

基于声场重建的三维音频技术

胡瑞敏, 张灵鲲, 王晓晨, 柯善发, 胡晨昊, 李登实

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

摘要: 近年来, 随着人们对视听体验的要求不断提高, 三维(空间)音频技术被广泛应用于娱乐影音等领域, 并取得显著进展。如何便捷地创造一个具有沉浸式空间听觉体验的环境是一直以来的研究热点。因此综述了主流三维声场重建的相关技术, 并分析了其优缺点, 主要讨论了三维声场的表达与维度、不同扬声器下的三维声场重建以及相应的混响均衡技术。最后, 归纳了三维声场重建技术目前存在的问题, 并展望未来发展的新方向。

关键词: 三维音频; 声场重建; 扬声器阵列; 球谐函数; 混响均衡

中图分类号: TN911 **文献标志码:** A

Review on Sound Field Reproduction Based Three-Dimensional Audio Technology

HU Ruimin, ZHANG Lingkun, WANG Xiaochen, KE Shanfa, HU Chenhao, LI Dengshi

(School of Computer Science, Wuhan University, National Engineering Research Center for Multimedia Software, Wuhan 430072, China)

Abstract: In recent years, spatial audio technology has been widely used and developed in a variety of fields. Many studies have been focused on the flexibility of creating an immersive spatial auditory experience. In this review, the advantages and disadvantages of the state-of-art approach have been analyzed. It mainly involves the dimension and representation of sound field, the reproduction of sound field, and corresponding room equalization technology. Finally, we discuss the current problems and predict possible promising directions for this research topic.

Key words: 3D audio; sound field reproduction; loudspeaker array; spherical harmonics; room response equalization

引言

自2009年《阿凡达》上映以来, 三维(Three-dimension, 3D)音视频技术受到了广泛的关注, 并得到了长足的发展。现在无论是从电影院里的杜比全景声, 还是家里的5.1或者7.1系统, 再到耳机上使用的空间音频技术, 都是为了能够让听音者产生较好的沉浸式音频体验。而MPEG-H标准也成为新的沉浸式空间音频标准^[1], 通过MPEG-H听音者可以身临其境的感受声音, 不仅仅是来自周围的声音, 还

基金项目: 国家重点研发计划(2017YFB1002803)资助项目; 国家自然科学基金(U1736206, 61761044)资助项目; 深圳市科技计划基础研究(JCYJ20170818143246278)资助项目。

收稿日期: 2021-06-29; **修订日期:** 2021-09-07

包括来自上方的声音,通过将声音渲染为可移动的物体而不是静态的物体来获得更真实的感知现实世界。法国电信2012年就已经提出了3D音频的定义^[2],其必须能够重建水平、高度以及距离三个自由度上的声像。然而现有的产品或者技术,离实现真正的3D音频还具有一定的差距,为了实现逼真的空间音频体验,需要大量的设备来进行回放,并且回放的效果还会受到设备的数目、布置以及回放环境的混响条件等因素的影响,这是制约着当前3D音频技术发展的重要问题^[3]。随着近年来电影产业以及VR/AR技术的发展,3D音频技术也成为了当前的一个热门研究课题。近5年来,在IEEE TASP、JASA、ICASSP、ICME等声学 and 信号处理知名期刊会议上发表的与声场重建相关的文章数量一直保持在80篇左右。国内外大量从事声学、音频信号处理、多媒体技术的相关学者进行了大量的研究工作。本文将综述其中有关3D音频技术的基本原理,以及相关信号处理方法在3D音频技术中的应用。

1 3D音频技术的基本原理

3D音频的基本原理是通过对物理空间中的声波进行还原重建,从而让身处其中的听音者再次感受到原始的声音场景,以此来实现沉浸式的空间音频体验。声音是由空间中的介质振动产生的,而振动的粒子之间就存在着力的作用,根据传统声学理论^[4],空间声场主要由声压(Sound pressure)和粒子速度(Particle velocity)两个物理量来表征,其中声压为标量,而粒子速度是空间中各点处声压的梯度,为矢量。由于粒子速度可以由声压来进行计算转换,在大多数研究中为了计算简便,主要以声压作为主要的研究对象。

由于声压能够对空间声场进行表征,其包含了发出声音的对象的距离、方位和宽度等信息,因此如何对空间中某一区域中的声压进行精确的重建对较好的3D音频体验具有重要意义。

声场重建系统可以重建特定场景下的声场。它被广泛用于为听众创建3D音频体验,基本实现方式为:在特定混响条件下,通过一组布置好的扬声器阵列来对声场(声压)进行重建。根据想要重建的物理声场,通过不同的理论计算出扬声器的驱动信号,并利用驱动信号回放出目标声学环境,从而实现听音者的沉浸式空间音频体验。基本的空间声场重建模型为^[5]

$$\min \|\boldsymbol{p} - \boldsymbol{p}_0\|^2 \quad (1)$$

$$\text{subject to } \boldsymbol{p} = \boldsymbol{H}(s) \quad (2)$$

式中: \boldsymbol{p} 为空间声场的量化指标,通常为声压的集合; \boldsymbol{H} 代表着扬声器阵列在当前环境下,从扬声器到目标声场的重建位置之间的空间传递函数,其表征了声波经过环境的反射以及衍射等作用后,到达最终位置处的滤波作用; s 为各扬声器的驱动信号。整个3D声场重建理论即围绕着如何在给定布置下,利用空间传递函数 \boldsymbol{H} ,求解出能够最佳逼近目标空间声场 \boldsymbol{p}_0 的扬声器驱动信号来进行研究。在很多研究中,为了简化计算,通常优先在理想条件下(即消声室无混响的环境下)进行声场重建,并且将放置在1.4 m以外的扬声器看作点声源,其辐射出的声波近似为球面波^[6]。然后才通过混响均衡的方法来对扬声器的驱动信号进行均衡^[5]

$$s = \boldsymbol{R}(s') \quad (3)$$

式中: \boldsymbol{R} 为对应的均衡函数,实现理想环境下扬声器驱动信号到混响环境下的驱动信号的映射。求解新的驱动信号 s' 使其在混响环境下依然能够达到理想条件下想要重建的声场效果。

通过以上分析可以发现,理想条件下3D声场的重建以及混响均衡对3D音频系统的实现有重要意义,这也是近些年3D音频技术发展的两个核心问题。

2 3D声场的重建

2.1 3D声场的表达与维度

球谐理论是声场分析及重建领域中很重要的一个理论。声场的球谐表达是波动方程的一组解^[7],它通过一组互相正交的球谐函数(Spherical harmonics function)对空间声场进行表达。对于空间中某区域中的任意一点 x 处的声场,可以通过球谐函数的线性组合进行表达^[4]

$$p(x; f) = \sum_{n=0}^{\infty} \sum_{m=-n}^n C_{nm}(f) j_n\left(\frac{2\pi}{c} f |x|\right) Y_{nm}(\hat{x}) \quad (4)$$

式中: $C_{nm}(\cdot)$ 为对应球谐函数 $Y_{nm}(\cdot)$ 的系数, c 为声波的传播速度, f 为当前计算频率, $|x|$ 为点 x 的到原点的距离, \hat{x} 为点 x 的方向向量, $j_n(\cdot)$ 为第1类 n 阶球贝塞尔函数, $Y_{nm}(\cdot)$ 为文献[7]定义的球谐函数。

$$Y_{nm}(\hat{x}) = \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}} P_{n|m|}(\cos\theta) e^{im\phi} \quad (5)$$

式中 $P_{n|m|}(\cdot)$ 为联合勒让德函数。

该函数的表现形式如图1所示。

球谐函数在单位球面上满足正交特性。利用正交性原理,可以得到对应球谐系数的计算方式,这也是基于球谐理论推导出的麦克风采集原理。当采用离散点处的声压对空间连续声场进行估计时,就得到空间连续声场的采集方法,其采集精度随着麦克风的数量的增加而提升

$$\tilde{C}_{nm}(f) = \frac{1}{j_n\left(\frac{2\pi}{c} f |x|\right)} \sum_{q=1}^Q p(Rx_q; f) Y_{nm}^*(\hat{x}_q) \omega_q \quad (6)$$

式中: $p(Rx_q; f)$ 即球表面上麦克风采集到的声压;为了弥补对无穷级数截断以后产生的误差, ω_q 为在各点处增加的权重。式(6)也能够用于计算虚拟声源的球谐表达系数。

尽管通过球谐函数已经能够对声场进行精确的表达,但是在实际使用球谐的表达过程中,需要对声场的无穷球谐级数的表达进行截断。一方面对无穷级数的截断会使得球谐表达产生误差,另一方面又需要确定需要用多少个球谐函数来对其进行表达,使得表达后的声场能够满足目标精度。这就涉及声场重建中的球谐表达维度以及截断误差的研究。

通过多少扬声器能够在特定范围内重建声场的问题本质上是无线通信环境下多通路场(Multipath fields)的空间维度问题^[8-9],该理论能够用于通信系统范围的扩大、容量增加、数据率提升以及降低误码率等。对无线通信环境下多通路的空间维度的研究主要分为两种理论,一种是基于统计学数据的统计模型,另一种是理论分析模型。

对于统计学模型,主要是根据采集或生成的大量数据中的参数特征进行分析,而其中比较常用到的特征包括了到达角^[10]、空间信号相关性^[11]和多径分布的变化^[12]等。另外,对于理论分析方法有一个基本的假设,即多离散远场源模型,该假设认为空间多通路场是由多个远场的源产生的信号/波的叠加而构成的。不同的散射源会产生不同的场分布,而多个散射源在目标位置处按照自身权重加权后线性叠加得到当前的空间多通路场,进而用于对空间场的维度进行分析。

在声场重建领域较为著名的理论是Kennedy等^[13]提出的理论,被广泛运用于确定所需重建声场的扬声器数目以及所需采集声场的麦克风数目。该理论给出了给定空间范围以及频率范围内,空间多路

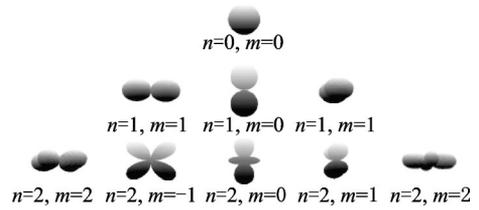


图1 各阶球谐函数示意图

Fig.1 Schematic of spherical harmonic function for each order

场的维度的一个上限 $N = \lceil ekr/2 \rceil$, 其中 e 为自然常数; k 为给定频率 f 对应的波数, 具体有 $k = \frac{2\pi f}{c}$, c 为声波的传播速度; r 为给定的空间范围半径, $\lceil \cdot \rceil$ 为上取整函数。当用于二维空间(平面区域)重建时该维度上限为 $2N + 1$, 当用于 3D 空间重建时, 该维度为 $(N + 1)^2$ 。

该方法推导出来的维度上界一直被广泛使用至今^[14], 根据该维度对球谐表达式进行截断后的截断误差能够保持在 4% 以下。虽然该方法给出了一个空间维度的上界, 但是并不代表着在任何应用场景下都必须具有那么高的维度才能够对空间场进行表达, 这能够为重建环境下初步确定扬声器以及麦克风的数量提供理论指导, 但是也存在进一步的降低空间, 后续很多研究都能够在一定的情况下突破这个上界^[15], 利用更少的扬声器或者麦克风实现声场的重建或者采集^[16]。

由于球谐表达的截断误差仅仅是由声场的近似表达产生的误差, 另外在声场重建过程中的各种理论计算也会产生一定的误差。这两种误差共同构成了声场重建过程中的误差, 而第 1 种误差(即截断误差)目前并未有新理论能够对声场的球谐表达进行替代, 一直被广泛应用到现在。第 2 种误差(即重建误差)主要受到扬声器的数目和排布以及相应的重建理论的影响。

2.2 标准扬声器阵列重建

第一个实际展示出的空间声场重建案例是在 1968 年由伊利诺伊理工学院的 Camras 完成的^[17]。在这个案例中, 通过布置在目标区域边界上的大量扬声器进行回放, 扬声器产生的声波在目标区域内叠加从而得到目标声场, 并且能够让听音者自由地在该重建环境中的区域内移动, 而保持相应的听觉体验。

为了能够更便捷地让听音者感受到原始声音场景下的沉浸式体验, 1985 年, Gerzon^[18]提出了最初的 Ambisonics 声场重建系统, 该方法使用 4 个扬声器组成的阵列对声场进行重建。该技术最早是开发来进行声场录制的, 后来被用到了声场重建方案中。该方案基于目标声场的一阶球谐系数分解而设计。

尽管最初的四通道 Ambisonics 系统能够一定程度上的重建声场, 然而仅能在空间单点附近达到 4% 左右的重建误差。为了进一步扩大重建的区域, 最初的 Ambisonics 被改进为高阶 Ambisonics (Higher orders of ambisonics, HOA) 系统^[19]。该方法通过对声场的更高阶的球谐表达进行重建, 可以在 1 kHz 的频率下通过 25 个扬声器在 0.2 m 半径的球体内以 3%~5% 的误差来重建声场。但是该方法也需要将扬声器均匀的布置到围绕目标区域的球面上, 来实现区域中心位置处的声场的重建^[20-21]。该方法在对平面波声场进行球谐分解的同时, 还对分解以后不同截断阶数的球谐系数对应的声场截断误差进行了分析, 根据经验法则给出了一个与目标重建波数 k 以及目标重建区域半径 r 有关的截断阶数建议: $N = \lceil kr \rceil$, 在这个截断阶数下能够保证大约 4% 的声场误差。进一步根据截断阶数判断所需要重建的声场的维度, 进而确定所需的最少扬声器数目。与 2.1 节中提到的空间多通路场给出的维度定理 $N = \lceil ekr/2 \rceil$ 具有非常小的差距, 对后续关于 Ambisonics 的研究都产生了较大影响^[22-24]。

该方法还有两个特别值得注意的地方。(1) 虽然按照给定的截断阶数大致能够确定声场的截断误差, 但在实际重建过程中, 如果按照所给出的 N 来确定对应的扬声器数目是不够的, 该数量下的扬声器并不一定能够产生那么多的声场模式(球谐函数), 从而难以保证在满足阶数条件后声场的重建误差。(2) 该方法通过求逆的方式来完成对扬声器的驱动信号的求解, 系数矩阵的条件数对结果影响较大, 这种情况下重建阵列产生的误差是难以控制的, 重建产生的误差可能会很大。而扬声器阵列的排布对系数矩阵具有很大的影响力, 进而产生了两种主要研究思路:(1) 以标准的球面排布的标准阵列模式进行声场重建来保证系数矩阵的条件数保持在较好的情况下;(2) 以不同的非标准阵列来研究声场重建的

效果,这种情况下需要探索具有较好条件数的扬声器阵列布置或者探索新的求解路线。

正因为文献[19]中的HOA方法需要通过求逆来获取扬声器的驱动信号,为了减弱求逆产生的影响,Poletti^[25]提出了一种减少求解扬声器驱动信号过程中求逆运算的简单源法(Simple source approach, SSA)。该方法假设声场是由覆盖在目标区域表面上的无穷多个简单声源辐射出的声波所构成的,并且给出了相应的证明,即封闭区域表面上的次级声源产生的声波可以表达区域内的任意声场。这种情况下目标声场是由球面上的简单声源产生的声场的积分构成,而球面上简单声源的构成是具有解析解的,两者结合直接推导出SSA的求解公式,避免了对扬声器产生的球谐模式矩阵的求逆,提升了基于球谐的声场重建方法的性能。

Ahrens等^[26]又在标准的二维圆形扬声器阵列的情况下进一步推导出了面向平面波重建的SSA。进一步证实了SSA在大量扬声器情况下的有效性。然而SSA是利用了目标区域表面的无穷次级声源来构成任意声场的理论基础推导而来,实际使用中需要将这个表达离散化,而离散化会产生怎样的误差却并未被较好分析。

Wu等^[27]将SSA拓展到了3D情况,分析了SSA在不同程度的离散化情况下产生的误差变化情况,同时又通过在计算离散分布的扬声器驱动信号权重时乘以一个面积权重来进一步提高该方法的性能。总之,SSA与传统基于求逆的最小二乘法相比,能够更好地从物理本质层面上来体现其解的意义,也在一定程度上降低了传统HOA算法的最小二乘解在求逆时的不稳定性。

除了经典的HOA算法以及SSA以外,还有一类最直观的声压匹配算法(Pressures matching, PM)。该类方法直接根据扬声器阵列在目标位置处产生的声压线性叠加后与目标声压值进行匹配,利用最小二乘法求解得到对应的扬声器驱动信号^[28-29]。然而这类在离散位置处匹配声压的方法,虽然能够在目标位置处达到比较好的声压匹配效果,但是会产生一些不良的后果,例如,在目标位置附近的声压不受控制,可能会产生较大误差。目前的研究主要还是以面向连续声场的声场重建为主。

在标准阵列上的基础理论研究在近些年并未发生本质改变,更多的目光聚集在如何通过便捷性的阵列对声场进行重建,或者在扬声器数目较少以及布置不佳的情况下更好地重建3D声场来实现听众的沉浸式听音感受,即非标准扬声器阵列下的3D声场重建技术。

2.3 非标准扬声器阵列重建

2.3.1 特定布置扬声器阵列的重建

为了提高用于声场重建的扬声器阵列布置的便捷性,有两种主要的思路来进行改进:(1)从球面扬声器阵列的理论为起点进行变换,找到能够更好布置扬声器的方案及其理论,其中最主要的是基于多圈扬声器的扬声器阵列;(2)直接从布置的便捷性角度来进行分析推导,优化特定布置下的重建效果,以逼近球面阵列的重建效果,其中常见的形状有矩形扬声器阵列、线性和平面扬声器阵列以及双平面扬声器阵列。

在分析声场重建理论时,Fazi等^[30-32]引入了泛函分析的框架。与上面提到的SSA的思想类似,该框架也是从目标重建区域表面的连续次级声源对内部声场的表达开始的,通过构造出的算子,并利用奇异值分解(Singular value decomposition, SVD),将内部声场的分布进行表达,最终实现用特征值对目标区域表面的次级声源强度(即扬声器驱动信号)的表达。分别在球面坐标系以及柱面坐标系上进行求解,可以得到标准的HOA解以及通过圆形扬声器阵列进行声场重建的解。这个方法在泛函框架的基础上实现了两个同心圆上以及球体上的声场重建的推导。

Zhang等^[33]在此基础上将该框架扩展到了3D声场重建上,该方法首先假设3D空间中有无穷多个连续分布在不同高度上的圆形扬声器阵列,并分析其中单一圆形扬声器阵列能够有效覆盖到的空间角

度,发现其中的高度模式主要由联合勒让德函数(即式(5)中的 $P_{n|m}(\cdot)$)决定,而离散化后的联合勒让德函数在最高阶数以及最高项(即 $n=N, m=N$)时主瓣宽度最窄,所有更小项的范围都能够覆盖其主瓣。选择主瓣的两个零点之间的范围作为圆形扬声器阵列的有效范围能够最大化的包括其有效范围,得到了两个同心圆之间的高度角差距 $\Delta\theta < 2/N$,其中 N 为圆形扬声器阵列对应的阶数(对应的扬声器数为 $2N+1$),从而确定了多圈扬声器的构造。在重建声源的时候选择需要激活的圈层来对声场进行重建,类似于传统Panning算法的思路,通过相邻的圆形扬声器阵列来进行重建。而不同于传统的HOA方法,在重建任意声场的时候都使用所有扬声器来进行重建。该方法在500 Hz的频率下通过15个扬声器在0.5 m范围内的球体内以1%左右的误差重建了声场。

另一类是从便捷性角度出发的扬声器阵列。其中Naoe^[34]提出了一种8个扬声器的矩形扬声器阵列,刚好把球面阵列上的扬声器转化到了一个立方体的8个顶点位置,跟家庭环境下的房间位置具有一致性。在相应的主观评测中,验证了17个方向上的声波重建效果,其中14个方位的效果都具有良好的定位效果。Ahrens等^[35]研究了平面以及线性扬声器阵列的重建性能,通过分布式的布置线性阵列来实现3D声场重建。与立方体形状的扬声器布置类似,Chen等^[36]设计了一种上下两层结构的扬声器阵列,分布在顶部和底部,通过设计的阵列分别对奇数阶和偶数阶的声场进行重建,与立方体形状的阵列相比,一方面增加了扬声器的数目,提高了扬声器的精度,另一方面又保证了顶部和底部的相对便捷性的布置。

尽管不同形状的扬声器阵列都被设计用来对3D声场进行重建,然而由于声场重建的理论基础,想要对空间声场进行重建,所用到的扬声器数目必须满足空间采样定理,必然需要用到数量较多的扬声器。为了进一步提高在扬声器数目较少或者扬声器布置不规则的条件下听音者的3D听音体验,还有很多研究聚焦于在给定的扬声器布置下来对3D声场进行重建,此时需要重建的不仅仅是3D声场本身,还有一些跟主观听音感受有关的物理线索。

2.3.2 任意非标准扬声器阵列的重建

在基于耳机的3D音频技术中,主要是通过采集到的头相关传输函数(Head related transfer function, HRTF)来对双耳需要接收到的听音信号进行重建^[37]。其中就涉及了主观听音的物理线索,例如双耳间时间差(Interaural time difference, ITD)和双耳间强度差(Interaural level difference, ILD)是对人类的定位感影响最明显的两个物理量^[38]。与基于耳机的3D音频技术类似,基于物理声场的重建中也有类似的对主观听音感受有明显影响的物理量。从文献[7]中可知,空间声场主要由声压(Sound pressure)和粒子速度(Particle velocity)两个物理量来表征,其中声压为标量,而粒子速度是空间中各点处声压的梯度,为矢量。由于粒子速度可以由声压来进行计算转换,在大多数研究中为了计算简便,主要以声压作为主要的研究对象。然而,粒子速度也被证明是与主观听音感受有关的物理量^[39-40]。

在22.2多声道系统下混算法^[40]中用到了这样的方法,当扬声器数目较少时,用3个包围了22声道系统中任意声道的扬声器来对其进行替代,实现由22声道到更低声道的下混,其核心原理为保持3个扬声器与对应的替代扬声器在中心点处产生的声压和粒子速度不变。在该研究中,该方法下混到的10声道和8声道系统都具有较好的主观听音效果。值得注意的是,该方法借鉴的是经典的基于向量的幅度平移(Vector-based amplitude panning, VBAP)算法^[41]中的思想,通过3个扬声器到中心点处的矢量的线性叠加来表达虚拟声源。VBAP算法已经在工业界得到了大量的应用,其主观听音效果的有效性也有粒子速度重建的贡献。

为了在声场重建系统中控制粒子速度。Shin等^[42]提出了一种能够对粒子速度进行控制的声场重建技术,单独通过粒子速度的重建来实现声场的重建。粒子速度是一个矢量,仅对粒子速度进行控制,由于维度变为了声压维度的3倍,在相同的扬声器阵列下,能够有效重建的点数会大幅度降低,由于该

算法仅仅对粒子速度进行重建,明显提高了整个声压场的重建误差。在后续工作中提出了一种基于声压和粒子速度的联合优化方法^[42-43],并分析了粒子速度如何对声场重建的误差产生影响。另外,此类基于粒子速度的方法都是通过对离散空间表面区域上的粒子速度进行控制,而并非控制连续空间内的粒子速度。为了保证声场重建的有效性,需要利用大量的扬声器控制大量点处的粒子速度,这种方式效率较低。Zuo等^[44]提出了一种对连续空间中的声压和粒子速度联合优化的算法,该方法与声场的球谐表达类似,推导出了连续空间内部粒子速度的球谐表达方法,通过联合优化声压和粒子速度的重建误差来达到目标区域的声场控制。除了粒子速度以外,声强(Intensity)也被认为是对主观听音效果有影响的物理线索,在文献^[45]中通过对声强的球谐表达的推导来实现对空间声强场的控制。

2.4 房间混响均衡

房间响应均衡(Room response equalization, RRE)已在理论上进行了研究并应用于实践中,以改善声场重建的质量,从而减少家庭环境对重建系统的影响。在RRE系统中,房间传递函数(Room transfer function, RTF)表征了从扬声器到收听者的声音传播路径,通过设计合适的均衡器来对RTF进行均衡来实现混响均衡。该均衡器可通过多种方法得到。基本思想是使用麦克风测量特定位置处的房间脉冲响应,然后通过计算其逆来获得均衡器。通过将该均衡器应用到理想环境(自由场条件,不考虑环境反射和衍射等的影响)下的声场重建方案来实现混响条件下的声场重建。

Elliott和Nelson^[46]提出了最早的多点均衡方法之一,在该项研究中提出了一个用于设计声场重建系统中混响均衡器的方法,在当前房间中的多个点处,通过调整混响脉冲响应的系数,使得当前房间脉冲响应与目标脉冲响应均方误差的和最小,从而得到混响均衡滤波器。该研究同时给出了基于固定均衡器的和自适应均衡器的方法。该方法得到了较为广泛的应用,并且还被许多其他方法所改进^[47-49],但存在一个明显缺点,即通过均方误差的最小化不足以描述房间脉冲响应的相似性,并且在某些位置也无法设计出达到要求的均衡滤波器。另外,在车载混响均衡的背景下,Elliott等^[50]通过考虑有误差的加权平均的多点均衡算法对该方法进行了改进,该解决方案在所选位置处的响应中实现了提升,并且在其他位置效果也没有发生明显衰减。

进一步还可以利用不同空间分布的房间响应之间的相似性,通过选择不同的距离度量对它们进行聚类来,实现多个位置处的混响均衡。Mourjopoulos^[51]提出了一个思路:将房间内具有较大相似度的RTF组合在一起,组合之后的RTF被分为了少数几类,就可以通过对较少数量的均衡器进行均衡来实现整个范围的均衡。使用LPC进行分析,然后通过全极点滤波器对RTF进行建模,从而设计出最小相位均衡器。该分类的结果可以用作空间均衡库,根据收听者的实际位置,实现房间内各种不同位置处的混响均衡。但是该方法依然存在明显的缺点:必须采集和记录大量房间响应和均衡器并且需要实时跟踪收听者的位置。在文献^[52-53]中通过将频率变形和模糊c均值聚类应用于混响均衡中,之后由Carini^[54-55]改进后,大幅度降低了计算复杂度。Turmchokkasam和Mitaim^[56]还在混响均衡中引入了加权的模糊c均值聚类,通过对不同的房间脉冲响应样本以不同方式加权来说明它们对RRE的不同影响。

文献^[52-53]的模糊c均值聚类方法也是“多点原型方法”的第一个例子。这类方法通过测量得到的不同位置的房间响应来提取原型响应,通过该原型响应表示必须被均衡的感知声学信息。然后在此原型响应的基础上,使用间接或直接的方法来设计单个均衡器^[57]。Cecchi等^[58]研究了确定原型响应的不同方法。具体来说,就是将模糊c均值方法与计算平均值、中值、min-max和均方根平均值的方法进行了比较,并在对频谱进行平滑处理后通过Kirkeby算法^[59]或LPC分析得到均衡器,进行最小相位均衡。在所考虑的条件下,平均值方法给出了最好的结果,其他方法也提供了相近的性能效果。之后基于平均值的原型提取方法也与文献^[54]的方法相结合,被RRE的应用验证了效果^[55,60],主观听力测试证实

了该方法获得的良好结果。Primavera等^[61]通过考虑群延迟均衡进一步提升了这种方法的效果。

近年来,便携式设备的使用量大幅度增加,达到了非常高的规模。然而,由于扬声器的特性及其与房间环境的相互作用,许多这些设备仅仅能够满足基本的音频要求。通过考虑这些设备的声学特性并通过信号处理技术,可以部分地改善这种情况。Cecchi等^[62]通过引入多点均衡方法来改善手机等便携式设备上扬声器的非理想响应,客观测试和主观听力测试结果证实了该算法对个人便携设备上的混响均衡具有一定效果。同年,Czyzewski等^[63]提出了用于在平板电脑上进行混响均衡的频率响应线性化的静态和自适应算法,主观听力测试结果表明了该方法能够改善听音者的感知评价。

Lecomte等^[64]提出的混响均衡算法中,通过SVD将待求逆矩阵中奇异值小于阈值的去掉,从而降低条件数的大小,避免在较低频率时导致过大的扬声器响应而造成损伤。同年,Cecchi等^[65]综述了在混响均衡领域存在的大量问题,并指出这些问题主要是由于房间脉冲响应的可逆性不稳定造成的,而可逆性不稳定问题依然是当前房间混响均衡相关技术中的重难点问题。

3 3D声场重建技术展望

3D声场重建技术通过信号处理的技术控制扬声器阵列的驱动信号进行回放,让身处阵列中的听音者感受到沉浸式的3D听觉体验。近些年,关于3D声场重建的技术不断发展,但在扬声器阵列不规则或数目较少情况下的算法性能和应用场景的拓展等方面仍有一些问题有待进一步深入研究。

(1) 3D声场的表达问题

为了对连续空间中的声场进行重建,需要对连续空间中的声场进行表达以后,针对表达的3D声场特征(例如球谐系数)进行重建,从而实现连续空间内的3D声场重建。虽然球谐表达理论给出了很好的声场表达式,然而实际应用中,截断以后的球谐系数的重建误差最小并不一定得到最小的声场重建误差^[16]。由于不同阶球谐系数对不同声场的重建误差的影响程度不同,因声场而异,较为复杂,难以有效分析,导致基于截断球谐系数逼近的声场方法存在进一步降低误差的空间。近年来有研究人员探索通过机器学习的方法来寻求更低的声场重建误差。通过数据驱动的方式来对声场进行表达,有机会进一步降低误差以及突破维度定理的限制,从而减少所需扬声器数目。

(2) 感知相关声场物理量的重建问题

传统3D声场重建技术都是对原始物理环境进行重建,是面向物理实验设计的。在重建沉浸式听觉体验的时候,精确的物理重建是一种途径,也可以在此基础上考虑与感知有关的物理量的重建,通过加权等方式综合考虑对主观听音感受的影响程度^[15,44-45],在扬声器排布不够规则或者数目较少的条件下尽可能地提高听音者的3D音频体验,这是未来研究的一个重要方向。

(3) 个人音频系统

随着声场重建的技术不断发展,通过大量扬声器已经能够在一定范围内进行具有较好主观听音感受的声场重建^[66]。近年,一个称为个人音频系统(Personal sound zone)的应用场景逐渐引起大家的关注。该场景通过在多个区域中同时控制声场,使得不同区域之间互不干扰,形成所谓的亮区和暗区。与当前的主动降噪耳机应用场景有类似之处,但该技术能够不佩戴耳机即可享受独自的空间。该技术的实现还需要面对环境混响的影响、听音者走动、亮区和暗区在面朝方向上的相互遮挡等问题,这也是未来的重要研究方向之一。

(4) 高阶扬声器

混响环境会对声场重建效果产生影响,同时重建效果也会因为环境的变化产生明显的影响,其中一个原因是扬声器产生的散射声波受到环境反射以后再次到达听音位置处。为了降低环境混响对声场重建的效果,有许多研究提出了新型的扬声器,即所谓的高阶扬声器(Higher order source, HOS)^[67]。

在高阶扬声器上可以有多个扬声器,并且可以有用于均衡的麦克风。其设计的目标是:一方面通过设计不同方向的声波模式来减少目标方向以外的声波,降低混响的影响;另一方面,高阶扬声器上具有多个扬声器,具有复杂的辐射模式,有机会替代大量的单一模式扬声器,从而可以实现扬声器数目的减少。

(5) 基于移动麦克风的房间混响均衡

传统混响均衡算法需要在多个位置处测量房间脉冲响应,耗费大量的人力和时间成本。现在的研究中有一种基于移动麦克风的房间脉冲响应测量方法^[68-69],通过特定路线和速度移动的麦克风阵列高效地采集房间信息,能够明显提升混响条件下3D声场技术在实际应用中的性能。

4 结束语

3D声场重建技术在虚拟现实和娱乐影音等领域有着广泛的应用前景,是近年来音频信号处理领域中重要的研究之一。尽管国内对3D声场重建技术的研究起步较晚,但国内一些科研机构也取得了许多研究成果。随着相关技术的不断发展,相信会有越来越多的问题被突破,并被应用到实际场景中。

参考文献:

- [1] HERRE J, HILPERT J, KUNTZ A, et al. MPEG-H 3D audio—The new standard for coding of immersive spatial audio[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2015, 9(5): 770-779.
- [2] 胡瑞敏,王晓晨,张茂胜,等. 三维音频技术综述[J]. *数据采集与处理*, 2014, 29(5): 661-676.
HU Ruimin, WANG Xiaochen, ZHANG Maosheng, et al. Review on three-dimension audio technology[J]. *Journal of Data Acquisition and Processing*, 2014, 29(5): 661-676.
- [3] KOYAMA S, CHARDON G, DAUDET L. Optimizing source and sensor placement for sound field control: An overview[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020(28): 696-714.
- [4] WILLIAMS E G. Fourier acoustics: Sound radiation and nearfield acoustical holography[M]. [S.l.]: Elsevier, 1999.
- [5] BETLEHEM T, ZHANG W, POLETTI M A, et al. Personal sound zones: Delivering interface-free audio to multiple listeners[J]. *IEEE Signal Processing Magazine*, 2015, 32(2): 81-91.
- [6] WARD D B, ABHAYAPALA T D. Reproduction of a plane-wave sound field using an array of loudspeakers[J]. *IEEE Transactions on Speech and Audio Processing*, 2001, 9(6): 697-707.
- [7] 马大猷. 现代声学理论基础[M]. 北京: 科学出版社, 2004.
MA Dayou. Fundamentals of modern acoustics theory[M]. Beijing: China Science Publishing & Media Ltd, 2004.
- [8] RAPPAPORT T S. Wireless communications: Principles and practice[M]. New Jersey: Prentice Hall PTR, 1996.
- [9] KOHNO R. Spatial and temporal communication theory using adaptive antenna array[J]. *IEEE Personal Communications*, 1998, 5(1): 28-35.
- [10] KERMOAL J P, MOGENSEN P E, JENSEN S H, et al. Experimental investigation of multipath richness for multi-element transmit and receive antenna arrays[C]//Proceedings of 2000 IEEE 51st Vehicular Technology Conference Proceedings. [S.l.]: IEEE, 2000, 3: 2004-2008.
- [11] LOZANO A, PAPADIAS C. Layered space-time receivers for frequency-selective wireless channels[J]. *IEEE Transactions on Communications*, 2002, 50(1): 65-73.
- [12] WOLNIANSKY P W, FOSCHINI G J, GOLDEN G D, et al. V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel[C]//Proceedings of 1998 URSI International Symposium on Signals, Systems, and Electronics. [S.l.]: IEEE, 1998: 295-300.
- [13] KENNEDY R A, SADEGHI P, ABHAYAPALA T D, et al. Intrinsic limits of dimensionality and richness in random multipath fields[J]. *IEEE Transactions on Signal Processing*, 2007, 55(6): 2542-2556.
- [14] MA F, ABHAYAPALA T D, ZHANG W. Multiple circular arrays of vector sensors for real-time sound field analysis[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, 29: 286-299.

- [15] ZHANG L, WANG X, HU R, et al. Optimization of sound fields reproduction based higher-order ambisonics (HOA) using the generative adversarial network (GAN)[J]. *Multimedia Tools and Applications*, 2020 (80): 1-16.
- [16] ZHANG L, WANG X, HU R, et al. Estimation of spherical harmonic coefficients in sound field recording using feed-forward neural networks[J]. *Multimedia Tools and Applications*, 2021, 80(4): 6187-6202.
- [17] CAMRAS M. Approach to recreating a sound field[J]. *The Journal of the Acoustical Society of America*, 1968, 43(6): 1425-1431.
- [18] GERZON M A. Ambisonics in multichannel broadcasting and video[J]. *Journal of the Audio Engineering Society*, 1985, 33 (11): 859-871.
- [19] WARD D B, ABHAYAPALA T D. Reproduction of a plane-wave sound field using an array of loudspeakers[J]. *IEEE Transactions on speech and audio processing*, 2001, 9(6): 697-707.
- [20] WU Y J, ABHAYAPALA T D. Theory and design of soundfield reproduction using continuous loudspeaker concept[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, 17(1): 107-116.
- [21] POLETTI M, FAZI F M, NELSON P A. Sound-field reproduction systems using fixed-directivity loudspeakers[J]. *The Journal of the Acoustical Society of America*, 2010, 127(6): 3590-3601.
- [22] LI D, HU R, WANG X, et al. Loudspeaker triplet selection based on low distortion within head for multichannel conversion of smart 3D home theater[J]. *Concurrency and Computation: Practice and Experience*, 2020, 32(13): e4796.
- [23] TAKIDA Y, KOYAMA S, UENO N, et al. Reciprocity gap functional in spherical harmonic domain for gridless sound field decomposition[J]. *Signal Processing*, 2020, 169: 107383.
- [24] ZHANG J, SUN H, SAMARASINGHE P N, et al. Active noise control over multiple regions: Performance analysis[C]// *Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.]: IEEE, 2020: 8409-8413.
- [25] POLETTI M A. Three-dimensional surround sound systems based on spherical harmonics[J]. *Journal of the Audio Engineering Society*, 2005, 53(11): 1004-1025.
- [26] AHRENS J, SPORS S. Analytical driving functions for higher order ambisonics[C]//*Proceedings of 2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.]: IEEE, 2008: 373-376.
- [27] WU Y J, ABHAYAPALA T D. Theory and design of soundfield reproduction using continuous loudspeaker concept[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, 17(1): 107-116.
- [28] BETLEHEM T, TEAL P D. A constrained optimization approach for multi-zone surround sound[C]//*Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.]: IEEE, 2011: 437-440.
- [29] POLETTI M. An investigation of 2-D multizone surround sound systems[C]//*Proceedings of Audio Engineering Society Convention 125*. [S.l.]: Audio Engineering Society, 2008.
- [30] FAZI F M, NELSON P, CHRISTENSEN J E, et al. Surround system based on three-dimensional sound field reconstruction [C]//*Proceedings of audio Engineering Society Convention 125*. [S.l.]: Audio Engineering Society, 2008.
- [31] FAZI F M. Sound field reproduction[D]. Southampton, United Kingdom: University of Southampton, 2010.
- [32] FAZI F M, NELSON P A. Application of functional analysis to the sound field reconstruction[J]. *Proceedings of the Institute of Acoustics*, 2007, 29(7):1-9.
- [33] ZHANG W, ABHAYAPALA T D. 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, 22(7): 1184-1194.
- [34] 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]//*Proceedings of 2008 Second International Symposium on Universal Communication*. [S.l.]:IEEE, 2008: 36-41.
- [35] 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.
- [36] CHEN P, SAMARASINGHE P N, ABHAYAPALA T D. 3D exterior soundfield reproduction using a planar loudspeaker array[C]//*Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.]:

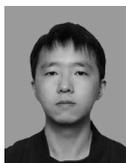
- IEEE, 2018: 471-475.
- [37] 殷福亮, 汪林, 陈喆. 三维音频技术综述[J]. 通信学报, 2011, 32(2):130-138.
YIN Fuliang, WANG Lin, CHEN Zhe. Review on 3D audio technology[J]. Journal on Communications, 2011, 32(2): 130-138.
- [38] ZHANG W, ABHAYAPALA T D, KENNEDY R A, et al. Insights into head-related transfer function: Spatial dimensionality and continuous representation[J]. The Journal of the Acoustical Society of America, 2010, 127(4): 2347-2357.
- [39] GANDEMER L, PARSEIHIAN G, BOURDIN C, et al. Perception of surrounding sound source trajectories in the horizontal plane: A comparison of VBAP and basic-decoded HOA[J]. Acta Acustica United with Acustica, 2018, 104(2): 338-350.
- [40] ANDO A. Conversion of multichannel sound signal maintaining physical properties of sound in reproduced sound field[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 19(6): 1467-1475.
- [41] LAITINEN M V, PIHLAJAMÄKI T, ERKUT C, et al. Parametric time-frequency representation of spatial sound in virtual worlds[J]. ACM Transactions on Applied Perception (TAP), 2012, 9(2): 1-20.
- [42] SHIN M, NELSON P A, FAZI F M, et al. Velocity controlled sound field reproduction by non-uniformly spaced loudspeakers [J]. Journal of Sound and Vibration, 2016, 370: 444-464.
- [43] BUERGER M, MAAS R, LÖLLMANN H W, et al. Multizone sound field synthesis based on the joint optimization of the sound pressure and particle velocity vector on closed contours[C]//Proceedings of 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). [S.l.]: IEEE, 2015: 1-5.
- [44] ZUO H, ABHAYAPALA T D, SAMARASINGHE P N. Particle velocity assisted three dimensional sound field reproduction using a modal-domain approach[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2119-2133.
- [45] ZUO H, SAMARASINGHE P N, ABHAYAPALA T D. Intensity based spatial soundfield reproduction using an irregular loudspeaker array[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 1356-1369.
- [46] ELLIOTT S J, NELSON P A. Multiple-point equalization in a room using adaptive digital filters[J]. Journal of the Audio Engineering Society, 1989, 37(11): 899-907.
- [47] BHARITKAR S, KYRIAKAKIS C. Immersive audio signal processing[M]. [S.l.]: Springer Science & Business Media, 2008.
- [48] BHARITKAR S, HILMES P, KYRIAKAKIS C. Robustness of spatial averaging equalization methods: A statistical approach [C]//Proceedings of Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers. [S.l.]: IEEE, 2002, 1: 184-187.
- [49] BHARITKAR S, HILMES P, KYRIAKAKIS C. Robustness of spatial average equalization: A statistical reverberation model approach[J]. The Journal of the Acoustical Society of America, 2004, 116(6): 3491-3497.
- [50] ELLIOTT S J, BHATIA L P, DEGHAN F S, et al. Practical implementation of low-frequency equalization using adaptive digital filters[J]. Journal of the Audio Engineering Society, 1994, 42(12): 988-998.
- [51] MOURJOPOULOS J N. Digital equalization of room acoustics[J]. Journal of the Audio Engineering Society, 1994, 42(11): 884-900.
- [52] BHARITKAR S, KYRIAKAKIS C. Perceptual multiple location equalization with clustering[C]//Proceedings of Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers. [S.l.]: IEEE, 2002, 1: 179-183.
- [53] KYRIAKAKIS C, BHARITKAR S. New factors in room equalization using a fuzzy logic approach[C]//Proceedings of Audio Engineering Society Convention 111. [S.l.]: Audio Engineering Society, 2001.
- [54] OMICIUOLO I, CARINI A, SICURANZA G L. Multiple position room response equalization with frequency domain fuzzy c-means prototype design[C]//Proceedings of IWAENC. Seattle, USA: [s.n.], 2008: 1-4.
- [55] CARINI A, CECCHI S, PIAZZA F, et al. Multiple position room response equalization in frequency domain[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 122-135.
- [56] TURMCHOKKASAM S, MITAIM S. Effects of weights in weighted fuzzy c-means algorithm for room equalization at multiple locations[C]//Proceedings of 2006 IEEE International Conference on Fuzzy Systems. [S.l.]: IEEE, 2006: 1468-1475.
- [57] KARJALAINEN M, PAATERO T. Equalization of loudspeaker and room responses using Kautz filters: Direct least squares design[J]. EURASIP Journal on Advances in Signal Processing, 2007(1): 1-13.

- [58] CECCHI S, PALESTINI L, PERETTI P, et al. Evaluation of a multipoint equalization system based on impulse response prototype extraction[J]. Journal of the Audio Engineering Society, 2011, 59(3): 110-123.
- [59] KIRKEBY O, NELSON P A, HAMADA H, et al. Fast deconvolution of multi-channel systems using regularisation[J]. Proceedings-Institute of Acoustics, 1996, 18: 2829-2832.
- [60] CECCHI S, PIAZZA F, CARINI A. Subjective evaluation of a multiple position room response equalization approach applied in frequency domain[C]//Proceedings of 2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA). [S.l.]: IEEE, 2011: 84-89.
- [61] PRIMAVERA A, CECCHI S, PIAZZA F, et al. Mixed time-frequency approach for multipoint room response equalization [C]//Proceedings of Audio Engineering Society Conference: 45th International Conference, Applications of Time-Frequency Processing in Audio. [S.l.]: Audio Engineering Society, 2012.
- [62] CECCHI S, VIRGULTI M, PRIMAVERA A, et al. Investigation on audio algorithms architecture for stereo portable devices [J]. Journal of the Audio Engineering Society, 2016, 64(1/2): 75-88.
- [63] CZYZEWSKI A, CIARKOWSKI A, KOSTEK B, et al. Adaptive personal tuning of sound in mobile computers[J]. Journal of the Audio Engineering Society, 2016, 64(6): 405-428.
- [64] LECOMTE P, GAUTHIER P A, LANGRENNE C, et al. Cancellation of room reflections over an extended area using Ambisonics[J]. The Journal of the Acoustical Society of America, 2018, 143(2): 811-828.
- [65] CECCHI S, CARINI A, SPORS S. Room response equalization—A review[J]. Applied Sciences, 2018, 8(1): 16.
- [66] HONG J Y, LAM B, ONG Z T, et al. Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications[J]. Building and Environment, 2019, 149: 1-14.
- [67] ELKO G W. Differential microphone arrays[M]//Audio Signal Processing for Next-Generation Multimedia Communication Systems. Boston, MA: Springer, 2004: 11-65.
- [68] HAHN N, SPORS S. Continuous measurement of impulse responses on a circle using a uniformly moving microphone[C]// Proceedings of 2015 23rd European Signal Processing Conference (EUSIPCO). [S.l.]: IEEE, 2015: 2536-2540.
- [69] MAZUR R, KATZBERG F, PHAN H, et al. Room equalization based on measurements with moving microphones[C]// Proceedings of 2017 Hands-Free Speech Communications and Microphone Arrays (HSCMA). [S.l.]: IEEE, 2017: 121-125.

作者简介:



胡瑞敏(1964-),通信作者,男,教授,博士生导师,研究方向:多媒体信号处理、人工智能、网络安全、云计算与大数据处理等,E-mail: hrm1964@163.com。



张灵鲲(1991-),男,博士研究生,研究方向:三维音频处理。



王晓晨(1980-),男,博士,研究方向:三维音频处理、音频编解码。



柯善发(1989-),男,博士研究生,研究方向:三维音频处理、音源分离。



胡晨昊(1991-),男,博士研究生,研究方向:三维音频处理、音频对象编码。



李登实(1977-),女,博士,研究方向:三维音频处理。

(编辑:张彤)