

面向多声道三维音频的和差压缩编码技术

董石, 胡瑞敏, 杨玉红, 王晓晨, 涂卫平

(武汉大学 国家多媒体软件工程技术研究中心, 湖北 武汉 430079)

摘 要: 分析了传统基于单声道对的和差(M/S)编码技术, 针对三维音频系统的空间特性将其拓展为基于三声道的和差编码技术(3D-M/S)。并提出一种基于 3D-M/S 的音频编码框架, 使其能够压缩任意声道配置的三维音频系统。实验结果表明与独立声道编码相比, 所提方法能够在相同码率下提高约 25% 的客观质量。并且在同等质量下, 所提方法引起编解码复杂度的增加约为基于 PCA 编解码方法的 1/3。

关键词: 三维音频压缩; VBAP; 和差编码; 多声道音频编码

中图分类号: TN911.21

文献标识码: A

文章编号: 1000-436X(2014)06-0148-06

Three-channel dependent mid/side coding framework for multichannel 3D audio

DONG Shi, HU Rui-min, YANG Yu-hong, WANG Xiao-chen, TU Wei-ping

(National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan 430079, China)

Abstract: The conventional mid/side coding method is discussed and generalized to a three-channel dependent M/S coding (3D-M/S) method. Moreover, a codec framework based on 3D-M/S is proposed which enables the 3D-M/S to work for three-dimensional audio systems with more than three channels. It is experimentally shown that proposed framework gets about 25% objective quality improvement compared with independent coding, and only increases about 1/3 complexity of the PCA method.

Key words: 3D audio compression; VBAP; M/S coding; multichannel audio coding

1 引言

随着三维影视市场的巨大成功, 三维音频技术受到了广泛关注并得到了迅速发展。三维音频系统因其能为观众提供更好的声音定位和临场感, 越来越多被引入与音频相关的应用当中取代传统环绕声系统。波场合成(WFS, wave field synthesis)、Ambisonics 和幅度矢量合成(VBAP, vector based amplitude panning)是目前发展最为完备的 3 种多声道三维音频理论, 其中 WFS 基于惠更斯原理重建原始声场信号, 德国 Fraunhofer 研究院的 IDMT 实验室和法国 IRCAM 等研究机构都开展了长期研究, 并尝试将 WFS 应用于剧场和音乐会的实时直

播; Ambisonics 利用球谐函数记录声场并驱动扬声器。其有严格的扬声器排布要求, 能够在扬声器中心位置高质量重建原始声场; VBAP 基于三维空间中的正弦法则, 利用空间中 3 个临近的扬声器形成三维声音矢量。由于算法简单, VBAP 也是最常用的多声道三维音频技术^[1]。如日本 NHK 公司的 22.2 多声道三维音频系统正是利用 VBAP 技术进行三维声像的生成, 2012 年面向三维音频的国际标准 MPEG-H 正在制定当中, 22.2 多声道三维音频系统也被其作为标准系统。

未来三维音频技术将逐渐走向成熟并取代现有的立体声和环绕声技术。当前三维音频系统的一个主要特点就是其庞大的声道数目。例如 WFS 一

收稿日期: 2013-05-03; 修回日期: 2013-06-30

基金项目: 国家自然科学基金资助项目(61231015, 61102127, 61201340, 61201169); 湖北自然科学基金资助项目(2011CDB451, 2012FFB04205)

Foundation Items: The National Natural Science Foundation of China(61231015, 61102127, 61201340, 61201169); The Natural Science Foundation of Hubei Province (2011CDB451, 2012FFB04205)

般包含了数十个甚至上百个声道, 22.2 多声道系统也有分 3 层排布的 24 个声道, 而 Ambisonics 系统虽然可以灵活地制定声道数目, 但由于较少的声道数会严重影响三维音质, 因此声道数一般也有几十个。相比于双声道立体声和 5.1 环绕声技术, 三维音频技术中声道数的增加带来了数据量的激增。来自 Fraunhofer 的研究报告显示, WFS 用于实时传输时需要高达 37 Mbit/s 的码率; 而对于未压缩的 22.2 多声道系统也需要 28 Mbit/s 的码率^[2]。2011 年, 殷福亮等对三维音频技术进行了综述, 指出三维音频巨大计算量和数据量是其发展的瓶颈问题^[3]。目前的存储介质和传输带宽难以满足三维音频庞大数据量的需求, 因此三维音频信号的压缩技术将成为三维音频领域的重要研究方向。

针对三维音频数据量激增的问题, 近期在三维音频压缩方面已经展开了一些非常有价值的工作。2007 年, Goodwin 等提出了一种基于主成分分析的参数编码多声道压缩框架^[4], 这种框架可以应用于增强特殊的音频场景信号并提高空间音频编码的顽健性。2008 年, Cheng 提出了一种空间压缩环绕声编码(S³AC)方法, 用于 Ambisonics 信号的参数编码^[5]。2009 年, Hellerud 使用声道间预测编码的方法去除 Ambisonics 声道间的冗余信息^[6], 这种方法具有较低的算法延时, 但是计算复杂度较高。2010 年, Pinto 利用时空域变换将 WFS 信号分解成平面波和瞬态波, 通过舍弃瞬态波, 并利用感知编码对平面波信号进行压缩, 来获得编码增益。由于其时空变换的精度取决于空间分辨率, 也即 WFS 声道数目, 所以编码效率会随着单位空间内声道数的增加而提高^[7,8]。2011 年, Cheng 又进一步提出了一种空间定位量化格点(SLQP, spatial localization quantization point)的参数编码方法, 并使用三维方位线索来压缩 VBAP 信号^[9]。由于提取了空间线索并通过下混技术减少了声道数目, SLQP 取得了较高的压缩效率。

上述模型和参数编码方法可以提供较高的压缩比, 但实际应用中音频编码器要兼顾算法复杂度和编码效率 2 个方面, 而且参数编码也只能在低码率下提高编码音质。所以本文考虑高质量/高比特率的应用情况, 并专注于传统和差编码方法(M/S, mid/side coding)。本文提出一种基于三声道的 M/S 编码方法(3D-M/S, three-channel dependent mid/side coding)和相应框架, 用于压缩 VBAP 三维音频系统

和 22.2 多声道系统的音频信号。首先分析了传统 M/S 编码技术的基本原理, 并基于传统方法原理提出新的变换矩阵, 将传统 M/S 编码方法拓展到三维空间情况, 实现三维空间中基本单元——三声道的信号冗余去除。进一步, 提出一种基于三声道组的三维音频编解码框架, 使 3D-M/S 能够压缩任意声道配置。最后, 将所提算法和独立声道编码(independent channel coding)、PCA 编码(PCA, principal component analysis)在算法复杂度和客观质量上进行了对比, 实验表明 3D-M/S 方法保留了传统 M/S 编码低复杂度、高效率的特点, 适用于三维音频直播通信传输的需求。

2 三维空间中差编码

2.1 传统立体声 M/S 编码

由 J D Johnston 提出的 M/S 编码技术^[10]被多种主流音频压缩标准采用, 如 MPEG1-Audio Layer 3(MP3)和 aacPlus (AAC+)。M/S 编码基于“立体声 2 个声道是强相关信号”的现象, 不直接编码原始双声道信号, 而利用 M/S 变换将原始声道转换为和声道与差声道进行编码。由于差声道信号动态范围小于原始信号, 编码所需的比特数更少, 因而获得编码增益。

本文通过立体声正弦模型来说明 M/S 的编码过程, 首先用矢量 $V_0 = (C_L, C_R)$ 来表示立体声道信号

$$\begin{aligned} C_L &= S \sin \theta \\ C_R &= S \cos \theta \end{aligned} \quad (1)$$

其中, S 为虚拟声源信号, θ 为音源的方位角, 代表音源能量分布于在 2 个声道的增益因子, 且 $\theta \in [0, \pi/2]$ 。M/S 编码的 2 个变换矩阵 M_0 和 M_1 为

$$M_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad M_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \quad (2)$$

实际编码时 C_L 和 C_R 为子带的能量值。当 2 个声道信号差异较大时, 和差变换后差信号的能量依然较大, M/S 编码使用矩阵 M_0 编码原始信号; 当 2 个声道非常相似时才会使用 M/S 编码, 因为变换为和差信号后会重新计算掩蔽阈值, 增加编码复杂度。如 MP3 标准中 2 个声道信号满足能量判别条

件 (3), 即当能量差异小于阈值 $Thr=2$ dB 时, 才使用矩阵 M_1 编码进行和差变换^[10]。

$$\left| 10\text{lb} \left(\frac{C_L}{C_R} \right)^2 \right| \leq Thr \quad (3)$$

为了方便讨论矩阵切换, 这里将判别条件表示为 2 个向量的距离。将 $C_L/C_R = \tan \theta$ 代入式(3)可得到

$$\frac{1}{\sqrt{1+10^{\frac{Thr}{10}}}} \leq \cos \theta \leq \frac{1}{\sqrt{1+10^{-\frac{Thr}{10}}}} \quad (4)$$

由式(4)可见当且仅当 $\theta \rightarrow \pi/4$ 即 $\cos \theta \rightarrow 1/\sqrt{2}$ 时条件(3)成立, 此时使用 M_1 矩阵编码。因为 θ 可以表示向量 V_0 的角度, 所以式(4)也可以表示为信号矢量 V_0 与矩阵 M_1 和矢量 $V_1 = (\sqrt{2}/2, \sqrt{2}/2)$ 的内积

$$\langle V_0, V_1 \rangle \geq Thr_v \quad (5)$$

此处的 Thr_v 是能量阈值 Thr 对应的矢量距离阈值

$$Thr_v = \frac{\sqrt{2}}{2} \left(\frac{1}{\sqrt{1+10^{\frac{Thr}{10}}}} + \frac{1}{\sqrt{1+10^{-\frac{Thr}{10}}}} \right) \quad (6)$$

式(5)即为能量判别条件的等价矢量表示, 其表明只有输入信号矢量 V_0 与变换矩阵 M_1 的和矢量 V_1 足够接近时才使用和差变换矩阵。这个矢量表示对下节讨论三维空间 M/S 编码将会十分有用, 因为三维空间 M/S 编码存在多于一个的变换矩阵及其和矢量。

2.2 三维音频 M/S 编码

传统立体声和环绕声系统中为了保证声像的稳定性, 只用最邻近的 2 个声道形成一个有方向的声像, 所以 2 个邻近的声道具有最大相似性, 因而 M/S 编码和其他参数编码也是基于 2 个声道单元去除冗余。

对于 Ambisonics 和 VBAP 等三维音频系统(如图 1 所示), 其声道在三维空间中的球面上排布, 具有三维特性, 与立体声和环绕声有很大差异。首先, 在球面排列的三维音频系统中, 三角形是能覆盖三维空间中区域的最小组合, 因此最少要使用 3 个声道才能在三维空间某一区域内的任意位置形成声像。其次, VBAP 算法只用邻近 3 个声道(C_1, C_2, C_3) 形成一个虚拟声像, 这 3 个声道之间具有最大相关性。基于三维音频的这种特点, 本文提出一种基于三声道的和差编码算法 (3D-M/S) 编码球面排布三维音频信号。

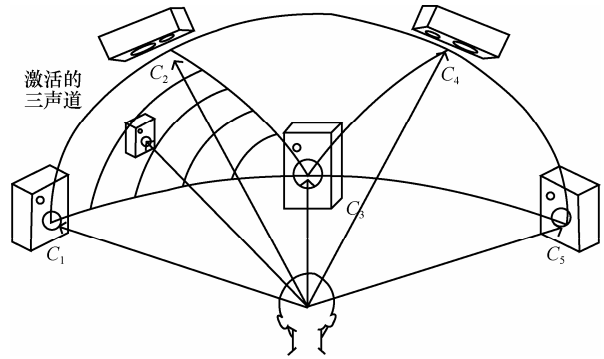


图 1 VBAP 系统在三维空间球面上的扬声器排布

此处按照传统 M/S 编码算法对三声道情况进行讨论, 基于三维空间中正弦模型, VBAP 三维音频系统信号可表示为矢量 $V_0 = (C_1, C_2, C_3)$, 其中

$$\begin{cases} C_1 = S \sin \theta \cos \varphi \\ C_2 = S \sin \theta \sin \varphi \\ C_3 = S \cos \theta \end{cases} \quad (7)$$

且 $\theta, \varphi \in [0, \pi/2]$, 它们确定了 3 个声道的增益因子和声像位置。声像位置可分为 3 种基本情况。第 1 种情况, 只通过一个声道产生声像, 对应于声像正好定位于在某一个声道所在位置, 即球面三角形的顶点, 此时和立体声中仅使用一个声道情况相同, 3 个声道信号完全不同, 因此使用 M_0 矩阵编码原始声道。第 2 种情况, 只通过 2 个声道产生声像, 此时对应于虚拟声源位于 2 个声道之间, 即球面三角形的边上。这种情况和传统的立体声相同, 因此使用 M_1 矩阵编码。但由于三声道中有 C_3^2 种两声道情况, M_1 变换矩阵需要相应的拓展成 3 种变换矩阵, 拓展方式如(8)中 M_1, M_2 和 M_3 所示遍历所有 C_3^2 种两声道情况。第 3 种情况, 3 个声道都用于产生声像。此时对应于虚拟声源位于 3 个声道之间, 即球面三角里面。为了实现这种情况下声道间冗余信息去除, 需要依据传统 M/S 编码原理设计新的变换矩阵 M_4 。其第一个矢量是和矢量, 其余矢量与和矢量正交且相互正交, 保证其余 2 个声道将变换为差声道。

$$M_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

$$\begin{aligned}
 \mathbf{M}_2 &= \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 2 & 1 & 2 \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 2 & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix} \\
 \mathbf{M}_3 &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 2 & 2 & 0 \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ 2 & -\frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
 \mathbf{M}_4 &= \begin{bmatrix} \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{3} \end{bmatrix} \quad (8)
 \end{aligned}$$

3D-M/S 编码的矩阵判别条件和传统 M/S 编码采用相同的方法设计：当 3 个声道中有且仅有 2 个声道满足传统 M/S 编码切换条件时，选择 \mathbf{M}_1 , \mathbf{M}_2 , \mathbf{M}_3 中和矢量与输入矢量距离最近的矩阵，得到一个能量较小的差声道；当 3 个声道中所有声道之间均满足传统 M/S 编码切换条件时，表明 3 个声道差异很小。此时采用变换矩阵 \mathbf{M}_4 ，得到 2 个能量较小的差声道。利用矢量距离的方式，上述 3D-M/S 编码切换条件可以表示为

$$mode = \begin{cases} \mathbf{M}_4, \langle \mathbf{V}_0, \mathbf{V}_4 \rangle \geq Thr_v \\ \mathbf{M}_i, \langle \mathbf{V}_{0i}, \mathbf{V}_i \rangle \geq Thr_v \\ \langle \mathbf{V}_{0i}, \mathbf{V}_i \rangle \geq \langle \mathbf{V}_{0j}, \mathbf{V}_j \rangle, \forall j \neq i \\ \mathbf{M}_0, \text{其他} \end{cases} \quad (9)$$

其中， $i, j \in \{1, 2, 3\}$ 且 $\mathbf{V}_{01} = (0, C_2, C_3)$, $\mathbf{V}_{02} = (C_1, 0, C_3)$,

$\mathbf{V}_{03} = (C_1, C_2, 0)$ 都是 \mathbf{V}_0 的子集。 $\mathbf{V}_1 = (0, \sqrt{2}/2, \sqrt{2}/2)$, $\mathbf{V}_2 = (\sqrt{2}/2, 0, \sqrt{2}/2)$, $\mathbf{V}_3 = (\sqrt{2}/2, \sqrt{2}/2, 0)$, $\mathbf{V}_4 = (\sqrt{3}/3, \sqrt{3}/3, \sqrt{3}/3)$ 分别是各变换矩阵的和矢量。

3 基于 3D-M/S 的多声道编码框架

3D-M/S 编码只适用于三声道情况，实际三维音频系统都包含更多声道数目。球面排布三维音频系统的多声道配置可归纳为图 2 所示的拓扑结构，每一个声道 C_i 对应一个扬声器。

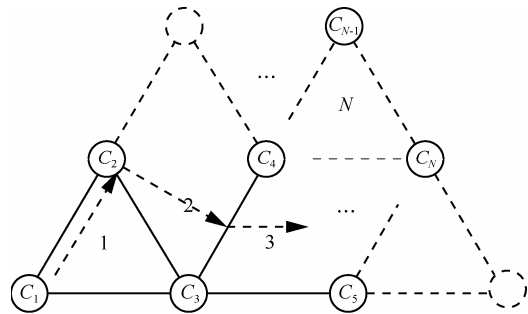


图 2 三维音频系统中多声道排布的一般情况

为了让 3D-M/S 编码能够去除多声道信号冗余，本文提出一种三维音频编码框架如图 3 所示，其中 C_M 为和声道， C_S 和 C_T 分别是第 2 个和第 3 个差声道。该框架对拓扑结构中声道按三角形逐个处理，直到所有的声道都被编码完为止。每个 3D-M/S 编码单元和前一个单元共用 2 个声道，即每个单元只加入一个新的声道。除第一个编码单元外，其他单元只将 3D-M/S 变换后第 3 个差声道 C_T 送入核心编码器，保证了编码声道数和原始输入声道数完全一致。由于 3D-M/S 变换矩阵均为正交矩阵，在解码端 3D-M/S 解码单元通过解码出的声道 C_T 和前一单元的 2 个声道，可以还原出原始声道信号。当变换矩阵为 \mathbf{M}_0 时， C_T 为原始声道，当变换矩阵为 \mathbf{M}_1 、

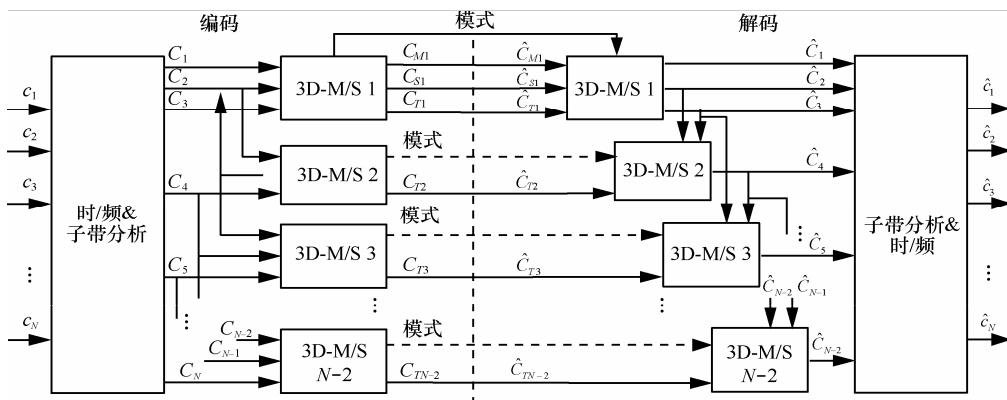


图 3 针对多声道的 3D-M/S 编解码框架

M_2 、 M_3 、 M_4 时, C_T 为差声道, 与编码原始声道相比, 整个框架通过编码差声道获得编码增益。

4 实验与性能分析

4.1 实验配置

PCA 是理论上最好的去相关变换算法, 因此本文选取了 3D-M/S、PCA 和独立声道编码 3 种编码方法, 分别在比特率、复杂度、客观音质上进行对比。复杂度通过各编码器在 PC(CPU: Intel Core2 Duo P8600 2.53 GHz, RAM:8 GB)上的运行时间来衡量; 客观音质则使用分帧信噪比和 ITU 标准推荐的客观质量 ODG 评分来衡量^[11]。实验采用图 2 所示的五声道配置(C_1, C_2, C_3, C_4, C_5)。测试所用的三维音频信号为合成的 VBAP 信号, 该信号使用 3 个 MPEG 测试序列(es01 语音, sc03 交响乐, si02 响板, 48 kHz 采样)作为虚拟音源, 虚拟声源的定位依据 VBAP 算法实现。由于实验中无法涵盖虚拟声源的所有可能位置, 因此实验采用空间中 3 种代表性的声源位置: 声源在三角形中从一个顶点移动到另一个顶点, 从一个顶点移动到对边, 以及从一条边移动到另一条边, 如图 2 箭头方向所示。

本文算法基于开源 AAC 编码器 FAAC-1.28 实现, 解码器基于对应的 FAAD2-2.7 实现。与传统 M/S 编码相同, 编码器在频域上按个子带进行 3D-M/S 和 PCA 编码。为了便于码率分析和算法实现, 核心编码器关闭了长短窗切换, 只采用固定的长窗。这会同等程度降低各种对比算法的信噪比, 但结果不失一般性; 为了避免 FAAC 动态带宽设置功能带来的影响, 实验中采用固定编码带宽 12 kHz 和 35 个均匀子带。

独立声道编码: 声道送入核心编码器单独编码。

3D-M/S 编码器: 通过原 AAC 编码器已有的子带能量进行矩阵判别并进行矩阵变换, 一个子带的变换矩阵索引使用 3 bit 来量化。

PCA 编码器: 计算每个编码单元信号的特征向

量, 并通过特征向量将三声道子带转化为去相关后的主要成分和次要成分, 然后送入核心编码器。遵循已有 PCA 多声道编码算法^[12], 每个子带用 24 bit 量化协方差矩阵, 最后传递到解码端重建特征向量并将主次成分还原为原始声道。

4.2 客观质量测试和复杂度测试

图 4 显示了 3 种编码方法对声像移动的声道编码后的分帧信噪比值。声源 1 位置移动的时间范围从 1~400 帧, 声源 2 位置移动的时间范围从 400~800 帧, 声源 3 位置移动的时间范围从 800 帧到最后一帧。首先, 图中所有的信噪比曲线都呈下降趋势, 这是因为瞬态信号和交响乐具有更加丰富的信息频谱细节, 使得它们更难被压缩。其次, 当虚拟声源位置接近 2 个声道间时(200 帧 C_1, C_2 之间, 400~800 帧 C_4, C_5 之间), 3D-M/S 比声道独立编码的方法具有更高的信噪比。特别是在接近 600 帧时, 3 个声道几乎完全相同, 此时通过变换矩阵 M_4 可以最大程度地去除冗余, 2 个差声道能量几乎为零。可以保留更多比特数编码和声道, 因此此时 3D-M/S 的效果优于 PCA 方法。

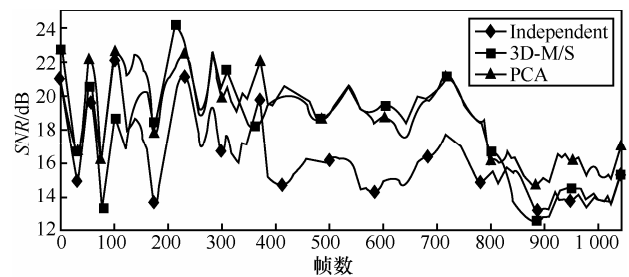


图 4 3 种方法在虚拟音源位置移动下的分帧信噪比

表 1 显示每种方法所用的比特数和复杂度。3 种方法采用大致相同的码率进行编码以进行对比, ODG 分数由整个信号计算所得。数据表明 PCA 和 3D-M/S 都获得了约 0.66 ODG 分的改善, 但是 PCA 算法涉及协方差矩阵计算, 复杂度提高了 29.9%。而 3D-M/S 只进行矩阵变换, 所以复杂度仅提高了

表 1 3 种方法的 ODG 质量、比特率、时间复杂度

方法	ODG 质量	比特率/信道/(kbit·s ⁻¹)									复杂度/s		
		C_1	C_2	C_3	C_4	C_5	P_1	P_2	P_3	平均	编码	解码	比特率
Ind	-2.558	61.6	60.8	60.3	58.7	58.7	—	—	—	61.2	2.382	0.223	100.0%
3D-M/S	-1.909		229.7		21.9	57.6	4.9	4.9	4.9	64.8	2.597	0.302	111.3%
PCA	-1.871		133.4		25.6	42.0	39.3	39.3	39.3	63.8	2.973	0.411	129.9%

注: 3D-M/S 和 PCA 编码中 C_1, C_2, C_3 为首个编码单元总比特率, C_4 和 C_5 为 C_T 声道。 P_1, P_2, P_3 是编码单元 1、2、3 的参数比特率。

11.3%, 是 PCA 方法的 37.8%。并且 PCA 的参数码率为每声道 39.3 kbit/s 远高于 3D-M/S 每声道的 4.9 kbit/s。这种高参数码率在 3 个声道不相关(比如声道记录不同内容, 声道间信号冗余非常小)情况下, 核心编码器无法获得编码增益, 高参数码率反而导致编码效率下降。但 3D-M/S 参数码率仅为每声道 4.9 kbit/s, 在其主要应用场景的高码率下, 其编码效率不会明显降低。

5 结束语

本文针对多声道三维音频系统提出了一种拓展和差编码技术, 将传统基于声道对的 M/S 编码方法拓展为基于三声道的 3D-M/S 编码, 并继承了传统 M/S 编码低复杂度的特点。对于 VBAP 和 Ambisonics 等球面排列的三维音频系统, 空间中三声道是形成声像和冗余存在的最小单元, 因此 3D-M/S 更符合三维音频的空间结构。结合所提编码框架, 3D-M/S 可以进一步对 3 个声道以上的三维音频系统进行压缩编码。实验表明, 与 PCA 编码和独立编码相比, 3D-M/S 在保持较低复杂度的同时提供了相当的压缩效率。考虑到三维音频技术的快速发展及其直播通信中终端处理能力, 在实际应用中低复杂度的音频编码技术会更有应用前景。

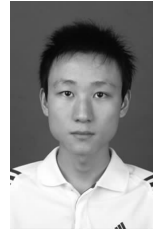
参考文献:

- [1] COOPERSTOCK J R. Multimodal telepresence systems[J]. IEEE Signal Processing Magazine, 2011,28(1):77-86.
- [2] SAKAIDA S, IGUCHI K, NAKAJIMA N, *et al.* The super hi-vision codec[A]. IEEE International Conference on Image Processing[C]. San Antonio, TX, 2007.21-24.
- [3] YIN F, WANG L, CHEN Z. Review on 3D audio technology[J]. Journal on Communications, 2011,32(2):130-138.
- [4] GOODWIN M M, JOT J. Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement [A]. IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Honolulu, HI,2007.9-12.
- [5] CHENG B, RITZ C, BURNETT I. A spatial squeezing approach to ambisonic audio compression[A]. IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Las Vegas, NV, 2008. 369-372.
- [6] HELLERUD E, SOLVANG A, SVENSSON U P. Spatial redundancy in higher order ambisonics and its use for low delay lossless compression[A]. IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Taipei, China, 2009.269-272.
- [7] PINTO F, VETTERLI M. Wave field coding in the spacetime frequency domain[A]. IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Las Vegas, NV, 2008.365-368.
- [8] PINTO F., VETTERLI M. Space-time-frequency processing of acoustic wave fields: theory, algorithms, and applications[J]. Signal Proc-

essing, IEEE Transactions on, 2010, 58(9): 4608-4620.

- [9] CHENG B, Spatial Squeezing Techniques for Low Bit-Rate Multichannel Audio Coding[D]. University of Wollongong, 2011.
- [10] JOHNSTON J, FERREIRA A J. Sum-difference stereo transform coding[A]. IEEE International Conference on Acoustics, Speech and Signal Processing[C]. San Francisco, CA,1992.569-572.
- [11] ITU-R. BS Recommendation 1387-1.Method for Objective Measurements of Perceived Audio Quality[S]. 1998.
- [12] DAI Y, HONGMEI A, KYRIAKAKIS C, *et al.* High-fidelity multichannel audio coding with Karhunen-Loevetransform[J]. IEEE Transactions on Speech and Audio Processing, 2003,11(4):365-380.

作者简介:



董石(1986-), 男, 湖北武汉人, 武汉大学博士生, 主要研究方向为信号处理和音频信号压缩编码。



胡瑞敏[通信作者](1964-), 男, 湖北武汉人, 武汉大学教授、博士生导师, 主要研究方向为多媒体信息处理、信源压缩、图像处理 and 安防监控技术。E-mail:hmr1964@163.com。



杨玉红(1974-), 女, 福建福安人, 武汉大学副教授、硕士生导师, 主要研究方向为语音信号处理和质量评价。



王晓晨(1981-), 男, 湖北武汉人, 武汉大学讲师, 主要研究方向为音频信号压缩和多媒体信号处理。



涂卫平(1974-), 女, 湖北荆州人, 武汉大学副教授, 主要研究方向为音频信号处理与压缩编码。