{\frac{3}{4}}}{MAX_QUANT}\right) \quad (1)$$

其中, max_mdct_line 是当前数据帧中谱系数值最大的一个; $ceiling()$ 是正无穷方向舍入最近整数的函数; MAX_QUANT 是可以编码如比特流的最大量化值,定义为 8 191。在迭代过程中,全局公共缩放因子 $common_scalefac$ 不能小于初始值 $start_common_scalefac$ 。然后,必须将所有比例因子全部置零。

音频频谱的量化公式如下所示:

$$x_quant = int\left(\left(|mdct_line(i)| \times 2^{\frac{sf_decoder - SF_OFFSET}{4}}\right)^{\frac{3}{4}} + MAGIC_NUMBER\right) \quad (2)$$

$MAGIC_NUMBER$ 定义为 0.405 4, SF_OFFSET 定义为 100, $mdct_line$ 是 MDCT 计算出来的某个频谱值。

为了在迭代循环里使用,比例因子“ $sf_decoder$ ”分成两个变量:

$$sf_decoder = common_scalefac - scalefactor + SF_OFFSET \quad (3)$$

除非输入的频谱数据全部为零,否则都要经过量化双循环结构的处理。外层迭代循环为量化噪声控制循环,负责控制在内层迭代循环中由于对频谱数据进行量化而引起的量化噪声,如下式所示:

$$mdct_scaled(i) = |mdct_line(i)|^{\frac{3}{4}} \times 2^{\frac{16}{3} \times scalefactor(sb)} \quad (4)$$

在进行量化之前,将各谱线与其所属的比例因子频带的比例因子相乘,完成所谓的噪声染色(噪声谱的不均匀化),即噪声能量在不同频率段上的重新分配。如图1所示,外层循环要先调用内层循环,内层循环包含三大元素,即经由相应比例因子频带的比例因子缩放的谱系数 $mdct_scaled(i)$,全局公共缩放因子 $common_scalefac$ 以及当前可用于编码的比特数。

首先,计算 MDCT 谱系数的量化值:

$$x_quant(i) = \text{int}(mdct_scaled(i) \times 2^{-\frac{3}{16}common_scalefac} + 0.4054) \quad (5)$$

然后,计算在当前条件下实际编码所需的比特数,如果按照 ISO 标准规定,取全局公共缩放因子的初始值为 $start_common_scalefac$,则在第一轮循环之后,所需比特数通常要大于可用的比特数。因此,在第一次循环后,我们往往需要对全局公共缩放因子 $common_scalefac$ 作出相应的调整,即以某一量化阶为基础逐渐增大因子的取值,重新计算所需的比特数,直到满足条件为止,便跳出内层比特率控制循环。

接下来就要计算各比例因子频带的量化噪声,如下式所示:

$$error_energy(sb) = \sum_i (|mdct_line(i)| - x_quant(i))^{\frac{4}{3}} \times 2^{-\frac{1}{4}(scalefactor(sb) - common_scalefac)^2} \quad (6)$$

如若某一比例因子频带的量化噪声超出该子带的掩蔽阈值 $x_{min}(sb)$,便对该比例因子加 1 调整,对应子带的所有谱系数须根据上式重新计算。

通常情况下,若所有比例因子频带的量化噪声都不超过上限值(即掩蔽阈值),便可跳出外循环。然而并非所有场合下都能达到这一条件,这时有必要引入如下的外循环异常终止条件,即所有的比例因子频带都已经被放大过,或相邻的两个比例因子频带的比例因子之差超过了 60。倘若是由于上述两个原因使得循环被终止,则需恢复先前循环过程中存储的比例因子,以得到正确的输出。

3 MPEG-4 AAC 量化模块的改进

从理论上说,标准算法中针对于量化模块环节所提出的内、外嵌套的循环结构是一种近似最优的算法结构,通过内循环和外循环反复地交替工作,使得编码器在比特数使用和噪声控制方面取得最佳平衡。但是,这样的最佳平衡是以较高的运算复杂度和较多的循环迭代次数为代价的。在实际应用中,我们自然不希望编码器耗费太多的功耗,浪费太多的运算时间。幸运的是,AAC 标准仅仅给出了一个编、解码的参考算法标准,设计者完全可以根据自己的需要,在不改变其算法基本框架的基础上做出一定程度的优化和改进工作,而且可以肯定地说,这方面工作的上升空间很大。目前,世界上已有不少机构和学者对其提出了自己的优化、改进算法和结构,并达到了预期的结果。

通过前一节的分析我们不难发现:在外循环过程中,主要是用公式(6)来计算量化后的量化噪声,并要求所得的值不能超过该子带的掩蔽阈值 $x_{min}(sb)$,否则要进行循环,直到满足要求为止。通过研究该计算公式以及它与整个量化模块的关系,寻求一个能使计算得到的量化噪声低于该子带的掩蔽阈值 $x_{min}(sb)$ 的方法,减少该计算过程。

AAC 标准中对于谱线的非均匀量化公式为

$$x_quanted = \text{int}(|mdct_line(i)|^{\frac{3}{4}} \times 2^{\frac{3}{16}(scalefactor(sb) - common_scalefac)} + 0.4054) \quad (7)$$

其中, $scalefactor$ 即为各比例因子频带的比例因子,而 $common_scalefac$ 则为所有尺度因子带所共享的公共缩放因子。不难看出,量化步长的大小是由各比例因子频带的比例因子和所有比例因子频带的公共缩放因子所决定。为了方便化简,我们需要对公式进行取整操作,即设取整误差为 e ,同时令 $q = 0.4054 + e$,于是可将量化公式化简为

$$x_quanted = |mdct_line(i)|^{\frac{3}{4}} \times 2^{\frac{3}{16}(scalefactor(sb) - common_scalefac)} + q \quad (8)$$

$$x_quanted - q = |mdct_line(i)|^{\frac{3}{4}} \times 2^{\frac{3}{16}(scalefactor(sb) - common_scalefac)} \quad (9)$$

$$\frac{x_quanted - q}{2^{\frac{3}{16}(scalefactor(sb) - common_scalefac)}} = |mdct_line(i)|^{\frac{3}{4}} \quad (10)$$

$$|mdct_line(i)| = \left(\frac{x_quant(i) - q}{2^{\frac{3}{4}(scalefactor(sb) - common_scalefac)}}} \right)^{\frac{4}{3}} \quad (11)$$

将式(11)代入式(6),经过化简得:

$$error_energy(sb) = \sum_i \left(((x_quant(i) - q)^{\frac{4}{3}} - x_quant(i)^{\frac{4}{3}}) \times 2^{-\frac{1}{4}(scalefactor(sb) - common_scalefac)} \right)^2 \quad (12)$$

由于 $q = 0.4054 + e$, 则 $0.4054 \leq q \leq 1$ 。所求得的 $error_energy(sb)$ 的值稍微放大一些可以表示为

$$error_energy(sb) = \sum_i \left(((x_quant(i) - 1)^{\frac{4}{3}} - x_quant(i)^{\frac{4}{3}}) \times 2^{-\frac{1}{4}(scalefactor(sb) - common_scalefac)} \right)^2 \quad (13)$$

又由于要求量化噪声要低于该子带的掩蔽阈值 $x_{min}(sb)$, 即

$$error_energy(sb) \leq x_{min}(sb) \quad (14)$$

所以

$$\sum_i \left(((x_quant(i) - 1)^{\frac{4}{3}} - x_quant(i)^{\frac{4}{3}}) \times 2^{-\frac{1}{4}(scalefactor(sb) - common_scalefac)} \right)^2 \leq x_{min}(sb) \quad (15)$$

经过简化

$$common_scalefac - scalefactor(sb) \leq \frac{2 \times \lg \frac{x_{min}(sb)}{\sum_i \left(((x_quant(i) - 1)^{\frac{4}{3}} - x_quant(i)^{\frac{4}{3}}) \right)^2}}{\lg 2} \quad (16)$$

由量化模块的外层循环流程图(如图 1)可知,首先进行的是内层循环,内层循环结束后我们可以得到满足要求所有比例因子频带的公共缩放因子 $common_scalefac$ 值,由公式(16)可知,在 $common_scalefac$ 一定的前提下,等号成立时 $scalefactor(sb)$ 的取值是最大的,即该比例因子频带的比例因子最大,使得最终各个比例因子频带的压缩比达到最大值。因此,外层循环可以利用公式(16)和(17)直接求出各比例因子频带的最佳比例因子。计算公式中的变量 $x_{min}(sb)$ 在进入循环过程前都已经得到, $x_quant(i)$ 是内层循环计算得到的最终结果,这些使得计算 $scalefactor$ 的计算更加简单和准确。

$$scalefactor(sb) = common_scalefac - \frac{2 \times \lg \frac{x_{min}(sb)}{\sum_i \left(((x_quant(i) - 1)^{\frac{4}{3}} - x_quant(i)^{\frac{4}{3}}) \right)^2}}{\lg 2} \quad (17)$$

MPEG-4 AAC 量化模块的改进流程图如图 2 所示。

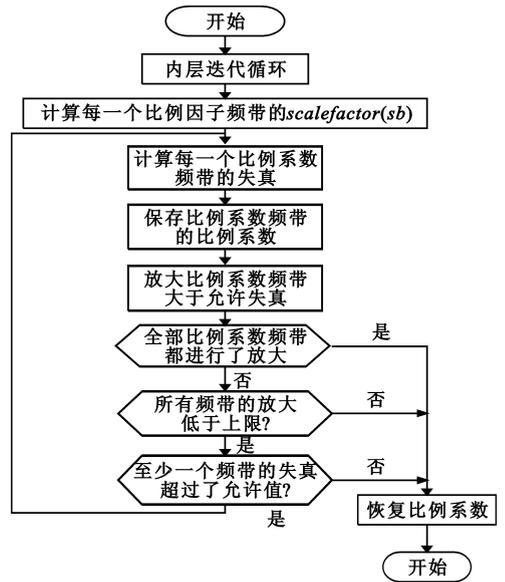


图 2 改进的外层循环流程图

Fig.2 The improved flow chart of outer circulation

在外层循环过程中,通过内层循环传出的量化频谱,以及在迭代循环前所求得的允许的最小掩蔽阈值能量来计算每个比例因子频带的比例因子。这时候所求得各频带的比例因子其实已经是最佳的数值,从而减少了外层循环的迭代次数,节省了编码时间。

4 实验结果

实验采用开源 FAAC 编码器,该编码器是 MPEG-4 AAC 参考代码的改进版本且已广泛应用于互联网的音频。实验是在低复杂度框架下,对 48 kHz、16 bit 采样的 MPEG 双声道音频测试序列进行测试。用 FAAC 编码器和改进的 FAAC 编码器对同一组音频数据进行编码处理,计算编码所需要的时间,实验结果如表 1 所示。

表 1 音频测试序列编码时间

Table 1 The encoding time of audio test sequence

测试序列	文件大小 /Mbyte	FAAC 编码 时间/s	改进的编码 时间/s	节省编码 时间/s
Es01	1.96	9.8	8.9	0.9
Es02	1.57	6.0	5.4	0.6
Es03	1.39	6.2	6.1	0.1
Sc01	2.00	11.9	10.6	1.3
Sc02	2.33	12.9	12.1	0.8
Sc03	2.11	12.3	11.7	0.5
Si01	1.46	7.4	6.6	0.8
Si02	1.41	7.2	6.2	1.0
Si03	5.10	27.1	25.7	1.4
Sm01	2.04	12.2	11.4	0.8
Sm02	1.84	8.9	8.0	0.9
Sm03	2.56	13.2	13.0	0.2

由表 1 的测试结果可以看出,改进后的编码器相对于原编码器有较大的提高,节省了编码时间。

另一个需要考虑的是此方法的改进对音频的音质是否造成影响。因此,实验还对原编码器和改进后的编码器生成的音频文件进行解码,通过 Adobe Audition 3.0 软件对两种编码器生成的文件波形、频谱、相位、相位谱等进行对比。以测试序列 Es01 为例,FAAC 编码器解码后音频文件的频谱图如图 3 所示。

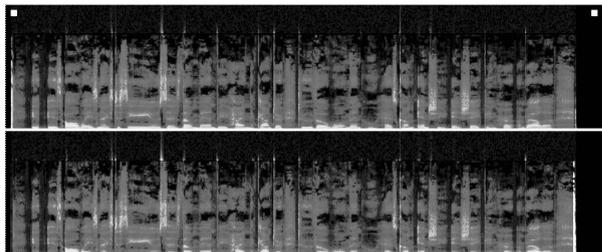


图 3 FAAC 解码的音频频谱图

Fig.3 The audio spectrum diagram of decoding FAAC

改进 FAAC 编码器解码后音频文件的频谱图如图 4 所示。

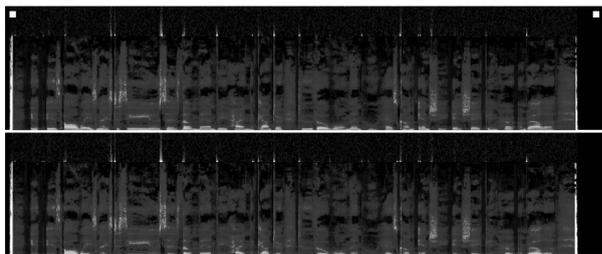


图 4 改进 FAAC 解码的音频频谱图

Fig.4 The improved audio spectrum diagram of decoding FAAC

由图 3 和图 4 可知,改进 FAAC 编码器与原 FAAC 编码器对音频文件编码效果基本相同,并没有影响音频文件的音质,达到了算法改进的目标。因此,该算法的改进是可行的,能有效地提高编码效率。

5 结 论

通过对 MPEG-4 AAC 音频编码器量化模块的研究,实现了一种减少量化时间的算法。通过实验证明了该算法的可行性,该算法是在不影响音频质量的前提下提出的,与文献[4]和文献[5]的思路和

方法有很大的不同,该算法对实时音频编码具有重要的研究意义,能够提高编码效率。该算法的提出只是针对外层循环的计算过程,而内层循环的计算没做任何改进,因此,如何有效地改进内层循环将是下一步研究的内容。

参考文献:

- [1] Painter T. Perceptual coding of digital audio[J]. Proceedings of the IEEE, 2000, 88(4): 451 - 513.
- [2] ISO/IEC IS - 14496 (Part 3, Audio), Information technology - coding of audio - visual objects[S].
- [3] Aggarwal A. Near - optimal selection of encoding parameters for audio coding[C]//Proceedings of 2001 International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, USA: IEEE, 2001: 3269 - 3272.
- [4] Olivier Derrien. A new quantization optimization algorithm for the MPEG advanced audio coder using a statistical subband model of the quantization noise[J]. IEEE Transactions on Audio, Speech, Language Processing, 2006, 14(4): 1328 - 1339.
- [5] Ali M T, Saleem M. Improved audio quality at 48 kbits/s for MPEG-4 AAC[C]//Proceedings of 2008 International Conference on Electrical Engineering, Hongkong: IEEE, 2008: 1 - 8.
- [6] Brandenburg K. OCF - A new coding algorithm for high quality sound signals[C]//Proceedings of 1987 IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, TX: IEEE, 1987: 141 - 144.

作者简介:

刘红甫(1983—),男,河南开封人,2009年获学士学位,现为硕士研究生,主要研究方向为音频编解码技术;

LIU Hong - fu was born in Kaifeng, Henan Province, in 1983. He received the M. S. degree in 2009. He is now a graduate student. His research concerns audio decoding technology.

Email: hb_ wh_ zc@163. com

张 聪(1968—),男,上海人,博士,教授、硕士生导师,主要研究方向为音频编解码技术与通信;

ZHANG Cong was born in Shanghai, in 1968. He is now a professor with the Ph. D. degree and also the instructor of graduate students. His research concerns audio decoding technology and communication.

梁瑞凡(1988—),男,湖北汉川人,硕士研究生,主要研究方向为音频编解码技术。

LIANG Rui - fan was born in Hanchuan, Hubei Province, in 1988. He is now a graduate student. His research concerns audio decoding technology.