文章编号:1004-9037(2014)05-0661-16

三维音频技术综述

胡瑞敏 王晓晨 张茂胜 李登实 王 松
高 丽 杨 乘 杨玉红

(1.武汉大学计算机学院,武汉,430072; 2.国家多媒体软件工程技术研究中心,武汉,430072)

摘要:三维(Three-dimension, 3D)多媒体技术,尤其是和3D视频相比有所差距的3D音频技术受到了广泛的关注。当前三维音频技术研究可分为基于物理声场重建的多声道音频技术和基于感知的声音场景重建的多声道音频技术两大类。物理声场重建技术的重要代表是基于球谐分解的声重放技术和波场合成技术(Wave field synthesis, WFS),基于感知的声音场景重建技术主要包括幅度平移技术(Amplitude panning, AP)和基于头相关传输函数的双耳重建技术(Head related transfer function, HRTF)。本文对上述4类三维音频技术及其对应的典型系统进行了介绍及对比分析,并对三维音频技术当前3大主要研究热点:空间听觉机制、三维音频压缩编码以及三维音频系统精简的现状与前沿技术进行了介绍。

关键词:三维音频;波场合成;头相关传输函数;幅度平移;空间听觉;三维音频系统精简

中图分类号:TN911.7 **文献标志码**:A

Review on Three-Dimension Audio Technology

Hu Ruimin, *Wang Xiaochen*, *Zhang Maosheng*, *Li Dengshi*, *Wang Song*, *Gao Li*, *Yang Cheng*, *Yang Yuhong*

(1. Computer School, Wuhan University, Wuhan, 430072, China;
2. National Engineering Research Center for Multimedia Software, Wuhan, 430072, China)

Abstract: With the increasing requirement for audio-visual experience, three-dimension (3D) multimedia technology, especially 3D audio technology that is less developed than 3D video, has received widespread attentions in recent years. Current researches of 3D audio can be divided into two kinds of technologies based on physical reconstruction of acoustic field and perceptual reconstruction of sound scene respectively. The representatives of the former are ambisonics, which is a full-sphere surround sound reproduction technique based on spherical harmonic decomposition, and wave field synthesis (WFS), which is a spatial audio rendering technique based on wave fronts synthesis. Perceptual reconstruction techniques of sound scene mainly include the amplitude panning (AP) technique and the binaural sound synthesis technique based on head related transfer function (HRTF). The above four kinds of 3D audio technologies and the corresponding typical systems are introduced and compared in this paper. The current hot researches of 3D audio techniques are also presented mainly from three aspects: spatial hearing perceptual mechanism, 3D audio coding and multichannel simplification of 3D sound system.

Key words: 3D audio; wave field synthesis (WFS); head related transfer function (HRTF); amplitude panning; specitial hearing; 3D audio coding

基金项目:国家高技术研究发展计划("八六三"计划)(SQ2014GX02D01144)资助项目;国家自然科学基金重点(61231015)资助项目;国家自然科学青年基金(61201169,61102127,61201340)资助项目。
收稿日期:2014-08-14;**修订日期**:2014-09-19

引 言

2009 年,3D 电影《阿凡达》获得史上最高的 27 亿美元票房,开启了 3D 多媒体时代,3D 音视频系统也成为全球各大家电制造商竞争的新焦点。但由于 3D 音频技术发展与 3D 视频技术发展的不对等,导致目前无论是在影院还是在家庭,主流的 3D 多媒体系统都是采用“3D 视频+立体声/环绕声”方案。根据 2012 法国电信研究院在 MPEG 标准会议上给出的定义,3D 音频应该保证重建的声像拥有水平、垂直和距离共 3 个自由度,而立体声或环绕声系统所重建的声像仅具备水平方向上的自由度,无法让声像脱离扬声器所在的平面,还未达到“2D”规格,与 3D 音频定义相差甚远。可见当前 3D 多媒体系统在重建对象时存在视觉感受和听觉感受不一致的缺陷,导致沉浸感和真实感不足,难以达到身临其境的效果。因此,3D 音频技术已成为近年来多媒体领域的一个热门研究内容。

早在 1934 年,Steinberg 和 Snow 就提出了“声音幕帘”概念,并开展了针对三维音频的初步研究。“声音幕帘”概念指的是:如果能在一个面上用大量麦克风(传声器)组成一个紧密网格(即面麦克风网格阵列),就可以采集到原始声源的方位信息和声场形状。根据惠更斯原理,只要能用同样结构的面扬声器网格阵列,就可以重建出原始声源的方位和声场形状,这就是声场物理重建技术的雏形。从 20 世纪 80 年代末开始,麦克风阵列采集技术的实现与发展推动了物理声场重建技术的发展,其中最著名的是基于球谐分解的声重放技术和波场合成(Wave field synthesis, WFS)两大技术。Ambisonics 技术诞生于 1973 年[1],其利用球谐函数对原始声源进行分解,再利用均匀排布在等距离球面上的二次声源(扬声器)来发送分解后的信号来实现球面包围区域内原始声场精确重建。1975 年,英国牛津大学的 Gerzon 实现了一阶球谐分解,在一阶条件下重建区域会退化到中心原点(即最佳听音点)处[2]。但随着球谐分解的阶数值变大,重建区域半径将随之变大[3],相应的技术研究称为高阶 Ambisonics (Higher order ambisonics, HOA)[4]。波场合成技术是荷兰代尔夫特理工大学的 Berkhout 于 1988 年提出[5-6],在已知原始声源方位和信号的情况下,WFS 利用连续分布的二次声源,通过 Kirchhoff-helmholtz 积分可求出各二次声源的驱动信号,实现对原始声源的声场进行精确重建。Ambisonics 和波场合成技术的最大优点是可以在较大的听音区域内精确重建原始声场,但是这两种技术都需要为此布置数量巨大的扬声器,而且 Ambisonics 技术对扬声器排布也有严格要求,这些都限制了这两种技术的实际应用。

20 世纪 90 年代末,研究者发现若考虑听觉系统的感知特性,则在声场重建时不必充分追求重建声场和原始声场的一致也可使人们获得较好的沉浸感和定位感,并在此基础上开展了基于感知的虚拟声场景重建技术研究。相对于物理声场重建技术,基于感知的虚拟声场景重建技术对于回放环境和设备的要求大幅降低,更容易得到实际应用,因此也得到了学术界和工业界的广泛关注,其中应用最为广泛的就是幅度平移技术(Amplitude panning, AP)和基于头相关传输函数(Head related transfer function, HRTF)的双耳重建技术。1961 年诞生的幅度平移是一种虚拟声像控制技术,通过将分配至扬声器的信号幅度进行调整,控制人耳感知到的声像的位置。利用幅度平移技术,可通过对立体声设备左右两个声道信号的幅度调制,重建听音点前方扇形区域内的虚拟声像,具有一定的声像位置的表达能力和环境渲染能力。基于 HRTF 的双耳重建技术则是利用实验测量得到的 HRTF 模拟某一空间声像的信号传播到双耳的过程,可以利用只有两个声道的耳机在回放时实现虚拟声像的感知重建,为移动环境下双耳 3D 音频提供了技术支持。

1 主流三维音频基础理论及应用

1.1 Ambisonics

1883 年 Kirchhoff 提出 Kirchhoff-Helmholtz 积分声场物理重建技术的数理基础,Kirchhoff-Helmholtz 积分表明:在一个自由源封闭区域 V 内任意一点的声压可以通过区域 V 表面 S 上的声压和声压的梯度计算得到[6],如图 1 所示。如果将 Kirchhoff-Helmholtz 积分在球极坐标系上进行,则可得到任一声源的球谐函数展开形式[7-8]

$$p(r,\theta,\phi,k)=\sum_{n=0}^{N}\sum_{m=-n}^{n}A_n^m(k)j_n(kr)Y_n^m(\theta,\phi) \tag{1}$$

式中:r,θ,ϕ 分别表示球面半径、高度角和方位角,k 表示波数且 $k=\dfrac{2\pi f}{c}$,f 表示声音信号的频率,c 表示声音在介质中传播的速度,N 表示球谐函数展开的阶数,$j_n(\cdot)$ 表示第一类球贝塞尔函数,

$Y_n^m(\theta,\phi)$表示球谐函数，它的定义为

$$Y_n^m(\theta,\phi)=\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\theta)e^{im\phi} \tag{2}$$

式中：P_n^m 为勒让德函数，$A_n^m(\cdot)$为球谐函数的系数。当 $N=1$ 时式(1)表示一阶球谐函数展开；当 $N\geqslant 2$ 时式(1)表示声场的高阶球谐函数展开；当 $N\to\infty$时式(1)表示声场的无穷阶阶球谐函数展开。Ambisonics 技术通过 N 阶球谐函数分解原始声源声场，利用均匀分布在球面上的二次声源(扬声器)精确重建球面内部声场。

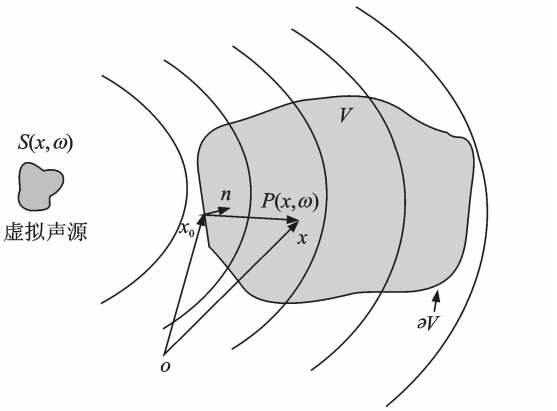

图 1　Kirchhoff-Helmholtz 积分图解

Fig. 1　Schematic illustration of Kirchhoff-Helmholtzintegral

1973 年 Gerzon 首次提出 Ambisonics 系统概念并给出了一阶 Ambisonics 系统模型。该系统在编码端利用原始声场的零阶和一阶球谐函数分解到 4 个声道 W,X,Y,Z，其中 W 声道的信号利用全向拾音器采集，携带全方向声信息，X,Y,Z 声道则分别利用指向性扬声器采集分别来自 x,y,z 三个轴向方向的声信号，携带着声场的方向信息，系统在重建端则利用 4 个声道的线性组合来得到各个扬声器的驱动信号。一阶 Ambisonics 系统对于低频信号重建效果较好，但对于高频信号重建效果不佳[6,9]，2001 年至 2009 年，澳大利亚国立大学的 T. D. Abhayapala 团队将 Kirchhoff-Helmholtz 积分在柱极坐标系上进行，提出声场的二维柱谐函数[2,10-12]，但仍存在重建区域半径小、高频信号重建效果不佳的问题。

为了提升 Ambisonics 的重建效果，法国电信的 J. Daniel 从 2003 年开始研究 HOA 理论。根据 Ambisonics 的原理，声场的分解需要进行球谐函数的无限阶展开，随着重建阶数的增加，Ambisonics 技术重建声场的区域逐渐增大，重建声音信号的频率逐渐增大。2005 年至 2012 年间，德国电信的 J. Ahrens 和 S. Spors 以及新西兰工业研究有限公司的 M. A. Poletti 等人继续针对 HOA 技术开展研究并搭建了实验系统[8,13-15]，使得 HOA 得到了极大的发展。其中 Poletti 基于 9 阶球谐函数，将 100 个扬声器摆放在半径为 2 m 的球面上。对于 500 Hz 的单频声源信号，该系统在水平听音平面可以精确恢复半径为 1 m 的圆形听音区域。对于 4 000 Hz 的单频声源信号，该系统在水平听音平面可以精确恢复半径为 0.1 m 的圆形听音区域。2014 年澳大利亚国立大学的 Wen Zhang 等针对 Ambisonics 技术要求在重建声场时扬声器必须在球面均匀排布的限制，研究了基于多环结构扬声器阵列的物理声场重建技术。多环结构扬声器阵列由多个平行的圆环形摆放的扬声器阵列构成，在圆环上的扬声器必须均匀分布，但平行的圆环间不用均匀分布。在此基础上，她构建了一个基于七阶 Ambisonics 的验证系统，该系统对于 3.1 m 处(60°,15°)，(135°,5°)方向传来的单个 500 Hz 单频声源信号，在半径为 0.5 m 的重建区域内重建声场与原始声场之间的平均相对平均平方误差分别为 0.001 和 0.01[16]。

在系统应用方面，一阶 Ambisonics 技术已有成熟的应用，Core Sound 公司的产品 TetraMic 使用四声道正四面体结构摆放的麦克风记录一阶 Ambisonics 声场，录音获得的 4 个声道可以以一阶 Ambisonics 系统回放也兼容单声道、立体声和 5.1 等多声道系统。Daniel Courville Ambisonic Studio 提出一种 Quad2B 软件，该软件可以将 4 个输入声道(左前，右前，左后，右后)转换成一阶 Ambisonics 四个输出声道。高阶 Ambisonics 系统也有一些相关产品，如 Blue Ripple Sound 有限公司最新使用三阶 Ambisonics 开发了 VST2 系列软件，可以对声场进行三阶 Ambisonics 编解码，平移，空间化，可视化，混响等相关操作。但高阶 Ambisonics 系统在精确恢复声场时所需的扬声器数目依然过分庞大，以在半径 1 m 区域内恢复最高频率 20 kHz 声场为例，需要对球谐函数进行不少于 37 阶的截断，而阶数 n 与声道数 L 之间的关系为 $L=(n+1)^2$，可见需要 1 369 个扬声器才能实现。扬声器数目的减少会严重影响重建声场的区域大小和信号频率上限，因此目前高阶 Ambisonics 系统主要还处于实验阶段。

1.2 波场合成

波场合成的物理原理可以回朔到 Huygens 原理。1678 年惠更斯在《光论》(《Traité de la Lumiere》)一书中提出了惠更斯原理,其内容是波从一个给定的波前传播,可被认为是由原始声源传播出来的或者由分布在波前的二次声源传播出来的[6]。1883 年,Kirchhoff 和 Rayleigh 通过对二次声源的分布增加一个自由度,将惠更斯原理推广到了一般情况,并给出了数学表达[6]

$$p(r,w)=\oint_S [G(\boldsymbol{X}\mid\boldsymbol{X}_0,\omega)\frac{\partial}{\partial \boldsymbol{n}}p(\boldsymbol{X}_0,\omega)- P(\boldsymbol{X}_0,\omega)\frac{\partial}{\partial n}G(\boldsymbol{X}\mid\boldsymbol{X}_0,\omega)]\mathrm{d}S \tag{3}$$

式中:$p(\boldsymbol{X},\omega)$表示区域 S 内部的声压,$P(\boldsymbol{X}_0,\omega)$表示区域 S 边界处的声压,$G(\cdot)$表示格林函数,k 表示波数,S 表示封闭区域的表面,$\boldsymbol{n}$ 表示指向区域 S 内部的法向量,$\partial/\partial\boldsymbol{n}$ 表示向量 $\boldsymbol{n}$ 方向的方向梯度。1993 年荷兰代尔夫特理工大学的 Berkhout 提出将 Kirchhoff-Helmholtz 积分中的封闭区域 V 简化成一个平面,使用有限等间隔排布的环状扬声器阵列可以实现在水平面重建合成的声场[6]。1995 年,MARINUS M. BOONE 研究团队提出了基于线阵列的波场合成技术,使用线扬声器阵列可以重建二维声场,并在 1999 和 2004 年继续将其扩展到和面扬声器阵列重建二维声场的情况[17-19]。

Kirchhoff-Helmholtz 积分从理论上给出了精确重建声场的方法,但并未给出重建频率和扬声器数量和排布的关系,2004 年,W. de Bruin 在其博士论文中指出利用数量有限、离散排布的扬声器可重建一定频率范围内声信号产生的声场,为了避免不可控的重建失真,扬声器的摆放需要满足空间奈奎斯特采样定律,即任意两扬声器间的距离必须小于最高频率声音信号的波长的一半[20]。2006 年开始,德国电信的 S. Spors 等又在此基础上进一步分析了数量有限,离散分布的线状和环状二次声源的声场重建误差模型[21-22]。然而,当时的波场合成技术研究并未利用麦克风采集原始声场。直到 2012 年,M. Cobos 在 S. Spors 研究成果基础上,提出利用紧凑摆放的麦克风阵列采集原始声场,通过对麦克风阵列信号进行时频分析,并利用波场合成技术重建了保持原有声场空间感的虚拟声音场景[23]。

在实际应用方面,波场合成系统一般使用线性扬声器阵列和面扬声器阵列,但是波场合成技术可应用于实际声场重建的系统较少。使用线性扬声器阵列的代表系统为 2001 年德国柏林工业大学电子演播室的 24 声道波场合成系统,该系统由线性有源扬声器阵列(FOSTEX 6302B)构成,扬声器之间两两间隔为 12.5 cm[24]。使用面扬声器阵列的代表系统为 2007 年德国柏林工业大学为 H0104 演讲大厅安装的一套包含 832 声道的波场合成系统,该系统为当时包含扬声器数目最多的波场合成系统,但是仍然不能满足空间奈奎斯特采样定律。该波场合成系统包含由 15 台装备支持 56 声道的 RME HDSP MADI 声卡的电脑构成的集群。和前面提到的 Ambisonics 技术类似,作为一个基于 Kirchhoff-Helmholtz 积分的物理声场重建技术,波场合成技术在较大的听音区域重建人耳可闻频率范围的声音信号需要数量巨大的扬声器(根据空间采样定律,针对最高频率为 20 kHz 的声信号,利用波场合成技术重建声场时扬声器间的间隔不能大于 0.85 cm,即使是一个半径 1 m 的环状阵列来恢复其内的平面声场,也需要至少 735 个扬声器),导致波场合成系统的应用更多地出现在三维声场重建实验和一些演示环境下,实际应用进展较为有限[25-27]。

1.3 幅度平移

幅度平移是一种高效的声像控制技术,通过将分配至扬声器的信号的幅度进行调整,控制人耳感知到的声像的位置。在立体声系统中利用声源与中心点正前方的夹角调制两个扬声器的信号的分配。

1961 年,哥伦比亚广播公司提出正弦法则,假设扬声器与中心点正前方夹角均为 φ_0,声源与正前方夹角为 φ(如图 2 所示),给左、右扬声器分配的权值分别为 g_1,g_2,则 g_1,g_2 可通过下式求解。但正弦法则要求头部必须保持静止,因此当头部存在略微偏转时无法使用。

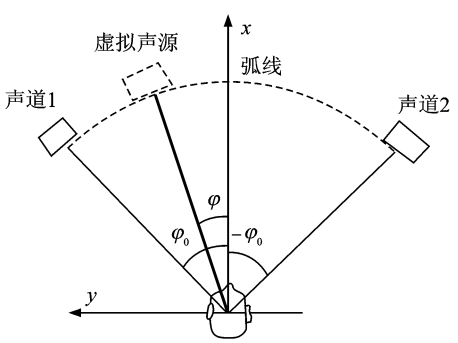

图 2 正弦法则配置图

Fig. 2 Configuration based on stereophonic law of sines

$$\frac{\sin\varphi}{\sin\varphi_0}=\frac{g_1-g_2}{g_1+g_2} \tag{4}$$

1973 年，罗马尼亚录音公司的 Bernfeld 提出了正切法则[28]，克服了正弦法则头部不能转动问题。Bernfeld 指出，两个扬声器增益之差与增益之和的比，应等于声像与正前方的夹角的正切值与扬声器与正前方的夹角的正切的比，公式如下

$$\frac{\tan\varphi}{\tan\varphi_0}=\frac{g_1-g_2}{g_1+g_2} \tag{5}$$

正切法则考虑了头部偏转的情形，对幅度平移技术造成了深远的影响。当扬声器与听音点在同一平面上时，正弦法则、正切法则能还原水平方向的声音事件。针对立体声系统，利用正弦法则或正切法则计算声像与听音位置正前方的角度，实现声像方向的调节；在环绕声系统中，选择声音事件方向最邻近两个扬声器作为幅度平移对象，计算增益因子调整声像方向。但是，当扬声器所在平面与用户头部不在同一平面上时正弦法则与正切法则均无法使用[29]。

针对三维空间中声源方向的还原，芬兰赫尔辛基理工大学的 VILLE PULKKI 于 1997 年提出了基于矢量的幅度平移（Vector-based amplitude panning，VBAP）技术，VBAP 利用 2 个或 3 个扬声器的位置的单位向量合成出虚拟声源的单位向量，达到重建声源方向的目的。假设声源和 3 个扬声器位于同一个球面上，将 3 个扬声器的单位向量视为基向量，虚拟声源的方向由它们的线性组合得到，如图 3 所示。设 3 个扬声器的单位矢量分别为 $\boldsymbol{l}_1,\boldsymbol{l}_2,\boldsymbol{l}_3$，每个扬声器的信号的增益为 g_1,g_2,g_3，声源单位矢量为 $\boldsymbol{l}_0$，则 $\boldsymbol{l}_0$ 可表示为

$$\boldsymbol{l}_0=g_1\boldsymbol{l}_1+g_2\boldsymbol{l}_2+g_3\boldsymbol{l}_3 \tag{6}$$

于是可求解增益得

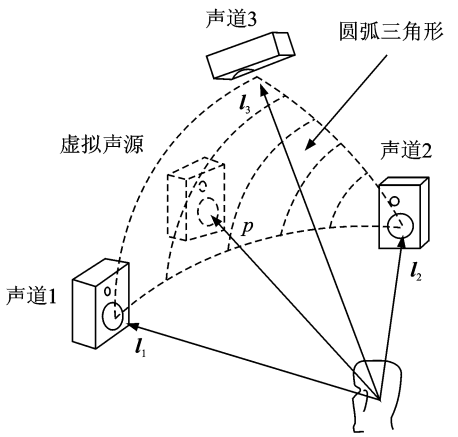

图 3　VBAP 原理图

Fig. 3　Vector base formulation of triplet-wise panning

$$g=[g_1\ g_2\ g_3]\begin{bmatrix}l_{11}&l_{12}&l_{13}\\l_{21}&l_{22}&l_{23}\\l_{31}&l_{32}&l_{33}\end{bmatrix}^{-1} \tag{7}$$

式中：$\boldsymbol{l}_i=[l_{i1},l_{i2},l_{i3}]$，$i=1,2,3$。

当扬声器在同一球面上时，VBAP 模型简单、计算高效，对 500～600 Hz 以下的声音的方向恢复较为准确。结合声音能量衰减规律，或者声音时间延迟的控制，可以将 VBAP 技术进一步拓展到非球面的情形。除了基于矢量的幅度平移技术之外，2009 年 Lossius 等提出基于距离的幅度平移技术（Distance-based amplitude panning，DBAP），在扬声器与听音点的距离和权值之间建立模型，利用扬声器与听音点的距离控制扬声器信号幅度。DBAP 技术的优势在于听音点可以位于扬声器阵列中的任何位置。

对虚拟声像的感知重建技术及系统在商业应用方面取得了巨大成功。双声道立体声系统可通过调整两个扬声器的信号的增益控制声像位置，为听音者形成声像在听音点前方扇形区域内运动的听觉感知。多声道环绕声系统则将声像运动范围由扇形区域扩展到圆形区域，典型的多声道环绕声系统有 Dolby（杜比）环绕声系统、DTS 环绕声系统等。其中杜比环绕声系统因其出色的声音渲染效果在影剧院、家庭环境中获得了广泛的使用。杜比实验室开发的 5.1 多声道环绕声系统已成为家庭影院系统的标配；杜比 7.1 环绕声系统（Dolby Surround 7.1）通过增加左后环绕和右后环绕两个声道，使得声音的包围感更加强烈。在 2009 年的国际消费电子展上，杜比展示了新的杜比定向逻辑技术，通过在前方增加两个高置声道将 7.1 声道升级为 9.1 声道，该系统能够提供一定的声像高度感知效果。

多声道环绕声系统中扬声器摆放在同一个水平面上，无法产生位于这个水平面之外的空中的声像。为了使听众对声像的感知从水平面拓展到真正的三维空间，音频技术的研究发展至三维音频层面，通过多层扬声器设置使听众获得三维音频享受。2005 年，日本 NHK 实验室正式提出 22.2 多声道原型系统，并被列入日本面向下一代超高清电视的三维音频标准。该系统中的扬声器阵列分为三层结构以实现三维沉浸效果，在三层扬声器共同作用下，在听音者周围重建三维声像，克服了环绕声在高度感知方面的不足。系统排布如图 4 所示，其中中间层设置了 10 个扬声器，上层设置了 9 个，下层屏幕前方设置 5 个（2 个为低频扬声器）。

22.2多声道系统能够在听音者周围重建出 3D 声像，在前方的屏幕区域内实现稳定的声音定位。2011 年 NHK 进一步将 22.2 声道下混为 10 声道或 8 声道，依然可以达到近似原 22.2 声道的重建效果。ISO/IEC 的 MPEG 标准工作组以 22.2 多声道系统为参考制定三维音频系统标准，并在 2011 年的需求提案中指出感知无失真压缩编码是该系统的重要需求之一。

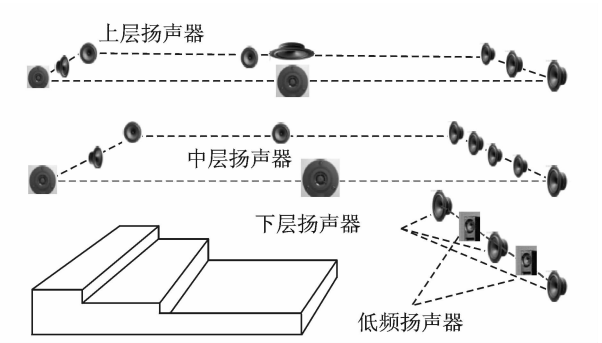

图 4 22.2 声道系统布置示意图

Fig. 4 Configuration of 22.2-channel sound reproduction system

2012 年，杜比公司最新提出了杜比全息声(Dolby atmos)多声道系统，通过将 64 个扬声器分层设计，综合利用多层扬声器渲染环境音和动态声效，使得影片的声音层次感更加明显，让观众更好地沉浸在影视体验中。

1.4 头相关传输函数

头相关传输函数即 HRTF 描述了自由场声波从声源到双耳的传输过程，反映了头部、耳廓和躯干等构成生理系统对声波综合滤波的结果，HRTF 是声源的水平角、高度角、距离、频率的连续函数，并且与生理结构和尺寸密切相关[30]。在自由场的情况下，HRTF 定义为

$$H_L = H_L(\theta,\phi,r,f) = p_L(\theta,\phi,r,f)/p_0(r,f) \quad (8)$$

$$H_R = H_R(\theta,\phi,r,f) = p_R(\theta,\phi,r,f)/p_0(r,f) \quad (9)$$

式中：p_L，p_R 分别为简谐点声源在左、右耳产生的复数声压；p_0 为头不存在时头中心位置处的复数声压；r 为声源到头中心的距离；f 为声波频率；水平角 $0°\leqslant\phi<360°$ 和高度角 $-90°\leqslant\theta\leqslant90°$ 表示声源的方向，其中 $\theta=0°$ 和 $90°$ 分别表示水平面和正上方，而($\phi=0°$，$\theta=0°$)和($\phi=90°$，$\theta=0°$)分别表示水平面上正前和正右方向。HRTF 的时域表示是头相关脉冲响应 $h_L(\theta,\phi,r,t)$，$h_R(\theta,\phi,r,t)$，简记为 HRIR。

利用 HRTF 技术生成三维音频信号的过程可描述为：原始单路音频信号 S 经过 HRTF 滤波器后，生成了带空间方位信息的左右耳音频信号 Y_L 和 Y_R，其中 $Y_L = H_L S$，$Y_R = H_R S$。然后可以采用耳机回放左右耳的信号，即可重建三维虚拟声像，当然也可以采用扬声器来重建三维虚拟声像，由于音频信号从扬声器播放后传递到人耳，会产生串声干扰，因此需要额外的串声消除过程，以保证到扬声器到达左右耳的信号仍然为 Y_L 和 Y_R[31]。

HRTF 技术中 HRTF 库的发展和完善以及定位准确性的提升是决定其重建的声音场景质量的关键所在。早期的 HRTF 库空间精度较低，1994 年，MIT 的 William G. Gardner 等对 KEMAR 人工头的 HRTF 进行测量(MIT HRTF 库)，测量距离为 1.4 m，得到了包含高度角 $-40°$～$90°$，水平角 $0°$～$360°$，共 710 个空间位置的 HRTF 数据库[32]。随后人们进一步考虑了人体测量参数对 HRTF 的影响，2000 年，加州大学的 V. R. Algazi 等对 KEMAR 人工头和 43 个真人的 HRTF 进行了测量(CIPIC HRTF 库)，测量距离为 1 m，测量的空间点数为 1 250，同时提供了人体测量参数[33]。2006 年，华南理工大学对中国人头模型的 HRTF 进行了测量。对 52 名受试者(男女各半)进行测量，建立了中国人样本的高空间分辨率的 HRTF 和受试者生理尺寸数据库[30]。2009 年，北大和中科院合作，首次在不同距离进行了测量，测量距离由 20 cm 变化到 160 cm，测量空间点数为 6 344(PKU&IOA HRTF 库)。该库对远场和近场都进行了测量，揭示了近场距离因素对 HRTF 的影响[34]。在定位准确性提升方面，针对通用 HRTF 产生的前后混淆及仰角误差问题，1998 年，南洋科技大学 Chong-Jin Tan 利用 HRTF 中某些频段对声音定位起关键作用这一现象，基于入射声方向对这些频段信号进行调整，提升定位效果[35]。2000 年，密歇根大学的 John C. Middlebrooks 等提出利用主观听觉比较从多套非个性 HRTF 中为听者选择一套最适合的 HRTF，并将其作为近似的个性 HRTF 来合成 3D 立体声，对定位精确度具有一定的改善效果[36]。2011 年，澳大利亚国立大学的 M. Zhang 等提出基于主成分分析获得具有强烈直接物理影响的重要的人体结构参数，为高效个性化 HRTF 生成提供了指导[37]。

1.5 主要技术对比分析

上述 4 类主流 3D 音频技术各有特色也各有

不足。表 1 给出了上述 4 类技术在内的主流音频重现系统及技术特性对比分析。在理想化的情况下，波场合成技术可以确保二次声源合成的声场与原始声场完全一致，因此听音者像在真实听音空间中一样感知和定位声源，并允许听音者在听音区域内部任意走动，声像不因人的位置的改变而发生改变。但波场合成技术要求大量的扬声器，对场地和器材有较高要求，系统费用昂贵[38]。

Ambisonics 技术的主要优势是编码和解码过程相互独立，编码时不必知道扬声器如何摆放，解码时可根据扬声器排布计算扬声器信号。另外 Ambisonics 编码方式是三维声场的真实重建，允许对三维声场的空间特征进行直接处理，例如，声场的旋转、反射。随着阶数的增加，球谐函数所携带的方向信息越来越精确，使得定位越来越精确，同时数据量会迅速增加。此外 Ambisonics 假设听音者的位置固定，因此有效听音区域有限。

幅度平移技术可以使用较少的扬声器实现声源的感知定位，基于幅度平移技术的多声道系统安装简单、快捷，实现高效，目前已有标准支持，有完善的系统、产品和成熟的解决方案，可以部署在影院、家庭等各种环境。但是幅度平移技术无法对原始声场进行精确的恢复，而且最佳听音区域十分狭窄。

HRTF 技术通过 HRTF 库虚拟生成某一空间声像到双耳的过程，实现了虚拟声像的感知重建。其优点在于只需要耳机即可实现 3D 音频的感知重建，但目前还存在通用 HRTF 库与人体个性化参数不能较好匹配的问题和头中效应较严重的问题，限制了 HRTF 技术的发展和推广。

2　三维音频研究前沿

2.1　空间听觉机制

2.1.1　基础与现状

在三维声场中，人类听觉系统对声源空间位置的感知包括 3 个方面：声源的水平方位，高度和距离。在水平面上，人类听觉系统对声源方位的定位主要依赖于 2 种双耳线索：双耳时间差（Interaural time difference，ITD）和双耳强度差（Interaural level difference，ILD）线索[39]：ITD 是由于声音到达左右耳所经过的距离不同而造成的，而 ILD 则是由于人头部对音频信号的遮蔽作用导致的；ILD 对 1 500 Hz 以上的声源定位起主要作用，ITD 则对 1 500 Hz 以下的低频声源的定位起主要作用。人耳通过对 ILD 和 ITD 线索的感知，就可以实现对声源方位的判断，但其感知灵敏度仍然具有一定的局限性。人耳能感觉到声像方位变化的阈值称为恰可感知差异（Just noticeable difference，JND），JND 越小则表明敏感度越强。影响双耳线索 JND 的因素是多方面的，主要包括声源频率、声源类型和声源方位等[40-41]。针对声源方位对 JND 影响的测量和分析结果表明，随着声源从中垂面方向向左右两侧移动，人耳对双耳线索的变化敏感度会明显降低[42-44]。针对双耳线索 JND 与信号频率之间关系研究的实验结果验证了人耳对 1 000 Hz 信号 ILD 线索的变化较不敏感，JND 值较大，在低频处略低，高频处偏高[45-49]。

表 1　主流音频重现系统/技术对比

Table 1　Comparison about sound reproduction systems/techniques

	环绕声	波场合成	Ambisonics	幅度平移	头相关传输函数
基本原理	幅度平移	Kirchhoff -Helmholtz 积分	球谐函数分解	幅度平移	基于听音实验和声传播
声道数量	较多	很多	很多	多	2
重建区域	小	大	较大	小	双耳
声场还原精度	低	高	高	中	高
空间感	较好	好	好	好	好
定位感	一般	强	强	强	强
缺点	声像范围有限，非3D音频	扬声器数目多，对场地和器材要求高，数据量巨大	低阶时有效听音区域和声场还原效果有限，扬声器排布要求严格	扬声器数量较多，数据量较大	存在由个性化导致的定位混淆和头中效应较严重的问题
适用范围	家庭、影院等	家庭、影院、演讲大厅	家庭、影院	家庭、影剧院等	个人桌面娱乐系统、移动环境应用

垂直方向上，主要通过在不同高度下的最小可听角度(Minimum audible angle, MAA)来测试人耳对垂直高度角的感知敏感度。高度角不同时，高度感知灵敏度有明显差异，当声源位于正前方或正后方时，随着声源高度的增加，声像的高度感知分辨率逐渐下降。当声源在人两侧时，感知声像的高度角随着声源高度的增加呈线性增加[50-53]。

听觉距离线索包括强度线索、混响线索、双耳线索、频谱线索、动态线索等。在自然听音环境中，强度、混响和双耳线索是人耳进行听觉距离定位的主要线索[54]。一般而言，当声源与听音点距离增加则在听音点处的声音强度下降。理想情况下，在自由场(无反射)下对于固定功率的点声源，强度与距离的关系符合平方反比定律，即：声源到听音点的距离翻倍则强度下降 6 dB。但声音强度只能够作为一个相对距离线索。混响线索是一种绝对距离定位线索。在存在声音反射面的混响环境中，直接到达听者的能量(直接声能量)，与通过一个或多个反射表面后到达听者的能量(混响能量)的比值，称为直混比(Direct reverberation ratio, DRR)，直混比能提供绝对距离信息[55]。除了强度和混响以外，双耳线索对于靠近头部的声源的距离感知也起着重要作用，尤其是在自由场近场情况下，双耳线索比强度线索更显著[56]。

国内对声源方位辨别的研究，中国科学院生理所的梁之安研究了声源定位与声源位置辨别阈的关系，用心理物理方法测定了人耳听觉的强度感知阈值和相位感知阈值[57]；2001 年第四军医大学的吴峰提出了一种测试人耳对声音空间定位能力的方法，通过测试 MAA 来判别受试者对声音的空间定位能力[58]；2005 年浙江大学的张彤在 MAA 的基础上开展听觉似动(Auditory apparent motion, AAM)，对没有空间位移的声音产生的运动错觉进行研究[59]2008 年武汉大学在国内首次对不同方位的双耳线索进行全频带的感知特性试验，获得较为精细的实验数据[60]。

2.1.2 前沿技术

现有的空间听觉研究结果证明声源的空间方位和频率对于双耳线索的临界可感知差异值 JND 具有重要影响。但现有实验得到的 JND 数据一般都是若干离散点的数据，缺乏全方位全频带的双耳线索临界感知特征数据。文献[61]同时在方位和频率两个维度上进行采样，针对不同方位和频率的声源进行测试，如图 5 所示，将双耳线索的 JND 数据由点或曲线拓展到曲面，从而能较全面地反映双耳线索 JND 的特性。该项研究为全面探究人耳感知特性提供基础数据和数学模型，也可为空间音频参数的感知编码提供支撑。

心理声学的研究表明，双耳听音既不是左耳和右耳分别听音的简单叠加，也不是左耳和右耳单独听音的平均，而是引入了新的信息，即空间信息。然而，现有感知熵理论建立在单耳听音模型基础上，无法度量双耳对空间信息的感知量。文献[62]从双耳听音出发，分析空间信息的物理层和生理层表示方法，建立双耳线索的生理感知模型，提出基于双耳线索生理感知模型的空间感知熵。

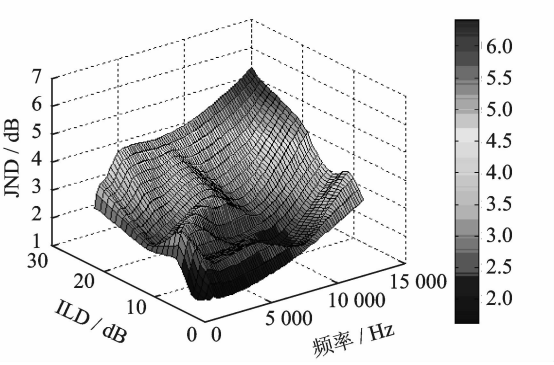

图 5 JND 随方位和频率变化的趋势

Fig. 5 Changing trends of JND

对多声道信号中的双耳线索来说，根据单符号离散信源定义，每一个双耳线索都可以看作某一集合中的事件，每个可能出现的数字(事件)都是信源中的元素，它们的出现往往具有一定的概率。以 ILD 为例，在其信源空间中，讨论信源 ILD 的信息量。

定义 1 信源空间$[L \cdot P]$

$$[L \cdot P]:\begin{Bmatrix} L: l_1, l_2, \cdots, l_i, \cdots, l_N \\ P(L): P(l_1), P(l_2), \cdots, P(l_i), \cdots, P(l_N) \end{Bmatrix}$$

定义 2 l 的不确定度为 $I(l)$，则 $I(l)=f[P(l)]$。

其中，$I(l)$为 l 所含的信息量，为其概率的函数。函数 $f[P(l)]$是 $P(l)$的单调递减函数，且具有可加性。通常情况下，取 $I(l)=-\log_2 p(l)$，$l\in L$，那么 L 的平均信息量

$$\mathrm{spe}(l)=-\sum_{i=1}^{N} p(l_i)\log_2 p(l_i) \tag{10}$$

spe(l)称为空间感知熵(Spatial perceptual entropy, SPE)。若将其看做是 L 的平均码字长，一旦确定了 L 上长度为 N 的信息，就可用 $N\cdot$ spe(l)来估算该段信息所需要的总比特数。相对

于两个声道感知熵相加的情况，对标准测试序列利用 SPE 计算的空间感知熵值平均降低约 45%，理论下限码率明显降低。

该方法针对感知熵理论无法解释空间感知信息量大小的问题，通过双耳听音机理分析，利用表示物理层信息量的双耳线索，建立双耳线索的生理感知模型，并借鉴感知熵理论，给出了空间感知熵 SPE 的计算方法，丰富和发展了空间音频编码技术，也可为多声道音频的高效压缩提供理论支撑。

2.2　多声道三维音频压缩编码

2.2.1　基础与现状

对于单声道音频信号压缩编码，为了去除声道内的冗余信息，MPEG 1-audio layer 3 (MP3)技术将人耳感知机理引入编码中，提出基于心理声学模型的感知音频编码技术，利用听觉掩蔽效应和子带编码技术，将各子带量化噪声保持在听觉阈值之下，从而以较低码率(128 kbps)获得近似透明的音质。

立体声和 5.1 声道相对于单声道音频能更好地重建水平面声音方位和场景信息，但如果直接采用传统单声道编码技术对立体声或 5.1 环绕声系统的信号进行编码会导致编码码率随声道数目增加成线性增长，给存储和传输带来巨大挑战。为此，在立体声编码领域，1992 年 Johnston 等人提出的和差矩阵编码技术(Mid/side coding, M/S)[63]被 Dolby 的 AC3 以及 MPEG 组织的 AAC 等国际标准广泛采用。1994 年德国 Fraunhofer 研究所首次利用人的感知特性进行参数编码，提出强度立体声(Intensity stereo, IS)编码技术[64]。针对 IS 在低频段编码质量降低的问题，2001 年美国新泽西州多媒体信号处理实验室的 C. Faller 和 F. Baumgarte 分析了传统的人耳空间听觉原理，提出利用人耳空间方位感知的两个重要参数——ILD 和 ITD 对双声道音频信号进行参数编码，给出了完整的双耳线索编码(Binaural cue coding, BCC)框架，仅使用 2 kbps 左右的参数码率就能重建立体声的空间声效[65]。2004 年，J. Breebaart 等人提出在 BCC 基础上简化空间线索提取的参数立体声技术(Parametric stereo, PS)[66]，成为空间音频编码领域第一个实用性的成果——AAC Plus v2，被认为是当时最高效优质的立体声编码方案。2005 年诺基亚研究所和爱立信研究中心提出了一种基于时域预测参数的编码技术，以 2.0～8.0 kbps的极低码率实现了立体声编码，该技术被国际语音音频压缩标准 AMR-WB＋采用[67]。2006 年法国电信提出了一种基于主成分分析(Principal component analysis, PCA)的双声道编码技术[68]。

多声道音频编码方面，1999 年 Dolby 公司提出基于预测矩阵的 Meridian lossless packing (MLP)技术，与独立声道未编码数据相比，MLP 压缩比可达 2:1～4:1，被广泛的用于 DVD 音频文件压缩。MLP 技术旨在对多声道数据进行无损编码，压缩效率较低[69]。2009 年 Hellerud 利用高阶 Ambisonics 临近声道信号有较强相关性的特点，用一个声道信号预测其临近声道，获得较大的编码增益[70]。2011 年英国萨里大学提出基于分析合成的子带参数提取技术[66,71]，采用闭环分析替代了传统开环分析技术，从而减小空间参数提取后的信号合成误差。2013 年，澳大利亚 Wollongong 大学的 Ritz 提出空间方位量化格点技术(Spatial localization quantization point, SLQP)，利用扬声器位置信息进行三维分解，估计声源的方位信息并进行参数化，把 3D 多声道信号下混为不含位置信息的较少声道，去除空间信息冗余[72-73]。

由于 BCC 在低码率下出色的空间音效重建表现，2005 年 ISO/MPEG 组织基于空间音频 BCC 编码技术提出 MPEG Surround 多声道压缩编码标准，以声道对为基础，采用树形结构提取各声道对的空间参数和声道间预测参数，逐对去除多声道的空间冗余信息[74]。针对 MPEG Surround 对所有声道对的空间参数采用同一量化表的问题，多名研究者根据人类感知分辨率提出非均匀的量化方法[75-76]。2013 年 MPEG Surround 技术被采纳成为 MPEG 组织 3D 音频多声道编码标准的一部分。

2.2.2　前沿技术

(1)3D-MS 矩阵编码

信号分组决定了三维音频多声道音频编码器的整体框架和压缩效率。现有三维音频多声道信号分组编码技术由于缺乏理论指导，停留在以声道对为基础的分组方法阶段，限制了分组后去冗余算法的效率，造成同等编码质量下三维音频编码器整体码率的上升。面向声道对的 M/S 编码无法完全去除三维音频信号组间的冗余，而以全部声道为基础的编码方式由于缺乏对信号相关性的局部性考虑造成信号混叠，影响三维音频的压缩效率和重建质量。针对以上问题，文献[3,77]借鉴传统 M/S

矩阵编码技术，将其拓展到高阶矩阵，提出基于 3 声道的 M/S 编码方法（Three-channel dependent mid/side coding，3D-M/S）和相应框架，用于压缩基于 VBAP 的三维音频系统和 22.2 多声道系统的音频信号。

传统 M/S 编码基于“立体声两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，因此量化后信号的熵更小，使得之后的 Huffman 编码所需的比特数更少，从而获得编码增益。M/S 编码的两个变换矩阵 $\boldsymbol{M}_0$ 和 $\boldsymbol{M}_1$ 为

$$\boldsymbol{M}_0=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\quad \boldsymbol{M}_1=\begin{bmatrix}\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2}\end{bmatrix} \tag{11}$$

三维声场中实际的声像位置可能出现在球面上 3 个扬声器覆盖的整个区域中任意位置，主要可分为 3 种基本情况：

(a)只通过一个声道产生声像，声像位于该扬声器所在位置，因此只需编码原始声道，得到第 1 个变换矩阵 $\boldsymbol{M}_0$。

(b)只通过两个声道产生声像。此时对应于虚拟声源位于两个声道之间，即球面三角形的边上。这种情况和传统的立体声相同，因此可以使用和差矩阵编码，只是三声道从两声道拓展成 3 种变换矩阵，$\boldsymbol{M}_1$，$\boldsymbol{M}_2$ 和 $\boldsymbol{M}_3$。

(c)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，即球面三角形上。为了实现这种情况下声道间冗余信息去除，需要借鉴传统 M/S 编码原理，保证三声道信号变换后得到两个差声道信号，因此设计新的变换矩阵 $\boldsymbol{M}_4$。其第 1 个矢量是和矢量，其余矢量相互正交且与和矢量正交。这几个矩阵共同组成基于 3 声道的和差变换矩阵

$$\boldsymbol{M}_0=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\quad \boldsymbol{M}_1=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\\ 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2}\end{bmatrix}$$

$$\boldsymbol{M}_2=\begin{bmatrix}\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2}\\ 0 & 1 & 0\\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2}\end{bmatrix}\quad \boldsymbol{M}_3=\begin{bmatrix}\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0\\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$\boldsymbol{M}_4=\begin{bmatrix}\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3}\\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0\\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{3}\end{bmatrix} \tag{12}$$

该方法根据三维空间相关性在高维度信号上拓展了传统 M/S 矩阵编码理论，在相同质量下可提高压缩效率 14%，且所提去冗余算法复杂度只为传统算法的 30%。

(2)SLQP 技术

针对三维音频空间音源丰富这一特点，澳大利亚 Wollongong 大学的 Ritz 等人提出 SLQP 技术，利用扬声器的方位信息，将各声道的信号强度在听音者的绝对坐标系中进行分解，并估计出虚拟声源方位信息。然后将所有声道下混，对提取出的声源位置信息进行格点量化传输到解码端。如图 6 所示，将每个声道信号视作一个矢量（如仅考虑声道

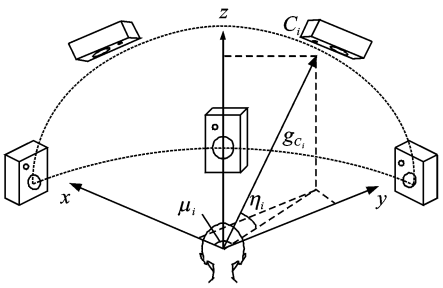

图 6 扬声器信号矢量与听音者绝对坐标

Fig. 6 Coordinates of loudspeakers and listening point

间强度差异），那么对于中心点的听音者而言，所有扬声器作用的结果可以视作一个和矢量，这个和矢量是所有声道信号矢量的矢量和。给定扬声器 C_i 在绝对坐标系中产生的信号矢量 $\boldsymbol{P}_{C_i}(k,n)$ 为

$$\boldsymbol{P}_{C_i}(k,n)=g_{C_i}(k,n)\cdot\begin{bmatrix}\cos\varphi_i\cdot\cos\theta_i\\ \sin\varphi_i\cdot\cos\theta_i\\ \sin\theta_i\end{bmatrix} \tag{13}$$

式中：$\boldsymbol{P}_{C_i}(k,n)$ 为扬声器 C_i 的频域信号，i 为扬声器的编号，k 为频域索引，n 为帧索引，θ_i 和 φ_i 分别代表扬声器 C_i 的高度角和方位角。若使用 N 个扬声器，则和矢量对应的虚拟音源的水平角、高度角计算

$$\tan\varphi_{\mathrm{sum}}(k,n)=\frac{\sum_{i=1}^{N}g_{C_i}(k,n)\cos\varphi_i\cos\theta_i}{\sum_{i=1}^{N}g_{C_i}(k,n)\sin\varphi_i\cos\theta_i} \tag{14}$$

$$\tan\eta_{\text{sum}}(k,n)=\frac{\sqrt{\left[\sum_{i=1}^{N}g_{C_i}(k,n)\cos\varphi_i\cos\theta_i\right]^2+\left[\sum_{i=1}^{N}g_{C_i}(k,n)\sin\varphi_i\cos\theta_i\right]^2}}{\sum_{i=1}^{N}g_{C_i}(k,n)\sin\theta_i} \tag{15}$$

SLQP 通过提取虚拟声像的空间位置信息来保证音源方位精确还原，同时去除空间信息冗余，能以 538 kbps 码率实现 16 声道三维音频信号的压缩[78]。

2.3　三维音频系统精简

2.3.1　基础与现状

如前所述，三维声场重建技术的本质是利用空间位置摆放的多个扬声器（即二次声源），通过在指定区域内进行声场叠加，重构出原始声场。然而，目前能实现三维声场重建的音频系统所需通道数通常要几十个，甚至要达到上百个。如何利用少量扬声器重建出原有大量扬声器构建的声场，显然成为三维音频系统普遍推广的关键问题，也是三维声场扬声器精简的目标。假设以原点 o 为球中心点建立球面坐标系，扬声器 y_l 位置坐标为 $y_l=(\sigma_l,\varphi_l,\theta_l)$，其中 σ_l 为扬声器到原点 o 的距离，φ_l 为水平角，θ_l 为高度角。接收点 x 坐标为 $x=(\rho,\theta,\phi)$，如图 7 所示。若扬声器 y_l 的声音信号为 $s_l(t)$，$l=1,\cdots,N$，且以点声源形式向外传播时，则自由声场中 N 个扬声器在接收点 x 处的声压通过傅里叶变化后如图 7 所示。

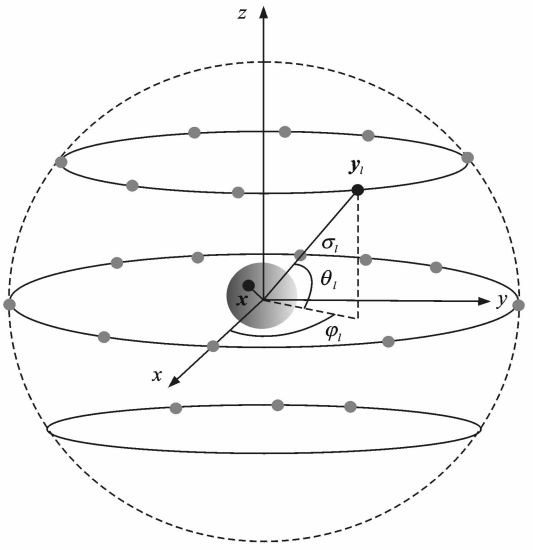

图 7　三维音频系统精简基本方法

Fig. 7　Schematic illustration of downmixing method in 3D sound system

$$p(x;\omega;k)=\sum_{l=1}^{N}A\frac{\mathrm{e}^{-\mathrm{i}k|y_l-x|}}{4\pi|y_l-x|}s_l(\omega) \tag{16}$$

式中：$k=2\pi f/c$ 是单位距离传播的声波数（称为波数），声波的频率为 f，传播速度为 c；$|y_l-x|$ 是扬声器 y_l 到接收点 x 的距离。假设单位距离上扬声器传播出去的声压与扬声器的输入信号成正比，且比例因子为 A；i 为虚数单位。

扬声器精简的目标就是用更少的扬声器（如 M 个扬声器，$M<N$）重建 N 个扬声器在接收点 x 处的声压。若此时扬声器 y'_j 的声音信号为 $q_j(t)$，$j=1,\cdots,M$，则 M 个扬声器在接收点 x 处重建声压表示为

$$\tilde{p}(x;\omega;k)=\sum_{j=1}^{M}A\frac{\mathrm{e}^{-\mathrm{i}k|y'_j-x|}}{4\pi|y'_j-x|}q_j(\omega) \tag{17}$$

且要求 $p(x;\omega;k)=\tilde{p}(x;\omega;k)$ 若要精确重建一个区域内的声场（$x\in\chi$），式(3)需要考虑区域 χ 内的所有接收点，即

$$\oint p(x;\omega;k)\mathrm{d}x=\oint\tilde{p}(x;\omega;k)\mathrm{d}x \tag{18}$$

如前所述，可以通过 WFS 和 HOA 技术实现对原始 3D 声场的重建，但是 WFS 所需扬声器的数量巨大，而 HOA 对扬声器的摆放要求苛刻。因此，2012 年 MPEG 发布了 3D 音频的需求（N12610），要求在保持用户听感基础上，通过精简扬声器数目来支持面向家庭电视直播等应用。

2.3.2　前沿技术

2011 年，日本 NHK 研究室 Ando 提出一种多通道音频系统精简方法[79]，并且在精简过程中能够保持听音点（即原点 O）处的声压不变。其基本思想是将 N 通道音频系统中每一个扬声器 y_l 都视为 M 通道音频系统中一个虚拟点声源，通过从 M 通道音频系统中选取能包围这个虚拟声源的 3 个扬声器 y'_1，y'_2 和 y'_3，用 3 个扬声器在听音点 O 产生出与虚拟点声源位置一致的虚拟声像。若 $q_j(t)=w_{lj}s_l(t)$，$j=1,2,3$，其中 w_{lj} 是第 j 个扬声器音频信号可以得到 y_l 扬声器音频信号的权值系数。则：

$$\frac{\mathrm{e}^{-\mathrm{i}k|y_l-o|}}{4\pi|y_l-o|}=\sum_{j=1}^{3}\frac{\mathrm{e}^{-\mathrm{i}k|y'_j-o|}}{4\pi|y'_j-o|}w_{lj} \tag{19}$$

式中：$\dfrac{\mathrm{e}^{-\mathrm{i}k|y_l-o|}}{4\pi|y_l-o|}$ 为原始声场第 l 个扬声器到听音点 o 处的声音传播函数 h_l（$l=1,\cdots,N$），$\dfrac{\mathrm{e}^{-\mathrm{i}k|y'_j-o|}}{4\pi|y'_j-o|}$ 为重建声场第 j 个扬声器到听音点 o 处的声音传播函数 $\tilde{h}_j$（$j=1,\cdots,N$）。

从 N 通道到 M 通道音频系统的精简过程其

实是关注如何设置权值转换矩阵 $\boldsymbol{W}=\begin{bmatrix} w_{11} & \cdots & w_{N1} \\ \vdots & & \vdots \\ w_{1M} & \cdots & w_{NM} \end{bmatrix}$，让原声场中 N 个扬声器到听音点 o 处的声音传播函数 $H_d=(h_1,\cdots,h_N)$ 与重建声场中 M 个扬声器到听音点 o 处的声音传播函数 $\widetilde{H}_d=(\tilde{h}_1,\cdots,\tilde{h}_N)$ 满足

$$\widetilde{H}_d\boldsymbol{W}=H_d \tag{20}$$

Ando 提出的这种多通道精简方法可以成功将 22.2 多通道音频系统声道数精简到 8 个通道。但是，一方面该方法未考虑双耳无法同时在听音点 o 处，另一方面该方法通过大量主观测试确定精简后音频系统（如 10 通道和 8 通道）的扬声器排布。2014 年，武汉大学提出了一种保持双耳声压低失真的 3D 多通道音频自动精简方法。图 8 描述了在精简过程中包括：(1)提取原始声场 N 通道音频系统的扬声器排布；(2)多通道迭代精简（图中虚线框部分）；(3)得到精简后重建声场不同通道下扬声器的排布。其中多通道迭代精简过程主要是选取符合精简条件的 4 个扬声器，并从其中剔除一个扬声器。通过不断重复此过程，以此实现扬声器数目精简。

假设人头中心位于坐标中心点 $o=(0,0,0)$ 处，且左右耳关于中心点对称，所以对右耳声压失真的讨论也适用于左耳声压失真。定义原始声场右耳 x_R 处声压与中心点处声压的差异 $\sigma(x_R-o,k,w)$ 和重建声场右耳 x_R 处声压与中心点处声压的差异 $\sigma'(x_R-o,k,w)$ 分别为

$$\sigma(x_R-o,k,w)\triangleq\frac{p(x_R,k,w)-p(o,k,w)}{p(o,k,w)}$$

$$\sigma'(x_R-o,k,w)\triangleq\frac{\widetilde{p}(x_R,k,w)-\widetilde{p}(o,k,w)}{p(o,k,w)} \tag{21}$$

通过计算声压失真相对误差，控制重建声场与原始声场在右耳处声压的影响，即

$$\varepsilon(x_R-o,k,w)=|\,\sigma(x_R-o,k,w)-\sigma'(x_R-o,k,w)\,|^2 \tag{22}$$

以 22.2 多通道音频系统为例，这种自动精简方法可以得到从 22 个通道到 8 个通道共 23 种多通道音频系统，且双耳处声压失真相对误差小于给定的阈值。表 2 为自动精简得到的 6 种 10 通道音频系统和 8 通道音频系统的所有方案，与 Ando 多通道精简方法中 10 通道和 8 通道音频系统进行 RAB 主观测试实验，评价声音定位和声音强度。其实验结果表明，在声音定位和声音强度两个方面自动精简方法与 Ando 方法的平均差异得分近似为零，而且自动方法的 8_2 和 8_3 排布方式的声音定位还稍优与 Ando 多通道精简方法下的 8 通道的声音定位。另外，该自动精简方法还可以减少通过主观经验选择扬声器排布所耗费的时间。

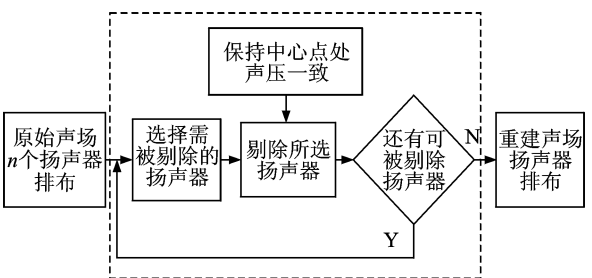

图 8 3D 多通道音频系统精简方法流程

Fig. 8 Downmixing procedurein 3D sound system

3 结束语

近年来，随着人们对视听体验要求的不断提升，3D 多媒体技术，尤其是和 3D 视频相比有所差距的 3D 音频技术受到了广泛的关注。当前三维音频技术研究可分为基于物理声场重建的多声道音频技术和基于感知的声音场景重建的多声道音频技术两大类。其中，物理声场重建技术基于 Huygens 原理和 Kirchhoff-Helmholtz 积分，通过

表 2 自动多通道音频精简方法所得 10 通道和 8 通道扬声器排布

Table 2 Configuration parameters using automatic downmixing method in 10-channel and 8-channel sound system

排布	重建声场扬声器排布（水平角，高度角）									
10-1	(0,0)	(30,0)	(90,0)	(150,0)	(180,0)	(270,0)	(90,45)	(270,45)	(0,90)	(90,−30)
10-2	(0,0)	(90,0)	(180,0)	(270,0)	(0,45)	(90,45)	(180,45)	(270,45)	(0,90)	(90,−30)
10-3	(0,0)	(30,0)	(150,0)	(180,0)	(270,0)	(45,45)	(90,45)	(135,45)	(270,45)	(90,−30)
8-1	(0,0)	(90,0)	(180,0)	(270,0)	(90,45)	(270,45)	(0,90)	(90,−30)	—	—
8-2	(0,0)	(90,0)	(180,0)	(270,0)	(90,45)	(225,45)	(315,45)	(90,−30)	—	—
8-3	(0,0)	(180,0)	(270,0)	(0,45)	(90,45)	(180,45)	(270,45)	(90,−30)	—	—

某种特定摆放的扬声器阵列实现对原始声源声场的精确重建，而感知声像重建技术则是利用人耳听觉特性实现对声音场景内声源的方位和距离信息的感知重建。

物理声场重建技术的重要代表技术是 Ambisonics 和波场合成技术，这两者均给出了可完整恢复某一区域内物理声场的扬声器阵列中个扬声器所需信号的求解公式，具有恢复效果精准、可恢复区域大的优点，尤其是 Ambisonics 技术，可以做到采集端和回放端各自独立，在回放时无需考虑是采用何种 Ambisonics 采集技术，同时还可向下兼容目前现有的立体声，5.1，7.1 等非 3D 音频回放系统。但在重建端，Ambisonics 技术受球谐函数截断方式限制，WFS 技术受二次声源排布必须满足空间采样定理要求限制，使得二者在重建声场时其所需的最小扬声器数目均随所重建声场的最高频率的提升而显著增加，如要重建的声场频率范围覆盖人耳可闻的频率范围，则扬声器均需上千个，同时 Ambisonics 技术对扬声器的摆放也有着苛刻要求，以上都极大的限制了物理声场重建技术的在适应娱乐方面实际应用。

基于感知的声音场景重建技术主要包括幅度平移技术和基于头相关传输函数的双耳重建技术。幅度平移技术通过调整各扬声器的信号幅度实现了原始声源方位信息的感知重建，其特点在于计算高效、系统实现相对于物理声场重建技术要容易很多，以此得到了广泛的应用，现有的立体声、多声道环绕声等系统也可看作是幅度平移技术的一种实现，对于基于幅度平移的 3D 音频系统，MPEG 标准组织也是在 22.2 声道系统上开展相关标准化工作。但幅度平移技术也存在重建听音区域狭小，重建效果区别大等问题，往往需要通过 3D 声场主观听音质量评测技术来保障重建效果，不过这也推动了人耳空间听觉机制的研究和 3D 声场主观评测技术的发展。另一方面，虽然基于幅度平移的 3D 音频系统所需声道数大大小于物理声场重建技术，但依然达到十数个或数十个，比当前常用的环绕声系统有了成倍地增加，因此针对此类系统的多声道 3D 音频压缩编码技术和声道数精简技术也成为最近的研究热点。不同于上述 3 类面向多音箱回放环境的 3D 音频技术，HRTF 技术主要面向耳机回放环境，由于是所有主流 3D 音频技术中对回放设备和条件要求最低的，所以特别适合于移动环境和单人欣赏。HRTF 技术将原始声场中声源到双耳受人体散射及衍射作用的影响的传播过程用头相关传输函数进行模拟，其中头相关传输函数库通过实验得到，其效果收到实验样本生理形态学参数、样本数量以及空间采样点密度等多方面条件影响，虽可用耳机实现原始声源的感知重建，但由于人体形态学参数的个性化差异会导致声像空间位置混淆和头中效应等问题。

综上所述，当前 3D 音频技术研究存在着多个不同的技术流派，各自均有不同的优势和缺陷。物理声场重建技术更适合于实验室和专业演示应用，未来需要针对其回放设备和环境要求过高的问题，在保证声场重建准确性的同时，精简扬声器数目并有效地扩大重建声场的最佳听音区域方面开展研究。基于感知的声音场景重建技术主要面向视听娱乐应用，三维音频技术要能为听者提供更好的三维空间沉浸感和方位感等听觉体验，有必要对人耳三维空间感知机理进行深入透彻的研究，指导三维声场信息的空间采样以及对空间方位信息的感知特性分析；此外，为了解决三维音频多声道系统由于声道数激增产生的海量数据导致在现有传输和存储环节带来的瓶颈问题，基于三维声场空间感知机理，分析声道间空间信息冗余，在保持三维音频空间感的同时，研究三维音频高效压缩编码和系统精简技术将对三维音频发展和实际推广应用具有重要意义；同时，针对现有的音频质量客观评价方法仍然停留在二维感知声场的问题，为了对三维音频系统所能提供的空间听觉体验进行有效的评价，面向三维音频感知声场的客观评价模型也将是未来的研究热点。

参考文献：

[1] Gerzon M A. Ambisonics: Part two: Studio techniques[J]. Studio Sound, 1975,8(17):24-30.

[2] Ward D B, Abhayapala T D. Reproduction of a plane-wave sound field using an array of loudspeakers [J]. Speech and Audio Processing, IEEE Transactions on, 2001,9(6):697-707.

[3] Bertet S, Daniel J, Moreau S. 3D sound field recording with higher order ambisonics-objective measurements and validation of spherical microphone[C]// Audio Engineering Society Convention 120. Paris, France: Audio Engineering Society, 2006:1-24.

[4] Daniel J, Moreau S, Nicol R. Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging[C]//Audio Engineering Society Convention 114. Amsterdam, Nether Lands: Audio Engineering Society, 2003:1-18.

[5] Berkhout A J. A holographic approach to acoustic control[J]. Journal of the Audio Engineering Society, 1988,36(12):977-995.

[6] Berkhout A J, de Vries D, Vogel P. Acoustic control by wave field synthesis[J]. The Journal of the Acoustical Society of America, 1993, 93(5): 2764-2778.

[7] Poletti M A. Three-dimensional surround sound sys-

tems based on spherical harmonics[J]. Journal of the Audio Engineering Society, 2005, 53(11): 1004-1025.

[8] Poletti M A. Three-dimensional surround sound systems based on spherical harmonics[J]. Journal of the Audio Engineering Society, 2005, 53(11): 1004-1025.

[9] Gerzon M A. Periphony: With-height sound reproduction[J]. Journal of the Audio Engineering Society, 1973, 21(1): 2-10.

[10] Ward D B, Abhayapala T D. Reproduction of a plane-wave sound field using an array of loudspeakers [J]. Speech and Audio Processing, IEEE Transactions on, 2001, 9(6): 697-707.

[11] Betlehem T, Abhayapala T D. Theory and design of sound field reproduction in reverberant rooms[J]. The Journal of the Acoustical Society of America, 2005, 117(4): 2100-2111.

[12] Wu Y J, Abhayapala T D. Theory and design of soundfield reproduction using continuous loudspeaker concept[J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2009, 17(1): 107-116.

[13] Daniel J. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format[C]//23rd International Conference: Signal Processing in Audio Recording and Reproduction. Copenhagen, Denmark: Audio Engineering Society, 2003: 1-20.

[14] Ahrens J, Spors S. An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions[J]. Acta Acustica United with Acustica, 2008, 94(6): 988-999.

[15] Ahrens J, Spors S. Applying the ambisonics approach to planar and linear distributions of secondary sources and combinations thereof[J]. Acta Acustica United with Acustica, 2012, 98(1): 28-36.

[16] Zhang W, Abhayapala T. Three dimensional sound field reproduction using multiple circular loudspeaker arrays: functional analysis guided approach [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 7(22): 1184-1194.

[17] Boone M M, Verheijen E N, Van Tol P F. Spatial sound field reproduction by wave field synthesis[J]. Journal of the Audio Engineering Society, 1995, 43(12): 1003-1012.

[18] De Vries D, Boone M M. Wave field synthesis and analysis using array technology: Applications of signal processing to audio and acoustics[C]//1999 IEEE Workshop on. New Paltz, NY, USA: IEEE, 1999: 15-18.

[19] Boone M M. Multi-actuator panels (MAPs) as loudspeaker arrays for wave field synthesis[J]. Journal of the Audio Engineering Society, 2004, 52(7/8): 712-723.

[20] De Bruijn W. Application of wave field synthesis in videoconferencing[D]. Delft: Delft University of Technology, 2004.

[21] Spors S, Rabenstein R. Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis[C]//120th AES Convention. Paris, France: Citeseer, 2006: 1136-1139.

[22] Spors S, Rabenstein R, Ahrens J. The theory of wave field synthesis revisited[C]//124th AES Convention. Amsterdam: Audio Engineering Society, 2008: 1-19.

[23] Cobos M, Spors S, Ahrens J, et al. On the use of small microphone arrays for wave field synthesis auralization[C]//45th International Conference: Applications of Time-Frequency Processing in Audio. Helsinki, Finland: Audio Engineering Society, 2012: 324-328.

[24] Weske J. Virtural sound localization using wave field systhesis technology in 24-channel system [M]. Chemnitz: Technique University of Chemnitz, 2001.

[25] Epain N, Friot E. Active control of sound inside a sphere via control of the acoustic pressure at the boundary surface[J]. Journal of Sound and Vibration, 2007, 299(3): 587-604.

[26] Naoe M, Kimura T, Yamakata Y, et al. Performance evaluation of 3D sound field reproduction system using a few loudspeakers and wave field synthesis [C]//Universal Communication, 2008. ISUC′08. Second International Symposium on. Osaka, Japan: IEEE, 2008: 36-41.

[27] Fazi F, Nelson P, Christensen J E, et al. Surround system based on three-dimensional sound field reconstruction[C]//Audio Engineering Society Convention 125. San Francisco, CA, USA: Audio Engineering Society, 2008: 203-207.

[28] Bernfeld B. Attempts for better understanding of the directional stereophonic listening mechanism[C]// Audio Engineering Society Convention 44. San Diego, CA, USA: Audio Engineering Society, 1973: 454-458.

[29] Pulkki V, Karjalainen M. Multichannel audio rendering using amplitude panning [DSP applications][J]. Signal Processing Magazine, IEEE, 2008, 25(3): 118-122.

[30] 彭昌友,黄青华.传输函数和平面波入射角对合成声场的影响[J].数据采集与处理,2014,29(2):327-333.
Peng Changyou, Huang Qinghua. Influence of transfer function and plane wave incidence angle on synthesized sound field[J]. Journal of Data Acquisition and Processing. 2014, 29(2): 327-333.

[31] 殷福亮,汪林,陈喆. 三维音频技术综述[J].通信学报,2011,2(32):130-138.
Yin Fuliang, Wang Lin, Chen Zhe. Review on 3D

audio technology[J]. Journal on Communications, 2011,2(32):130-138.

[32] Gardner B, Martin K. HRTF measurements of a KEMAR dummy-head microphone[J]. Massachusetts Institute of Technology, 1994,280(280):1-7.

[33] Kahrs M, Brandenburg K. Applications of digital signal processing to audio and acoustics[M]. Heidelberg, Germany: Springer, 1998.

[34] Qu T, Xiao Z, Gong M, et al. Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap[J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2009,17(6):1124-1132.

[35] Tan C, Gan W. User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems[J]. Electronics letters, 1998, 34(25): 2387-2389.

[36] Middlebrooks J C, Macpherson E A, Onsan Z A. Psychophysical customization of directional transfer functions for virtual sound localization[J]. The Journal of the Acoustical Society of America, 2000,108(6):3088-3091.

[37] Zhang M, Kennedy R A, Abhayapala T D, et al. Statistical method to identify key anthropometric parameters in HRTF individualization[C]//Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on. Edinburgh, USA: IEEE, 2011:213-218.

[38] Lossius T, Baltazar P, de La Hogue T. DBAP-distance-based amplitude panning[M]. Ann Arbor, MI: MPublishing, University of Michigan Library, 2009.

[39] Strutt J W, Rayleigh B. The theory of sound[M]. Volumn 1. New York: Dover Publications, 1877.

[40] Dunai L, Hartmann W M. Frequency dependence of the interaural time difference thresholds in human listeners[J]. The Journal of the Acoustical Society of America, 2011,129(4):2485-2499.

[41] Bernstein L R, Trahiotis C. Lateralization produced by interaural intensitive disparities appears to be larger for high-vs low-frequency stimuli[J]. The Journal of the Acoustical Society of America, 2011,129(1):15-20.

[42] Hershkowitz R M, Durlach N I. Interaural time and amplitude JNDs for a 500 Hz tone[J]. The Journal of the Acoustical Society of America, 1969, 46(6B): 1464-1467.

[43] Domnitz R H, Colburn H S. Lateral position and interaural discrimination[J]. The Journal of the Acoustical Society of America, 1977,61(6):1586-1598.

[44] Yost W A, Dye Jr R H. Discrimination of interaural differences of level as a function of frequency[J]. The Journal of the Acoustical Society of America, 1988, 83(5):1846-1851.

[45] Klumpp R G, Eady H R. Some measurements of interaural time difference thresholds[J]. The Journal of the Acoustical Society of America, 1956,28(5): 859-860.

[46] Mills A W. Lateralization of high-frequency tones[J]. The Journal of the Acoustical Society of America, 1960,32(1):132-134.

[47] Grantham D W. Interaural intensity discrimination: Insensitivity at 1000 Hz[J]. The Journal of the Acoustical Society of America, 1984, 75(4): 1191-1194.

[48] Gabriel K J, Koehnke J, Colburn H S. Frequency dependence of binaural performance in listeners with impaired binaural hearing[J]. The Journal of the Acoustical Society of America, 1992,91(1):336-347.

[49] Hu R, Dong S, Wang H, et al. Perceptual characteristic and compression research in 3D audio technology[C]//9th International Symposium on Computer Music Modelling and Retrieval. London, UK: Springer, 2013:82-98.

[50] Ando A, Matsui K. Perception of sound image elevation in various acoustic environments[C]//40th International Conference: Spatial Audio: Sense the Sound of Space. [S.l.]: Audio Engineering Society, 2010:460-466.

[51] Perrott D R, Saberi K. Minimum audible angle thresholds for sources varying in both elevation and azimuth[J]. The Journal of the Acoustical Society of America, 1990,87(4):1728-1731.

[52] Grantham D W, Hornsby B W, Erpenbeck E A. Auditory spatial resolution in horizontal, vertical, and diagonal planes[J]. The Journal of the Acoustical Society of America, 2003,114(2):1009-1022.

[53] Barreto A, Faller K J, Adjouadi M. 3D sound for human-computer interaction: Regions with different limitations in elevation localization[C]//Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility. Pittsburgh, USA: ACM, 2009:211-212.

[54] Zahorik P, Brungart D S, Bronkhorst A W. Auditory distance perception in humans: A summary of past and present research[J]. Acta Acustica United with Acustica, 2005,91(3):409-420.

[55] Mershon D H, King L E. Intensity and reverberation as factors in the auditory perception of egocentric distance[J]. Perception & Psychophysics, 1975,18(6): 409-415.

[56] Brungart D S. Auditory localization of nearby sources. III. stimulus effects[J]. Journal of the Acoustical Society of America, 1999, 106(6): 3589-3602.

[57] 梁之安,杨琼华,林华英.声源定位与声源位置辨别阈[J].声学学报,1966,3(1):27-34.
Liang Zhian, Yang Qionghua, Lin Huaying. Localization of sound source and difference threshold of sound source position[J]. ACTA Acoustica, 1966,3(1):27-34.

[58] 吴锋,高下,李志宏,等.声音空间定位测听系统的设计与实现[J].第四军医大学学报,2001,22(7):656-658.
Wu Feng, Gao Xia, Li Zhihong, et al. Devising and initial realization of testing hearing system for sound location[J]. Journal of the Fourth Military Medical University, 2001,22(7):656-658.

[59] 纪丽红.动态听觉机制的探讨[D].杭州:浙江大学,2005.
Ji Lihong. Research on dynamic hearing mechanism [D]. Hangzhou: Zhejiang University, 2005.

[60] Shuixian C, Ruimin H, Yutian L, et al. Frequency dependence of spatial cues and its implication in spatial stereo coding[C]//Computer Science and Software Engineering, 2008 International Conference on. Wuhan, China: IEEE, 2008:1066-1069.

[61] 涂卫平.双耳强度差线索感知特性分析及应用研究[D].武汉:武汉大学,2011.
Tu Weiping. Analysis of binaural intensity difference and application [D]. Wuhan: Wuhan University, 2011.

[62] Shuixian C, Ruimin H, Naixue X. A multimedia application: Spatial perceptual entropy of multichannel audio signals [J]. EURASIP Journal on Wireless Communications and Networking, 2010,2010:1-13.

[63] Johnston J D, Ferreira A J. Sum-difference stereo transform coding[C]//Acoustics, Speech, and Signal Processing, IEEE International Conference on. San Francisco, USA: IEEE,1992,2:569-572.

[64] Herre J, Brandenburg K, Lederer D. Intensity stereo coding[C]//Audio Engineering Society Convention 96. New York, USA: Audio Engineering Society, 1994:9402-9496.

[65] Faller C, Baumgarte F. Efficient representation of spatial audio using perceptual parametrization[C]//Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on. New Platz, NY: IEEE, 2001:199-202.

[66] Breebaart J, van de Par S, Kohlrausch A, et al. Parametric coding of stereo audio[J]. EURASIP Journal on Applied Signal Processing, 2005, 2005: 1305-1322.

[67] Makinen J, Bessette B, Bruhn S, et al. AMR-WB +: A new audio coding standard for 3rd generation mobile audio services[C]//Acoustics, Speech, and Signal Processing (ICASSP'05), IEEE International Conference on Philadelphia, US: IEEE, 2005:1109-1112.

[68] Briand M, Martin N, Virette D. Parametric representation of multichannel audio based on principal component analysis[C]//Audio Engineering Society Convention 120. Paris, France: Audio Engineering Society, 2006:1305-1308.

[69] Gerzon M A, Craven P G, Stuart J R, et al. The MLP lossless compression system[C]//17th International Conference: High-Quality Audio Coding. New York, US: Audio Engineering Society, 1999: 243-260.

[70] Hellerud E, Solvang A, Svensson U P. Spatial redundancy in higher order ambisonics and its use for lowdelay lossless compression [C]//Acoustics, Speech and Signal Processing, ICASSP 2009, IEEE International Conference on. Taipei, China: IEEE, 2009:269-272.

[71] Elfitri I, Gunel B, Kondoz A M. Multichannel audio coding based on analysis by synthesis[J]. Proceedings of the IEEE, 2011,99(4):657-670.

[72] Cheng B, Ritz C, Burnett I. A spatial squeezing approach to ambisonic audio compression[C]//Acoustics, Speech and Signal Processing, ICASSP 2008, IEEE International Conference on. Las Vegas, NV: IEEE, 2008:369-372.

[73] Cheng B, Ritz C, Burnett I, et al. A general compression approach to multi-channel three-dimensional audio[J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2013,21(8):1676-1688.

[74] Breebaart J, Disch S, Faller C, et al. MPEG spatial audio coding/MPEG surround: Overview and current status[C]//Audio Engineering Society Convention 119. [S.l.]: Audio Engineering Society, 2005:1-17.

[75] Choi S J, Jung Y, Kim H J, et al. New CLD quantization method for spatial audio coding[C]//Audio Engineering Society Convention 120. [S.l.]: Audio Engineering Society, 2006:422-426.

[76] Kim K, Beack S, Seo J, et al. Improved channel level difference quantization for spatial audio coding[J]. Electronics and Telecommunications Research Institute Journal, 2007,29(1):99-102.

[77] Dong S, Hu R, Wang X, et al. Expanded three-channel mid/side coding for three-dimensional multichannel audio systems[J]. EURASIP Journal on Audio, Speech, and Music Processing, 2014,2014(1): 1-13.

[78] Cheng B, Ritz C, Burnett I, et al. A general compression approach to multi-channel three-dimensional audio[J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2013,21(8):1676-1688.

[79] Ando A. Conversion of multichannel sound signal maintaining physical properties of sound in reproduced sound field[J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2011, 19(6): 1467-1475.

作者简介:胡瑞敏(1964-),男,教授,研究方向:多媒体信号处理,E-mail:hurm1964 @gmail.com;王晓晨(1980-),男,博士,研究方向:音频编解码;张茂胜(1981-),男,博士,研究方向:三维音频处理;李登实(1977-),女,博士,研究方向:三维声场重建;王松(1983-),男,博士,研究方向:三维音频处理;高丽(1981-),女,博士,研究方向:三维音频编码;杨乘(1979-),男,博士,研究方向:三维音频编码;杨玉红(1974-),女,副教授,研究方向:语音频质量评价、语音频信号处理及编码。