中国科学: 物理学 力学 天文学 2022 年 第 52 卷 第 4 期: 244303
SCIENTIA SINICA Physica, Mechanica & Astronomica physcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS

评 述 声学科学与技术前沿论坛专题

空间声原理及其相关的物理、听觉问题

谢菠荪*

华南理工大学物理与光电学院, 声学研究所, 广州 510641
*联系人, E-mail: phbsxie@scut.edu.cn

收稿日期: 2021-09-28; 接受日期: 2021-12-14; 网络出版日期: 2022-03-21
国家重点研发计划(编号: 2018YFB1403800)资助项目

摘要 空间声的目的是捡拾、传输和重放声场的空间信息, 以产生期望的空间听觉事件和感知. 空间声技术近年快速发展, 是声学领域的前沿课题, 并在众多科学研究、工程技术和文化领域有广泛的应用. 声场的物理分析重构与空间听觉机理是空间声的两个重要基础, 信号处理是实现空间声的技术手段. 因而空间声的研究在声场、听觉、实际实现方面涉及声学(物理)、人类听觉感知、信号处理等跨学科的基础与技术问题. 本文评述空间声的原理与发展, 主要科学与技术问题、研究进展及代表性应用, 并探讨了今后的技术和应用发展方向.

关键词 空间声, 空间听觉, 声重放

PACS: 43.60.+d, 43.55.+p, 43.66.+y

1 引言

声音是传播信息的重要媒介, 听觉(包括空间听觉)是人类获取外部信息的重要手段之一. 声源产生的辐射以波动的形式在空间传输, 包括直达和环境反射、散射的途径传输, 形成复杂的空间声场. 在可听声频率范围内, 声场包含有声源与声学环境的时间、空间听觉信息. 倾听者用双耳接收声场信息. 内耳对双耳声信息进行初步处理(如频率分析)并转换为听觉神经脉冲后, 再经高层神经系统综合处理, 形成各种听觉事件, 包括空间听觉事件(如对声源的定位, 对声学环境和场景的综合感知等)[1,2].

自从19世纪爱迪生发明留声机以来, 利用物理手段忠实地记录和重放声音信息一直是人类的技术追求. 20世纪以来, 物理学(声学)、电子与信息技术的发展, 使声音重放技术得到较大的发展, 并在实际中得到广泛的应用. 空间声的目的是捡拾(模拟)、传输(或记录)并重放声场的空间信息, 从而产生期望的空间听觉事件和感知[3]. 传统上, 空间声主要应用在文化和娱乐领域. 近年, 信号处理、通信、计算机与互联网技术的发展, 空间声技术得到很大的发展, 已拓展应用到一系列的科学研究、工程技术领域. 空间声是目前声学领域一个活跃的、有广泛应用前景的前沿方向.

空间声涉及广泛的基础科学与应用技术问题. 其中声源(扬声器)辐射、空间声场到倾听者双耳(鼓膜)传输主要涉及声学(物理)问题; 听觉感知涉及人类听觉信息处理的生理和心理机理的问题; 空间声的实现涉及信号与信息处理的技术问题. 声场的物理分析与

引用格式: 谢菠荪. 空间声原理及其相关的物理、听觉问题. 中国科学: 物理学 力学 天文学, 2022, 52: 244303
Xie B S. Principle of spatial sound and related physical and auditory problems (in Chinese). Sci Sin-Phys Mech Astron, 2022, 52: 244303, doi: 10.1360/SSPMA-2021-0273

© 2022 《中国科学》杂志社

人类空间听觉的机理是空间声的两个重要基础, 信号处理是实现空间声的重要手段. 而从应用方面看, 空间声会涉及更广泛的领域. 因而空间声是一个跨越声学、听觉生理与心理、信号与信息处理的交叉领域.

本文将评述空间声及其相关的物理、听觉问题. 包括空间听觉与空间声的原理、进展、科学与技术问题、应用和今后的发展方向等.

2 声场与空间听觉

声场定义为媒质中声波存在的空间. 空气可近似为均匀、各向同性的线性流体, 其中的声场可以用声压$p(\mathbf{r}, t)$作为空间位置$\mathbf{r}$和时间t的函数的唯一描述. 作为一种机械波, 声波满足波动方程或亥姆霍兹方程, 并遵循经典波动的一切规律, 包括叠加与干涉、反射、衍射、散射等. 倾听者进入声场后, 其自身的生理结构(如头部、耳廓等)对声波的散射作用, 将声场信息转换为双耳声压信号.

声场到双耳声压的转换在物理上是由一对头相关传输函数(HRTF)所描述, 它们定义为自由场情况下点声源到双耳的归一化传输函数(亥姆霍兹方程的格林函数)[2]:

$$H_\alpha(r_S, \Omega_S, f) = \frac{P_\alpha(r_S, \Omega_S, f)}{P_0(f)}, \tag{1}$$

其中, P_α表示声源在耳(例如鼓膜)所产生的声压, α = L和R分别表示左和右耳; P_0表示头部移开后声源在原头中心位置处产生的自由场声压. 一般情况下, HRTF与声源相对头中心距离r_S、方向Ω_S、频率f有关. 在大声源距离(与波长比较)的远场近似下, HRTF渐近地和距离无关. 并且, 由于不同个体的头部、耳廓等的生理形状与尺寸的差异 (对声波的散射不同), HRTF是和个体有关的. HRTF描述了倾听者生理结构对声波的散射和衍射或等效的声学滤波过程, 对有关听觉的物理分析非常重要.

听觉系统对双耳接收到的声压信息进行综合处理, 形成各种听觉事件与感知. 其中空间听觉事件包括对单声源的定位、多声源的听觉感知、对环境反射声的综合感知、声场景的综合感知等. 人类空间听觉的机理对于空间声的研究与发展有重要的作用. 但由于听觉系统对双耳声信息的处理涉及高层神经系统的生理信息处理, 目前其机理并不完全清楚, 因而在空间声研究中涉及听觉处理部分很多是采用心理声学的方法. 即将高层神经系统的听觉生理信息处理部分作为一个“黑匣”处理, 通过实验研究直接得到声场或双耳声压中的物理量与听觉感知(心理量)之间的映射关系.

声源定位包括方向定位和距离感知. 物理学家Rayleigh最早研究了听觉系统对单声源方向定位的物理与心理声学机理, 提出了经典的“双因素理论”. 也就是, 对中垂面以外的声源, 声波到双耳的传输存在双耳时间差(ITD); 同时由于头部对声波的散射作用, 存在双耳声级差(ILD). 与方向有关的ITD与ILD是声源方向定位的因素.

在简化的头部模型下, 将头部简化成半径为a的刚性球体, 双耳简化为球面上相对的两点. 根据亥姆霍兹方程对球散射问题的解, 可以求出不同方向远场声源(入射平面波)在左耳和右耳产生的声压, 并由此计算出ITD和ILD[2]. 图1(a)是取头部半径a=0.0875 m计算得到0.25, 0.5和1 kHz三个不同频率的简谐远场(平面波)声源所产生的ITD随水平面声源方位角θ_S的变化曲线. 其中方位角θ_S=−90°, 0°, 90°和180°分别表示正右、正前、正左和正后方向. ITD定义为右耳声压相对左耳声压的相延时. 由于对称性, 图中只给出了声源方向在左半水平面的结果. 可以看出, 虽然ITD和频率有关, 但随声源方向变化的规律是一致的. 在正前和正后方向ITD都是零; 随着声源偏离正前或正后, ITD的量值增加, 并在侧向达到最大.

图1(b)给出了计算得到的不同频率简谐远场(平面波)声源产生的ILD随水平面声源方位角θ_S的变化曲线. 其中ILD定义为左耳声压级与右耳声压级之差, 并采用ka作为归一化频率标度, k为波数. 同样由于对称性, 图中只给出了声源方向在左半水平面的结果. 在低频时ka很小, ILD也很小, 且随θ_S变化平缓. 随着频率的增加, 在ka大于1时ILD逐渐增加, 表现出与方向θ_S以及频率的复杂关系. 这是声波经过头部散射与衍射并在耳的位置叠加干涉的结果.

但由于头部的近似前后和上下对称性, ITD与ILD因素并不能唯一决定声源的空间方向. 经过一个多世纪的大量持续研究, 目前对方向定位的机理已比较清楚, 是多种因素的综合结果[1]. 其中频率在1.5 kHz以下的低频ITD (双耳相延时差)和3–4 kHz以上的高频

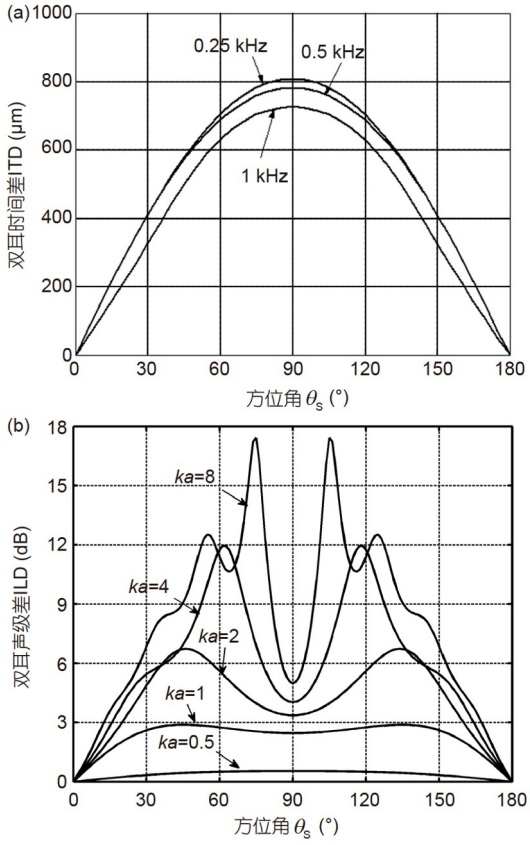

图 1 球形头部模型得到的水平面不同方向简谐声源产生的ITD与ILD[3]. (a) ITD; (b) ILD

Figure 1 ITD and ILD calculated from a spherical-head model for harmonic sound source at various horizontal directions [3]. (a) ITD; (b) ILD.

ILD是声源的侧向定位因素. 而头部、耳廓等对5–6 kHz以上的高频声波散射引起的耳接收到声压谱的改变(谱因素)和头部转动引起的双耳声压(特别是ITD)的变化是前后和垂直方向定位的因素. 以上这些与定位相关的物理量都可通过式(1)定义的HRTF进行严格分析, 上述刚球头部模型是一种简化的HRTF计算模型.

值得指出的是, 不同频段、不同的定位因素有一定的协同作用, 使得感知到的声源方向定位信息更加完整和稳定. 同时, 不同因素提供的方向定位信息是有一定冗余的, 当有部分信息缺失甚至发生冲突时, 听觉系统仍有可能按照主导信息或一致性好的信息进行正确的方向定位. 例如, 已有实验证明[4], 当声信号包含低频成分时, 低频ITD对定位起主导作用而可忽略冲突的ILD因素. 最近的研究结果表明, 当谱因素不存在甚至和动态因素冲突时, 动态因素仍然能一定程度地实现前后和垂直定位, 反之亦然[5]. 当然, 如果缺失或冲突的因素过多, 定位质量就会受到影响甚至出现完全不能定位的情况.

人类对声源的距离感知存在偏差. 实验结果表明, 感知距离通常可以用实际距离的幂函数来表示[6]. 距离感知是多个因素综合作用的结果[7]. 自由场情况下, 随距离变化的声音主观响度是一个相对的距离感知因素; 空气吸收引起声波的高频衰减(谱改变)是弱的距离感知因素; 头部对声波散射产生的近场HRTF效应, 从特别是中垂面外近场(1.0 m以内)声源产生的与距离有关的低频(3 kHz 以下)ILD, 提供了重要的绝对距离感知信息. 而在反射声环境中, 与声源距离有关的直达/混响声能比提供了重要的绝对距离感知信息. 但是, 在不同的声源方向、距离、频率, 不同因素对距离感知贡献的差别较大, 目前有关距离感知信息的冗余性、主导因素方面的研究与结论不如方向定位完善.

当同时存在多个不同空间位置的声源时, 会根据声源的空间位置、各声源信号的相关性、幅度与延时关系而出现不同的空间听觉事件. 对于相关声源信号, 如果各声源信号存在通路声级差或通路时间差, 就会出现合成定位的心理声学现象[1,3]. 对于水平面适当的方向、两个或更多的扬声器布置, 甚至垂直平面的适当扬声器布置, 也会出现类似的合成定位现象.

通路声级差产生合成定位的机理主要是因为它产生了部分的听觉定位因素, 特别是低频定位因素. 利用听觉定位因素的冗余性而产生虚拟源的感知. 其中在水平面, 适当的扬声器布置和通路声级差可以在倾听者双耳叠加声压中产生不同的低频ITD, 利用低频ITD对侧向定位的主导作用而产生不同方向的虚拟源[3]. 而在中垂面, 适当的扬声器布置和通路声级差可产生不同的动态ITD (随头部转动)的变化, 从而产生不同垂直方向的虚拟源[8]. 因而水平和垂直方向的合成定位机理是不同的. 另外, 对水平面左右扬声器布置和具有瞬态特性的宽带信号, 通路时间差(不超过1 ms)也会产生扬声器之间的合成虚拟源. 该现象的研究目前只是停留在心理声学实验现象的层面, 并不能从物理原理上给予解析. 但动物实验的一些迹象表明, 这可能和高层神经生理对声音信息的处理有关[9].

另一个重要的多声源合成空间听觉事件是优先效

应. 当两个或两个以上声源产生的具有瞬态特性声波到倾听者存在相对延时, 如果相对延时超过一定的低限和不超过一定的高限(低限在几个ms的量级, 上限在40 ms的量级, 取决于信号的特性), 听觉系统会感觉到声音来自于信号相对超前的声源, 而在空间定位上几乎感觉不到信号相对落后的声源单独存在. 也就是说两个声会被感知成融合的空间听觉事件. 当相对延时超过上限时, 听觉系统就会感觉到回声. 优先效应对人类听觉感知有重要意义. 在实际环境(房间)经常是有反射声的, 只要反射声相对直达声的延时在优先效应的范围内, 反射声并不影响听觉系统对声源的定位. 优先效应一直都是人类听觉研究的重要内容[10], 它不能用纯物理的机理进行解析, 它应该是发生在神经生理处理的层面.

除了优先效应, 环境反射声会产生其他特殊和重要的空间听觉感知. 在音乐厅中, 早期的侧向反射和后期的扩散反射分别会产生感知声源展宽和被声音包围的感觉, 即听觉声源宽度和包围感, 这对音乐厅的音质至关重要[11]. 实验表明, 听觉声源宽度与包围感和双耳声压的相关性密切相关, 低的相关性对应宽的听觉声源宽度与良好的包围感. 这是从物理量到听觉感知映射的典型例子. 另外, 当把去相关的信号馈给适当布置的多个扬声器重放时, 也会产生感知声源展宽和被声音包围的感觉, 这种心理声学实验现象可用于声重放中模拟音乐厅的听觉感知.

对复杂声场中空间听觉感知的研究(如多声源信息的分离、目标信息的注意力集中等)正在开展.

3 空间声的原理与进展

从原理上看, 空间声主要基于三个原理与方法[3], 包括物理声场的精确重构方法、心理声学与物理声场的近似重放方法与双耳声信号的精确重放方法. 也可以据此对空间声进行分类. 基于三类方法的空间声都有较长的历史, 但近年来基本理论与实验以及实际应用方面都有很大的发展.

3.1 物理声场重构方法

这类方法在一定的空间区域内实现物理声场的精确重构, 倾听者可以在重放声场中获得期望的声音空间信息, 从而产生相应的空间听觉事件. 这类方法在物理上是严格的, 在听觉感知效果上应该是完美的.

波场合成(Wave Field Synthesis, WFS)是声场重构的典型代表[12]. 其物理基础是经典的惠更斯原理, 数学上由Kirchhoff-Helmholtz积分表示. 即闭合边界表面S'内无源区域V的声场由边界上的声压及其梯度所决定:

$$P(\mathbf{r},f)=-\iint_{S'}\left[\frac{\partial P(\mathbf{r}',f)}{\partial n'}G(\mathbf{r},\mathbf{r}',f)-P(\mathbf{r}',f)\frac{\partial G(\mathbf{r},\mathbf{r}',f)}{\partial n'}\right]\mathrm{d}S',\quad \mathbf{r}\in V, \tag{2}$$

其中, $\mathbf{r}$为边界内任意场点或接收位置的矢量; $\mathbf{r}'$为边界表面任意点的位置矢量; f为频率. $G(\mathbf{r}, \mathbf{r}', f)$为三维亥姆霍兹方程的自由场格林函数, $\partial/\partial n'$为边界表面的内法向导数. 式(2)的积分是对整个边界表面进行.

由于$G(\mathbf{r}, \mathbf{r}', f)$在物理上表示位于$\mathbf{r}'$的单极(点)声源在场点$\mathbf{r}$产生的声压, $\partial G(\mathbf{r}, \mathbf{r}', f)/\partial n'$表示位于$\mathbf{r}'$的偶极声源在场点$\mathbf{r}$产生的声压. 因而式(2)表明, 表面$S'$内任意场点$\mathbf{r}$的声压可通过连续、均匀分布在边界表面的单极和偶极次级声源重构. 而单极和偶极次级声源的驱动信号分别正比于边界表面上声压的内法向导数和声压.

以上理想的声场重构虽然概念上完美, 但实际的波场合成需要在此基础上简化. 第一, 同时采用单极及偶极次级声源实现声场重构是复杂的, 特别是偶极次级声源在实现上有一定的困难. 而单极次级声源可以用通常的扬声器系统近似实现. 在一些特殊的边界条件下可只采用单极次级声源实现声场重构. 这时式(2)简化为

$$P(\mathbf{r},f)=-\iint_{S'}\frac{\partial P(\mathbf{r}',f)}{\partial n'}G(\mathbf{r},\mathbf{r}',f)\mathrm{d}S',\ \mathbf{r}\in V. \tag{3}$$

第二, 如果将三维空间的声场重构简化为水平面的声场重构, 则会出现次级声源辐射特性不匹配(用点声源代替直线声源)的问题. 第三, 实际中只能采用有限个分立布置的扬声器实现声场重构, 相当于对连续的次级声源分布进行了有限个空间采样, 这会导致重构声场的空间混叠和边缘效应. 因而, 对理想声场重构的简化会带来许多物理问题. 经过多年的发展, 目前对波场合成的理论已相对成熟[13], 对上述问题的物理结论已比较明确.

Ambisonics是声场重构的另一典型代表. 它最早

主要是作为一种心理声学与物理声场的低阶(一阶)近似重放方法引入的[14,15]. 后来的研究表明, Ambisonics的物理实质是将目标声场进行方向谐波(多极模态)展开, 然后在扬声器重放中逐级逼近目标声场[16–19]. 对水平面Q阶Ambisonics重放, 相当于按声源的方位角的傅里叶级数的前Q阶项逼近目标声场, 需要$(2Q+1)$个独立信号(方位角谐波)和$M \geq (2Q+1)$个扬声器重放. 如果目标重构声场是从水平方位角θ_S入射的平面波, 采用水平面圆周上均匀布置的M个扬声器重放时, 方位角为θ_i的第i个扬声器的信号振幅或增益可表示为方位角的傅里叶级数的前Q阶项的形式:

$$A_i = A_{\text{total}}\left[1 + 2\sum_{q=1}^{Q}[\cos q\theta_i \cos q\theta_S + \sin q\theta_i \sin q\theta_S]\right], \quad q = 1, 2, ..., Q. \tag{4}$$

对三维空间的L阶Ambisonics重放, 相当于按声源方向的球谐函数分解的前L阶项逼近目标声场, 需要$(L+1)^2$个独立信号(球谐函数分解项)和$M \geq (L+1)^2$个扬声器. 图2是作者所在实验室建立的空间Ambisonics实验系统扬声器布置的照片. 它包括28+1个扬声器, 分别布置在低仰角层、水平层和高仰角层, 另加一正上方的扬声器.

作为两类典型的声场重构方法, 波场合成和Ambisonics是亥姆霍兹方程在不同物理和近似条件下的结果, 或者理解为声场的不同表示. 波场合成源于声场的Kirchhoff-Helmholtz表示, 而Ambisonics源于声场的空间谐波表示. 最近也有研究提出空间小波作为基函数实现声场重构[20]. 波场合成与Ambisonics既相互关联, 又有一定的差异[21]. 在一定的条件下, 它们都可以重构不同方向的目标入射平面波场, 这一点非常重要. 因为根据空间傅里叶变换, 无源区域的任意声场可以分解为不同入射方向平面波的叠加. 可以重构不同方向的入射平面波场意味着可以重构任意的声场. 通过时间反演技术和聚焦源技术, 或者模态匹配技术, 有可能重构有限距离声源产生球面波或柱面波场[17,22,23].

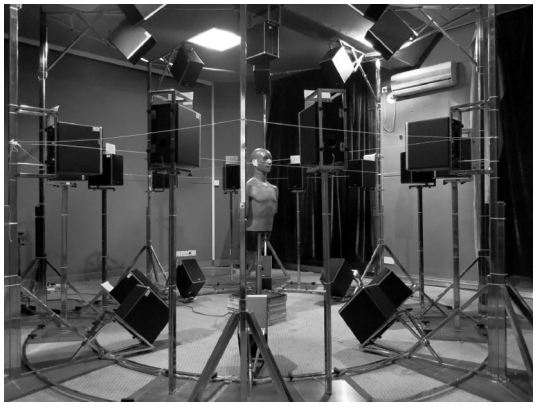

图 2 空间Ambisonics实验系统28+1个扬声器布置[3]

Figure 2 28+1 loudspeaker configuration for an experimental system of spatial Ambisonics [3].

波场合成通过控制边界上的声压或声压梯度, 理论上可以在边界内的整个空间实现声场重构. 图3(a)给出了仿真得到的布置在无限长直线上扬声器所重构水平面声场振幅. 图中的x-y坐标位于水平面, $+x$和$+y$方向分别指向正前和正左方向. 无限个扬声器(次级点声源)布置在x'=3.0 m的直线上, 相邻扬声器间隔为0.15 m. 目标声场为相对正前方向30°水平入射、频率f = 1.0 kHz的平面波. 图中给出的是$-3.0 \leq x \leq 3.0$ m, $-3.0 \leq y \leq 3.0$ m区域内的重构声场. 事实上, 无限长直线上的扬声布置是可以在$x \leq 3.0$ m的半水平空间内实现声场重构. 从图中可以清楚地看出重构声场的近似平面波阵面特性(当然, 重构声场会出现一定的幅度误差).

Ambisonics是通过模态控制而实现声场重构的, 理论上可以在一定的局部区域、一定的频率以下准确重构目标声场. Shannon-Nyquist空间采样理论给出, 对水平面Q阶Ambisonics重放, 实现准确重构声场圆区域半径r_0与上限频率或波数$k_{\max}$满足以下的关系[24]:

$$k_{\max} r_0 = Q. \tag{5}$$

随着Ambisonics的阶数增加, 准确重构目标声场的区域半径和上限频率增加, 重放效果得到改善, 但系统变得复杂. 对空间Ambisonics重放有类似的关系, 只要将式(5)的Q替换为空间Ambisonics的阶数L, r_0理解为准确重构声场的球形区域半径.

图3(b)给出了仿真得到的均匀布置在半径为5.0 m水平圆周上的18个扬声器作Q = 8阶水平Ambisonics重构声场振幅. 目标声场为相对正前方向30°水平入射、频率f = 1.0 kHz的平面波. 可以看出, 大约在半径$r \leq r_0$ = 0.45 m的圆形(图中虚线内)区域, 可以重构出目标的平面波阵面.

基于物理声场重构的空间声方法最大的问题是在物理上受到Shannon-Nyquist空间采样理论的限制. 企图在一定的听音区域内实现全可听声频带的声场重构

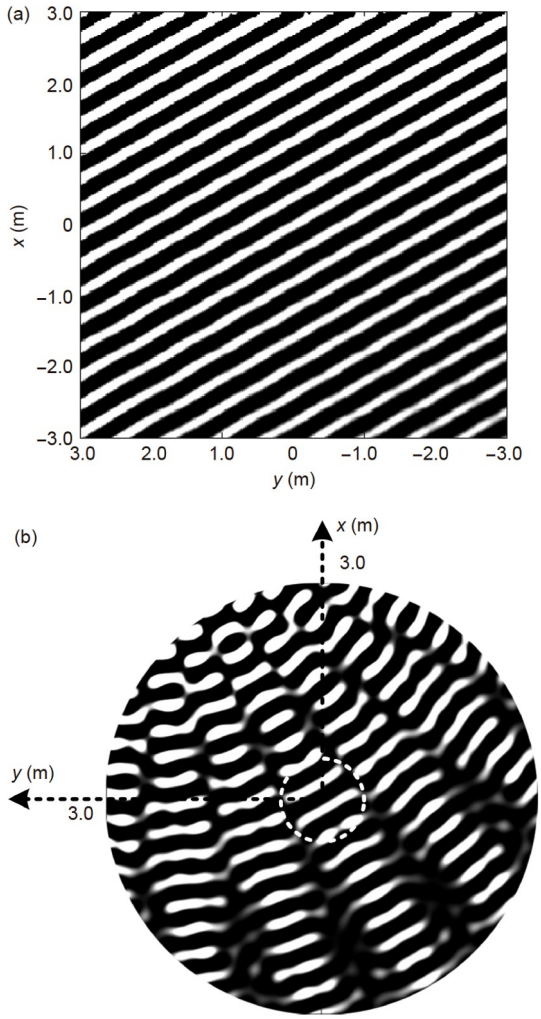

图 3 水平面波场合成和Q = 8阶Ambisonics的重构平面波声场振幅[3]. (a) 波场合成; (b) Ambisonics

Figure 3 Simulated amplitude of the reconstructed plane wave field of wave field synthesis and Q = 8 order horizontal Ambisonics [3]. (a) Wave field synthesis; (b) Ambisonics.

是非常复杂和困难的, 超过Shannon-Nyquist给出的频率上限, 重构声场会出现空间混叠错误. 对波场合成, 空间采样理论要求在边界上扬声器之间的距离不能超过半波长. 考虑到实际扬声器的尺度, 扬声间隔通常在0.1–0.3 m之间, 准确重构声场的上限频率大约在1.7–0.57 kHz的范围. 而对空间Ambisonics, 五阶重放至少需要36个均匀布置扬声器, 也只能在平均头部尺度(半径0.0875 m)范围内、3.1 kHz以下的频率实现准确的声场重构. 企图在同样尺度范围内实现上限频率为20 kHz的精确重构, 理论上需要采用高达33阶的Ambisonics重放, 需要超过1000个扬声器, 相应的系统非常复杂并难以实现. 按目前的技术水平, 基于物理声场重构远未达到20 Hz–20 kHz的全可听声频带范围. 一般情况下, 只能从低频0.1–0.2 kHz开始, 到高频2–3 kHz, 最多不超过4 kHz以下的频率范围实现声场重构. 因而基于物理声场重构的空间声最终也需要利用心理声学原理产生各种空间听觉事件, 例如利用低频定位因素的主导作用. 考虑到听觉系统对水平方向的分辨率高于垂直方向的分辨率, 有研究提出在Ambisonics中用较高阶的水平方向谐波和低阶的垂直方向谐波实现声场重构, 以简化重放系统[25–27]. 也有研究通过缩小局域声场重构的区域而换取更高的声场重构上限频率. 这些心理声学方法可以减少重放的定位错误, 但空间混叠引起的音色改变依然存在[28]. 正是由于其复杂, 过去很多基于物理声场重构的空间声是用于科学研究实验的, 但近年也逐渐用于影院和家用声重放.

总体上, 过去二十多年, 国际上在以波场合成、高阶Ambisonics为代表的声场重构方面做了大量的研究工作[3], 包括声场分析、空间采样与重构理论、信号处理与实现、性能的主客观评价等, 这些目前仍是国际研究的热点之一. 这类方法是与声学领域有源声场控制非常相似和紧密相连的[29]. 并且空间声场重构在很多方面是可以和其他物理场(如电磁场)的重构相类比的. 事实上, 因为声波和电磁波都遵从波动方程或亥姆霍兹方程, 量子力学中电子的运动遵循薛定谔方程, 这些类比是空间二阶偏微分方程在类似边界条件下有类似解、反映类似物理现象的结果. 因而在空间声场重构的理论研究借鉴了物理学其他领域的一些方法(如全息的方法), 所以精确的声场重构有时称为声学全息方法. 而Ambisonics的一些分析方法与空间转动变换是和量子力学中的角动量理论非常类似的[30].

3.2 心理声学与物理声场的近似重放方法

这类方法的重放声场在物理上与目标声场并不相同, 但在一定条件下有可能是目标声场的一种(粗略)近似. 其基本思路是利用特定的心理声学原理, 在一定程度上得到与目标声场类似的空间听觉事件或感知. 事实上, 这和利用光学的三色原理在电视产生彩色图像的思路类似. 最重要的是声场的简化可使系统简化, 使得这类方法与技术得到广泛的实际应用.

空间听觉心理声学原理在这类空间声的简化中起

着关键的作用. 其一, 在系统资源有限的条件下, 可以根据听觉上的相对重要性, 对不同方向的空间信息作适当的取舍或不同的近似, 从而实现系统的简化. 其二, 利用各种空间听觉信息的冗余性, 只要重放出部分的、起主导作用的声音空间信息, 就可以产生期望的空间听觉感知. 例如在水平面声重放中, 只要能产生期望的双耳时间差及其随头部转动的变化, 就可近似重放出期望的声音空间信息. 其三, 可以利用多声源(扬声器)合成空间听觉的原理产生各种期望的听觉感知. 例如利用一对适当的扬声器布置和信号, 可以合成介于真实扬声器布置之间的虚拟源, 并不需要在虚拟源方向布置真实的扬声器. 而利用多扬声器布置和去相关的信号馈给也可以模拟音乐厅包围感[31].

传统的两通路立体声是这类方法的典型代表, 早在20世纪30年代就开始了这方面的研究, 50年代末到60年代初就开始在家用声重放中广泛应用. 其基本思路是, 在系统资源有限的条件下, 优先考虑重放水平面前方一定角度范围的声音空间信息. 它采用前方对称的左右扬声器布置(相对倾听者的标准张角为60°), 通过改变左右扬声器的信号关系(相对振幅和延时)而产生两扬声器之间的空间信息. 图4是仿真得到的两通路立体声重放的叠加声压振幅分布图. 扬声器距离原点($x = 0$, $y = 0$)为2.5 m, 信号为0.5 kHz的正弦波; 左右扬声器的信号振幅相等. 在左右扬声器附近, 重放波阵面近似为单个扬声器辐射的球面波. 在左右对称中线点附近的区域(图中两虚线之间的区域), 叠加波阵面是和正前方入射球面波(远场可近似为平面波)的情况类似的; 但在左右偏离中线很小距离的区域, 重放的波阵面明显偏离了前方入射平面波的情况. 所以, 两通路立体声只能在很窄的区域和一定频率以下重构目标声场, 它主要是利用心理声学原理产生目标的空间听觉事件.

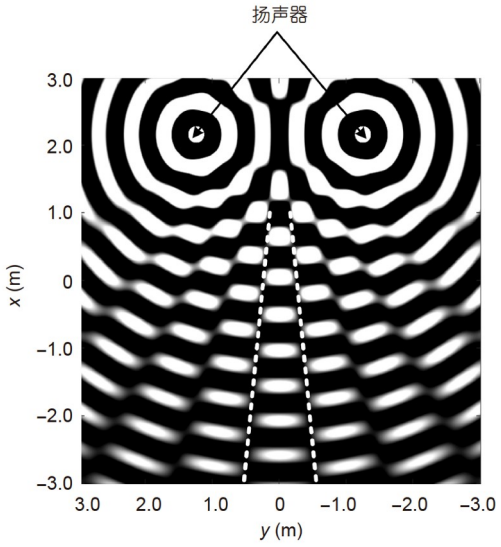

图 4 两通路立体声重放的叠加声压振幅分布(0.5 kHz正弦信号, 左右扬声器信号振幅相等)[3]

Figure 4 Wavefront amplitude of the superimposed sound pressures in stereophonic reproduction of 0.5 kHz-sinusoidal signal with identical amplitudes [3].

一般情况下, 利用合成定位的心理声学原理, 计算出两扬声器信号所产生的低频(0.7 kHz以下)双耳叠加声压, 计算出双耳相延时差并与真实声源的情况比较, 可以得到两通路立体声重放的虚拟源方位角θ_I满足以下的正弦定理[3,32]:

$$\sin\theta_I = \frac{E_L - E_R}{E_L + E_R}\sin\theta_0 = \frac{E_L/E_R - 1}{E_L/E_R + 1}\sin\theta_0, \tag{6}$$

其中方位角90°, 0°和−90°分别表示水平面正左、正前和正右方向. θ_0为左右扬声器之间的半张角. E_L和E_R分别为左右扬声器的信号振幅. 因而改变左右扬声器信号之间的振幅比, 就可以产生左右扬声器之间不同方向的虚拟源.

两通路立体声只能重放水平面前方附近一定角度范围内的声音空间信息. 20世纪70年代初国际上曾大力发展4通环绕声系统与技术[33]. 它采用左前、右前、左后、右后4个方形布置的扬声器重放, 企图产生整个水平面的声音空间信息, 但最后没有成功[34]. 失败的主要原因是在系统(可用通路数)资源有限的条件下, 仍将水平面不同方向的声音空间信息等同处理, 结果反而导致重要的空间信息丢失. 但这时期的失败为后来的发展提供了很好的经验. 80年代末到90年代初, 国际上开始发展用于影院和家用伴随图像重放的多通路水平面环绕声技术. 这类技术对不同方向声音的空间信息进行不同精度处理. 考虑到伴随图像重放中, 倾听者的注意力集中在前方, 因而采用非均匀的扬声布置, 以相对精确地重放前方的声音空间信息, 粗略重放后方的声音空间信息. 5.1通路环绕声和7.1通路环绕声是这类技术的典型. 其中5.1通路环绕声已被国际电信联盟推荐为通用的、伴随不伴随图像的多通路声标准并得到广泛的应用[35]. 5.1通路环绕声包括

前方左L、中C、右R, 侧后方左环绕LS、右环绕RS共5个全频带的通路, 加上一个可选择的低频效果通路. 图5是国际电信联盟推荐的家用5.1通路环绕声的五个全频带扬声器布置.

三维空间声增加了垂直方向的声音空间信息, 理应得到更好的感知效果. 但多通路三维空间声需要更多的通路, 扬声器布置更加多样化, 因而也更复杂. 已提出了9.1通路、11.1通路、13.1通路、22.2通路等甚至更多通路的扬声器布置和系统[36–38]. 考虑到伴随图像重放和听觉系统具有较高的前方声音空间信息分辨率, 这些系统一般采用了非均匀(前方范围相对密)的扬声器布置, 这相当于对空间方向进行了非均匀的采样. 按照Shannon-Nyquist空间采样理论, 这些扬声器布置不足以准确地重构目标声场, 需要采用心理声学原理合成各种空间听觉事件. 其中水平方向合成定位的机理与技术已比较成熟(低频双耳时间差主导); 本文作者及课题组在垂直方向的合成定位机理方面进行了深入研究, 证明低频动态因素在各种多通路空间声中的作用[8,39].

总体上, 基于心理声学与物理声场近似的空间声技术目前正在向多通路三维空间声发展, 且重放通路也在增加. 这一方面改善了重放声场的精确性, 减少了声音空间信息的损失, 使重放效果得到改善. 但另一方面系统却越来越复杂, 这始终是一对矛盾. 实际的系统是舍弃了部分声音空间信息, 采用心理声学原理简化后的一种折中的方案, 它所重放的声场最多是理想声场的一种粗略近似. 因而应根据实际应用的要求和基于心理声学原理, 考虑可用的资源(如媒体的信号传输、记录能力)、听音区域的大小等, 合理选择独立通路数、扬声器布置与信号馈给, 尽可能保留重要的声音空间信息, 以产生期望的主观空间听觉感知. 这是立体声和多通路声发展的主要思路, 也一直是多通路声(包括水平和空间环绕声)研究主要的核心问题. 过去研究在这方面已做了大量的工作, 但随着技术的发展, 不断有相关的基础(特别是心理声学基础)与实际的技术问题需要解决.

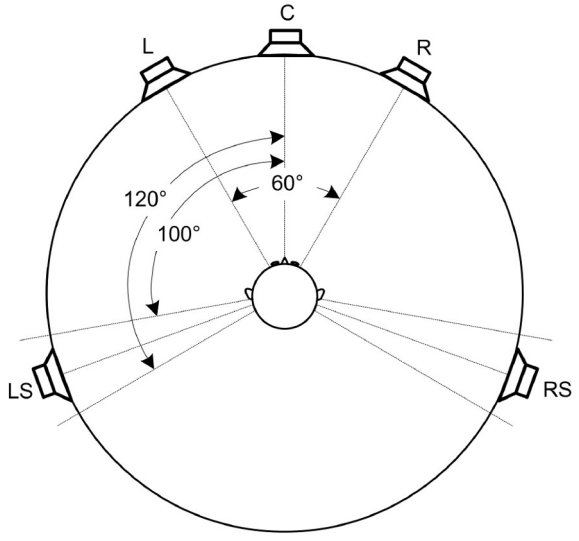

图 5 5.1通路环绕声的五个全频带扬声器布置[35]

Figure 5 Configuration of five loudspeakers with full audible bandwidth in 5.1-channel surround sound [35].

3.3 双耳声信号的精确重放

在任意复杂的声场, 人类都是通过双耳接收声音信息的. 这类方法通过精确重构双耳声信号而重放声音的空间信息, 从而产生相应的空间听觉事件. 这类方法的硬件结构比较简单, 且处理恰当的情况下可得到真实自然的空间听觉事件.

最直接的双耳重放方法是先采用一对放置在人工头听觉仿真模型双耳处的传声器在现场捡拾得到双耳声信号, 经放大后再用耳机重放. 双耳声信号也可以通过模拟得到. 也就是用信号处理的方法, 模拟出声源的辐射、声源到双耳的物理传输过程, 从而得到双耳声信号并用耳机重放. 这是虚拟听觉重放的基本原理. 最简单的情况, 当声源和倾听者头部都固定不动时, 自由场声源到双耳的传输可看成是一个线性时不变的过程, 并用式(1)定义的HRTF描述. 把单通路输入信号 $E_0(f)$用一对左右耳的HRTF滤波即可产生双耳声信号 $E(r_S, \Omega_S, f)$. 通过耳机重放, 即可产生相应位置的虚拟源感知[2,40]:

$$E_\alpha(r_S,\Omega_S,f) = H_\alpha(r_S,\Omega_S,f)E_0(f). \tag{7}$$

在现实中, 倾听者头部的运动会改变声源到双耳的传输, 从而改变双耳声压. 如果进一步采用头踪迹跟踪器实时地检测倾听者头部的位置, 并动态调整模拟声源到双耳的信号处理, 则可以在双耳重放中产生动态的空间信息. 这就是动态虚拟听觉重放或虚拟听觉环境的基本原理[41]. 另外, 通过适当的串声消除处理, 也可以将双耳信号转换为扬声器重放[42].

虚拟听觉重放近年发展很快. HRTF是实现虚拟听

觉重放的基础数据, 信号处理是实现虚拟听觉重放的重要工具, HRTF的物理与听觉特性分析与近似表示是简化信号处理的途径. 过去三十年, 国内外在HRTF与虚拟听觉重放方面做了大量的研究工作. 其中远场HRTF的获取(包括测量与数值计算)技术已比较成熟, 国际上已有多个课题组对人工头和真人受试者的远场HRTF(声源距离大于1.0 m)进行了测量, 建立了相应的数据库[43–45]. HRTF的基本特性、各种虚拟听觉重放信号处理方法等方面的研究也有基本的结论, 部分技术已进入实用. 当然, 虚拟听觉重放技术也存在一定的缺陷, 如重放虚拟源方向畸变、效果因人而异等. 有关HRTF与虚拟听觉重放的基本原理和2009年以前的研究进展参考文献[46], 更详细可参考文献[2]. 而近十年国内外在这方面的研究也有更深入的发展.

需要强调的是, 上述三类空间声方法在物理上是紧密关联的. 在声场空间采样与重构的框架上, 三类方法甚至是统一的[3,47]. 声场(声压)是空间、时间的连续函数. 第一类方法的物理本质是声场的空间采样与精确重构, 即利用空间分立扬声器(离散空间采样)布置实现目标声场的重构. 第二类方法是在不能实现声场精确重构的条件下, 实现声场的粗略近似重构, 同时利用心理声学原理实现声场空间信息的重放. 第三类方法是实现双耳(空间两点)的声压重构, 根据声学的互易原理, 这也是一种声场的精确采样与重构方法. 正因为如此, 不同类型空间声方法的许多理论基础是共通的, 可以相互借鉴. 而在实际中, 不同类型的原理与方法也可以混合应用, 达到相互补充的目的.

4 空间声实现的相关技术问题

传统上, 完整的空间声系统包括信号的产生、传输(与记录)以及重放三个环节, 对空间声的实现非常重要, 并涉及许多物理与技术的问题.

空间声信号可以通过模拟或捡拾的方法产生. 模拟是根据不同类型空间声的原理以及期望的重放声场和听觉感知, 在计算机或专门的设备中用信号处理的方法产生所需要的空间声信息及信号, 各种基于物理和心理声学原理的方法都有可能用于信号的产生, 如分立-对信号馈给法、基于振幅矢量的信号馈给(VBAP)方法和Ambisonics的方法[48]. 国际上在产生反射环境主观听觉感知的技术方面(如混响信号和去相关方法)有不少的研究[49,50]. 这种情况下, 声场空间信息是“人工造的”. 这类方法较多地用在影视节目和流行音乐节目的制作.

空间声信号可以通过适当的传声器技术在现场捡拾得到. 对两通路立体声和多通路声, 可以利用传声器的指向特性和声波传输到不同位置传声器的距离差异, 产生具有和声源位置有关的振幅(声级)差和时间差的空间声信号, 从而得到声音的空间信息. 利用适当的宽间隔的传声器布置, 可以捡拾得到(音乐厅内)低相关扩散混响声信号. 已发展了多种不同两通路立体声和多通路水平面环绕声的传声器捡拾技术[51], 这些技术比较成熟, 较多地用在大型管弦乐的录声. 目前正在发展多通路三维空间环绕声的传声器捡拾技术[52], 这可看成是多通路水平面环绕声传声器技术的推广. 但由于垂直方向的听觉感知机理与水平方向的差异, 其中一些原理与技术性的问题还有待解决. 传声器阵列是近年发展的一类捡拾技术[53,54]. 它通过特定规则布置(如球表面布置)的多个传声器对声场进行物理采样, 得到准确代表声场物理信息的捡拾信号. 把这些捡拾信号转换为基于物理声场重构的重放信号(如Ambisonics重放信号), 在一定物理条件下可以准确重构出物理声场. 信号的转换的实质就是不同正交基函数表示的声场信号之间的转换, 其数学方法与量子力学的表象变换非常相似. 传声器阵列捡拾信号也可以转换为多通路声和双耳信号重放. 必须指出的是, 实际的空间声信号模拟与捡拾已不单纯是一个科学与技术的问题, 也涉及艺术的问题.

信号传输或记录是空间声系统链的一个重要的环节. 前者是用物理手段将信号传递到不同的终端; 后者是采用物理手段将空间声信号存储在永久载体上. 它们的作用是将节目信号分配给倾听者. 早期是采用模拟技术进行信号传输和记录的, 包括模拟广播技术、模拟唱片、模拟磁性录声等. 模拟技术的信息传输和记录容量有限, 这曾经是制约多通路空间声发展的一个瓶颈. 为了用有限的信号传输和记录容量传递和重放声场的空间信息, 在20世纪70–80年代曾发展了矩阵环绕声的方法[33,55], 也就是将多通路的环绕声信号矩阵编码(线性组合)成少量通路(如两通路)信号传输, 重放时再通过矩阵解码(线性组合)或自适应非线性解码得到多通路声重放. Dolby实验室发展的Dolby Stereo和Pro-Logic就是这类技术的典型. 但是信号编

码、解码过程不可避免造成空间信息的损失. 当然, 矩阵环绕声的方法近年也有新的发展[56].

近三十年, 数字信号处理技术快速发展, 为多通路空间声的实际应用提供了重要的条件. 目前空间声的信号主要是通过数字媒体记录和传输. 为适应各种媒体的传输带宽, 发展了多种不同的基于心理声学的数字压缩编码以及各种无损编码技术[57]. 同时, 信号的传输与记录媒体也发生了很大的变化. 现代的光学记录媒体(光盘)、固体记录媒体、计算机与互联网系统、5G通信技术等都可用于空间声信号的传输与记录. 顺便指出, 记录声频信号的技术需求曾很大地促进了信息记录技术及其相关物理研究的发展, 包括磁记录技术和光学记录技术. 由于现代传输与记录手段主要涉及信号与信息处理问题, 在此不再详述.

重放是空间声系统的最后一个环节. 实际重放多数是在室内环境进行. 倾听者双耳所接收到的是扬声器的直达声和重放环境的反射声. 重放房间的反射声(注意不是目标声场如原音乐厅的反射声)会破坏原目标声场的空间信息的重放. 为了简单起见, 很多情况下空间声的理论分析忽略了重放房间反射声的影响. 虽然在重放房间反射到达听音位置的时间与强度满足优先效应的条件下, 重放房间反射对虚拟源定位的影响较小, 但对其他听觉感知如音色、主观包围感等有重要的影响. 因而实际中需要采取适当的声学措施(包括吸声处理、适当设计扬声器布置等)减少这些影响. 过去对重放房间的声学问题进行了大量的研究工作, 一些国际组织也制定了一些声学设计标准. 也有研究提出目标声场重构与有源反射声消除相结合的方法, 如利用Ambisonics同时进行目标声场重构和听音室房间反射消除[58]. 但由于通常(家用)小尺度房间内的声学与听觉问题比较复杂, 且近年三维空间环绕声的发展, 这方面还有许多工作要做.

上面所述的是传统的基于通路(Channel-based)的空间声系统结构, 即预先确定所采用的重放原理、重放配置(如扬声器数目与布置等), 在节目制作阶段完成信号模拟, 然后将制作好的节目分配给用户. 如果重放的原理和配置与节目制作设定的不同, 则需要对节目信号进行转换处理. 过去已发展了各种信号转换的技术, 包括向下、向上混合(将较多通路信号转换为较少通路信号重放或相反), 但转换过程不可避免会造成空间信息的损失[59]. 也有采用双耳合成的方法将扬声器重放信号转换为耳机重放或相反[2], 因而基于通路的结构缺乏灵活性. 随着空间声应用的多样化, 对重放原理、配置有不同的要求. 特别是对于多通路三维空间声系统, 其重放配置的可选择性的增加, 灵活性的问题更显突出.

为了适应不同应用的要求, 近年在空间声的系统结构方面发生了较大的变化, 发展了基于目标(Object-based)的空间声系统结构. 它将具有相同空间特性的声音成分组合成一系列声音目标传输, 同时传输包括用于描述各声音目标特性(如声音空间位置)的元数据. 重放时选择适当的重放原理, 并根据实际的配置, 选择适当的信号模拟方法, 按元数据提供的信息, 将信号馈给扬声器或耳机重放. 基于目标的结构将信号模拟从节目制作阶段(信号传输之前)后移到重放阶段(信号传输之后), 可适用于基于不同原理的、不同配置的重放, 为各种空间声技术的灵活、集成、综合应用提供了基础.

国外的研究已发展了一些基于目标的三维空间声技术与系统, 它们的原理和结构是相似的, 例如Dolby Atmos, AuroMax等. 国际标准化组织(ISO)与运动图像专家委员会(MPEG)已经制定新一代的基于目标的三维空间声标准MPEG-H Spatial Audio[60]. MPEG-H采用的基于目标与通路混合的模式, 支持各种不同的多通路声、Ambisonics和虚拟听觉重放等. MPEG-H标准对重放原理、配置、信号绘制方法等都是灵活和开放的. 目前各种基于目标的空间声技术与标准支持的最大重放通路(扬声器)数目在几十到一百的量级.

空间声的性能可通过主观和客观的方法进行评价. 客观的方法是通过仿真或测量重放声场或双耳声压及相关因素的物理误差, 具有相对简单、重复性好的特点. 但由于空间声的设计中利用了大量的心理原理, 客观的物理测量并不能完全反映实际的听觉效果. 目前对空间声主要是采用主观评价的方法, 通过心理声学实验评价实际的听觉效果. 主观评价方法复杂耗时, 需要仔细的实验设计、足够的受试者和合适的统计方法才能得到有意义的结果. 这是目前的认识水平下一种不得已的方法. 国际电信联盟已制定了一些主观评价的标准[61,62]. 更进一步的应该是基于双耳听觉模型的客观评价方法, 也就是将人类听觉信息处理的模型引入客观评价, 这方面已有一定的进展[63], 但还有许多问题需要研究.

5 国内有关空间声的研究

在国际上相关研究与实际应用的推动下，近年在国内也引起了有关研究的热潮. 但国内其他单位在空间声重放声场和心理声学方面的系统研究工作并不多. 主要的代表性工作有中国科学院声学研究所与北京大学合作，通过实验测量建立了KEMAR人工头的近场HRTF数据库[64]；中国科学院声学研究所李军锋研究员对虚拟听觉重放的研究[65]；以及中国科学院声学研究所杨军研究员在多区域声场控制方面的工作[66].

华南理工大学的课题组在谢兴甫的带领下从1958年就开始了空间声的基础研究工作, 20世纪70–80年代曾对两通路立体声和多通路声的声场和听觉分析进行了大量的前瞻性基础研究工作[15,33,67]，为80年代两通路立体声技术在国内的普及应用提供了重要的基础. 近二十年，该课题组(与北京邮电大学管善群合作)在空间声重放及其物理、听觉原理方面开展了系列性的基础研究工作. 特别是在声场空间采样、重构与听觉近似的理论框架上，发展了不同物理和听觉原理空间声的统一理论[3]. 国际上与华南理工大学的相关工作已系统总结成两本专著出版[2,3]. 近年华南理工大学对基于目标的扬声器三维空间声重放进行了深入的研究，特别是从理论和实验上证实了垂直方向定位信息的冗余性，证实了头部转动引起的动态因素对听觉垂直定位的作用，为三维空间声重放提供了理论和实验基础[5]. 在此基础上设计了重放垂直方向虚拟源的多通路声信号馈给法[30,68]. 同时也采用双耳听觉模型对Ambisonics音色进行了分析，提出捡拾和重放的设计方法[69].

另一方面, 华南理工大学对HRTF及虚拟听觉重放进行了大量的基础与应用研究工作. 早在2005年, 课题组设计了第一代的HRTF测量系统, 建立了52个中国人受试者样本的远场HRTF数据库[70]. 2015年华南理工大学余光正等人[71]设计了第二代的HRTF快速测量系统, 通过实验测量, 建立了目前国际上唯一的真人受试者的近场HRTF数据库(56名受试者，7个声源距离，每个距离685方向). 研究建立了测量HRTF的空间采样理论[72]，得到重构方向连续HRTF所需要的测量角分辨以及从少量测量方向近似恢复高方向分辨率的HRTF方法[73]. 而对HRTF的物理、听觉特性系统分析的研究中，得到了HRTF听觉高频谱分辨率，证明对5–6 kHz HRTF幅度谱进行2.0–3.5等效矩形听觉滤波器带宽的平滑不会带来可听的效果，为简化虚拟听觉重放信号处理提供了基础[74]. 同时，通过对中国人受试者样本的远场HRTF数据库进行分析，证明不同个体HRTF具有一定相似性特性，大多数受试者的HRTF可分成7类并用相应的类中心代表[75]. 对近场HRTF所包含的距离定位因素及与其他因素的协同作用进行了深入的研究.

另外，华南理工大学课题组建立了用于听觉科学研究的动态虚拟听觉环境的科学实验平台，为开展相关的基础研究提供了基础[76]. 该系统的实验证实动态虚拟听觉重放引入了头部转动引起的动态因素对虚拟听觉重放中产生头外定位感知的重要性. 华南理工大学在虚拟听觉重放(虚拟环绕声)方面也取得了具有重要自主知识产权的技术系列[77–80]，也就是用少量扬声器或耳机重放多通路声的技术，并且部分技术已实现产业化.

6 空间声的应用

文化娱乐是空间声的传统应用领域，如影院、厅堂和家用声重放(包括数字电视、家庭影院等)、汽车内声重放等. 特别是近十多年数字影院与演观厅堂的发展，大大推进了空间声技术的应用[81]. 而新传播媒体的出现，包括互联网与流媒体以及各种手持播放设备(如智能手机，iPad等)的普及，为空间声技术在消费电子领域提供了广阔的应用空间.

虚拟现实与人工智能是近年快速发展的高技术领域, 声音作为一种重要的信息载体, 在这方面有重要的作用. 空间声技术已开始应用于各种场景的智能模拟、各种虚拟训练、线上教育系统等. 而传统的语言通信与信息系统是单通路的，为了提高语言清晰度，包括空间声的语言通信系统已成为新的发展方向，并应用到各种民用与军用通信系统、航空与航天通信系统、远程会议系统等.

在有关听觉科学的基础研究中，经常需要人工控制声场及双耳接收到的声学信息. 空间声技术已成为人类听觉心理和生理的科学实验的一个重要实验工具, 可以方便地控制各种实验条件. 近年许多听觉实验是在包括空间声的虚拟现实平台上进行的. 而在环境

声学领域, 可以用空间声的方法模拟出期望的物理声场或双耳声信号, 从而进行相关的物理和听觉感知的科学研究, 包括室内音质的辅助设计(可听化)[82]、环境噪声的评价, 这可以看成是包括空间声的虚拟现实技术在科学研究和工程领域的一个特殊应用. 在医学方面, 空间声技术已用于包括听觉诊断、人工听力的恢复及其评估等[83].

以上仅是空间声众多应用的一部分, 但足以说明其广阔的应用前景.

7 总结、问题与趋势

空间声已有较长的研究历史, 已发展了三类不同物理与听觉原理的空间声技术, 但空间声目前仍是声学、听觉、信号与信息处理等跨学科领域的前沿课题. 应用的发展也对空间声提出了更高的科学与技术要求. 空间声的进一步发展将涉及声场的物理机理、人类空间听觉与感知机理、声信号与信息处理方法、各种应用等一系列科学与技术问题, 它们将成为今后的研究与发展方向. 当然, 这些问题并非是完全独立的, 它们是相互关联和交叉的.

(1) 声场捡拾与重构的物理问题

虽然有关声场捡拾与重构的理论已相对成熟, 声学和物理其他领域所发展的一些数学方法也可用于声场重构. 但受到Shannon-Nyquist空间采样理论和技术条件的限制, 目前最多在一定尺度范围内、在0.1–4 kHz的频率范围内实现声场准确的重构. 这已经涉及40倍的频率上、下限比, 从物理的角度, 已经相当不容易. 如果企图在大的区域、更宽的频带实现声场精确捡拾与重构会带来很多特殊的物理问题, 需要不同的声场物理和信号处理方法, 如声场的稀疏采样与重构方法, 甚至需要突破现有电声换能器的物理原理与结构. 这方面将是一个综合的、有相当难度的物理难题.

(2) 空间听觉的机理问题

人类空间听觉的机理是实现空间声的基础. 空间声的发展有赖于空间听觉机理的研究发展, 这对基于三种不同原理和方法的空间声重放都非常重要. 空间听觉的机理对空间声性能的评价也是不可缺少的.

听觉定位机理是其中的一个重要部分, 其中水平方向或侧向定位机理相对成熟. 垂直定位的心理声学机理, 特别是动态因素对垂直定位的作用, 需要进一步深化研究. 而距离感知是多个因素综合作用的结果, 各因素起作用的条件及其相互作用需要深入研究. 而各种听觉定位信息是具有一定的冗余性, 这些冗余性是简化空间声、使其能够实现或降低实现代价的基础. 听觉定位信息的冗余性, 特别是不同条件下距离感信息的冗余性需要深入研究. 复杂声学环境的听觉感知的研究, 特别是反射声环境的听觉感与空间声的关系, 环境反射声信息的冗余性等, 对实际空间声的发展也有重要的作用.

目前空间听觉机理的研究主要集中在心理声学的层面. 空间听觉的神经生理机制、模型的基础研究正在开始, 这方面的研究涉及听觉的生物物理问题, 将对空间声的认识产生质的飞跃.

(3) 空间声实现技术及信号处理发展

虽然三类空间声技术已有一定基础, 但应用技术的发展会带来大量的物理、听觉及信号处理的综合问题, 包括声场捡拾与重放、各种期望听觉感知的模拟、不同应用的性能与复杂性合适的折衷问题、基于物理和听觉原理的信号处理的简化及实现问题等. 特别是目前产生不同方向虚拟源的技术相对成熟, 但产生不同距离听觉感知的技术有待发展, 产生其他空间听觉感知和重放音色的优化的技术需要深入研究.

信号处理是现实空间声的重要技术手段. 虽然各种时、频域的数字信号处理理论与技术已比较成熟, 但空间声涉及的不仅是时频域的信号处理, 更是涉及空间域的多通路(多维)信号处理, 并且与声场的物理、感知特性密切相关. 只有结合声场物理与听觉机理的信号处理才是合理的. 这些特征给空间声的信号处理带来许多新的问题.

(4) 空间声的评价与听觉感知模拟

客观测量评价只能部分反映出空间声重放的感知性能. 主观实验评价虽然能反映空间声的主观性能, 但容易产生偏差. 基于双耳听觉模型的客观评价是今后的一个重要发展方向, 但其中的核心部分——听觉模型的建立需要大量的基础研究工作.

(5) 空间声与现代媒体和通信技术问题

新一代的空间声技术应该是在现代媒体与通信技术的环境下实现的, 包括计算机、互联网、5G通信等. 空间声与这些现代媒体的结合会带来许多通信与信息处理方面的新问题. 这方面的研究与发展将大大推动空间声应用技术的发展.

(6) 实际应用问题

空间声涉及多领域、多层次的实际应用. 各种不同的实际应用给空间声带来新的技术问题. 一方面, 对某一特殊的应用, 可能对空间声技术及其实现方法提出特殊的物理与听觉要求. 而在每一个应用领域, 还需要和该领域的其他技术相互配合. 例如, 在虚拟现实应用, 就需要和视觉场景的生成技术密切配合. 这些问题的研究对应用技术的发展有重要的作用.

总体上, 空间声是一个跨学科的研究领域, 同时有广阔的应用前景. 其发展将带动相关的声学、人类听觉的基础研究与应用技术的发展. 空间声发展涉及大量底层的基础研究问题, 这方面的基础研究与应用基础研究, 使其将在相关核心技术与自主知识产权方面有大的突破.

致谢 感谢中国声学学会特别是张春华理事长的支持, 感谢华南理工大学课题组同事、学生的合作.

参考文献

1 Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization (Revised edition). Cambridge: MIT Press, 1997
2 Xie B S. Head-related Transfer Function and Virtual Auditory Display. 2nd ed. Boca Raton: J Ross Publishing, 2013
3 Xie B S. The Principle of Spatial Sound (in Chinese). Beijing: Science Press, 2019 [谢菠荪. 空间声原理. 北京: 科学出版社, 2019]
4 Wightman F L, Kistler D J. The dominant role of low-frequency interaural time differences in sound localization. J Acoust Soc Am, 1992, 91: 1648–1661
5 Jiang J L, Xie B S, Mai H M, et al. The role of dynamic cue in auditory vertical localisation. Appl Acoust, 2019, 146: 398–408
6 Zahorik P, Brungart D S, Bronkhorst A W. Auditory distance perception in humans: A summary of past and present research. Acta Acust Unit Acust, 2005, 91: 409–420
7 Kolarik A J, Moore B C J, Zahorik P, et al. Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. Atten Percept Psychophys, 2016, 78: 373–395
8 Rao D, Xie B S. Head rotation and sound image localization in the median plane. Chin Sci Bull, 2005, 50: 412–416
9 Yin T C. Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. J Neurosci, 1994, 14: 5170–5186
10 Litovsky R Y, Colburn H S, Yost W A, et al. The precedence effect. J Acoust Soc Am, 1999, 106: 1633–1654
11 Ando Y. Auditory and Visual Sensations. New York: Springer-Verlag, 2009
12 Berkhout A J, de Vries D, Vogel P. Acoustic control by wave field synthesis. J Acoust Soc Am, 1993, 93: 2764–2778
13 Spors S, Rabenstain R, Ahrens J. The theory of wave field synthesis revisited. In: Proceedings of 124th Audio Engineering Society Convention. Amsterdam, 2008. Paper 7358
14 Gerzon M A. Ambisonics in multichannel broadcasting and video. J Audio Eng Soc, 1985, 33: 859–871
15 Xie X F. The 4-3-N matrix multi-channel sound system. Chin J Acoust, 1982, 1: 201–218
16 Xie B S, Xie X F. Analyse and sound image localization experiment study on multi-channel planar surround sound system. Chin J Acoust, 1996, 15: 52–64
17 Daniel J, Moreau S. Further study of sound field coding with higher order Ambisonics. In: Proceedings of 116th Audio Engineering Society Convention. Berlin, 2004. Paper 6017
18 Poletti M A. Three-dimensional surround sound systems based on spherical harmonics. J Audio Eng Soc, 2005, 53: 1004–1025
19 Poletti M A. A unified theory of horizontal holographic sound systems. J Audio Eng Soc, 2000, 48: 1155–1182
20 Scaini D, Arteaga D. Wavelet-based spatial audio format. J Audio Eng Soc, 2020, 68: 613–627
21 Daniel J, Nicol R, Moreau S. Further investigations of high-order Ambisonics and wavefield synthesis for holophonic sound imaging. In: Proceedings of 114th Audio Engineering Society Convention. Amsterdam, 2003. Paper 5788
22 Spors S, Wierstorf H, Geier M, et al. Physical and perceptual properties of focused virtual sources in wave field synthesis. In: Proceedings of 127th Audio Engineering Society Convention. New York, 2009. Paper 7914

23 Choi J W, Kim Y H. Integral approach for reproduction of virtual sound source surrounded by loudspeaker array. IEEE Trans Audio Speech Lang Process, 2012, 20: 1976–1989

24 Ward D B, Abhayapala T D. Reproduction of a plane-wave sound field using an array of loudspeakers. IEEE Trans Speech Audio Process, 2001, 9: 697–707

25 Favrot S, Marschall M, Kasbach J, et al. Mixed-order ambisonics reecording and playback for improving horizontal directionality. In: Proceedings of 131st Audio Engineering Society Convention. New York, 2011. Paper 8528

26 Grandjean P, Berry A, Gauthier P A. Sound field reproduction by combination of circular and spherical higher-order ambisonics: Part I—A new 2.5-D driving function for circular arrays. J Audio Eng Soc, 2021, 69: 152–165

27 Grandjean P, Berry A, Gauthier P A. Sound field reproduction by combination of circular and spherical higher-order ambisonics: Part II—Hybrid system. J Audio Eng Soc, 2021, 69: 166–181

28 Winter F, Wierstorf H, Hold C, et al. Colouration in local wave field synthesis. IEEE ACM Trans Audio Speech Lang Process, 2018, 26: 1913–1924

29 Hamdan E C, Fazi F M. A modal analysis of multichannel crosstalk cancellation systems and their relationship to amplitude panning. J Sound Vib, 2021, 490: 115743

30 Magariyachi T, Mitsufuji Y. Analytic error control methods for efficient rotation in dynamic binaural rendering of Ambisonics. J Acoust Soc Am, 2020, 147: 218–230

31 Damaske P, Ando Y. Interaural crosscorrelation for multichannel loudspeaker reproduction. Acta Acust Unit Acust, 1972, 27: 232–238

32 Leakey D M. Some measurements on the effects of interchannel intensity and time differences in two channel sound systems. J Acoust Soc Am, 1959, 31: 977–986

33 Xie X F. The Principle of Stereo (in Chinese). Beijing: Science Press, 1981 [谢兴甫. 立体声原理. 北京: 科学出版社, 1981]

34 Thiele G, Plenge G. Localization of lateral phantom sources. J Audio Eng Soc, 1977, 25: 196–200

35 ITU-R Report BS.775-3. Multichannel stereophonic sound system with and without accompanying picture. In: Proceedings of International Telecommunication Union. Geneva, 2012

36 ITU-R Report BS.2159-7. Multichannel sound technology in home and broadcasting applications. In: Proceedings of International Telecommunication Union. Geneva, 2015

37 Theile G, Wittek H. Principles in surround recordings with height. In: Proceedings of 130th Audio Engineering Society Convention. London, 2011. Paper 8403

38 Hamasaki K. The 22.2 multichannel sounds and its reproduction at home and personal environment. In: Proceedings of 43rd Audio Engineering Society International Conference. Pohang, 2011

39 Yi K, Xie B. Local Ambisonics panning method for creating a virtual source in the vertical plane of the frontal hemisphere. Appl Acoust, 2020, 165: 107319

40 Wightman F L, Kistler D J. Headphone simulation of free-field listening. I: Stimulus synthesis. J Acoust Soc Am, 1989, 85: 858–867

41 Saviojia L, Huopaniemi J, Lokki T, et al. Creating interactive virtual acoustic environments. J Audio Eng Soc, 1999, 47: 675–705

42 Bauck J, Cooper D H. Generalized transaural stereo and applications. J Audio Eng Soc, 1996, 44: 683–705

43 Gardner W G, Martin K D. HRTF measurements of a KEMAR. J Acoust Soc Am, 1995, 97: 3907–3908

44 Algazi V R, Duda R O, Thompson D M, et al. The CIPIC HRTF database. In: Proceeding of 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics. New York, 2001. 99–102

45 Majdak P, Balazs P, Laback B. Multiple exponential sweep method for fast measurement of head-related transfer functions. In: Proceeding of 122nd Audio Engineering Society Convention. Vienna, 2007. Paper 7019

46 Xie B S. Head-related transfer function and virtual auditory display (in Chinese). Sci Sin-Phys Mech Astron, 2009, 39: 1268–1285 [谢菠荪. 头相关传输函数与虚拟听觉重放. 中国科学: 物理学 力学 天文学, 2009, 39: 1268–1285]

47 Xie B S. Spatial sample, interpolation for HRTF and signal mixing for multichannel surround sound (in Chinese). Acta Acoust, 2007, 32: 77–82 [谢菠荪. 头相关传输函数空间采样、插值与环绕声重放. 声学学报, 2007, 32: 77–82]

48 Pulkki V. Virtual sound source positioning using vector base amplitude panning. J Audio Eng Soc, 1997, 45: 456–466

49 Välimäki V, Parker J D, Savioja L, et al. Fifty years of artificial reverberation. IEEE Trans Audio Speech Lang Process, 2012, 20: 1421–1448

50 Xiang N, Trivedi U, Xie B S. Artificial enveloping reverberation for binaural auralization using reciprocal maximum-length sequences. J Acoust

Soc Am, 2019, 145: 2691–2702
51 Rumsey F. Spatial Audio. Oxford: Focal Press, 2001
52 Lee H. Multichannel 3D microphone arrays: A review. J Audio Eng Soc, 2021, 69: 5–26
53 Rafaely B. Analysis and design of spherical microphone arrays. IEEE Trans Speech Audio Process, 2005, 13: 135–143
54 Zotkin D N, Duraiswami R, Gumerov N A. Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays. IEEE Trans Audio Speech Lang Process, 2010, 18: 2–16
55 Julstrom S. A high-performance surround process for home video. J Audio Eng Soc, 1987, 35: 536–549
56 Tsingos N, Chabanne C, Robinson C, et al. Surround sound with height in games using Dolby Pro Logic IIz. In: Proceedings of 129th Audio Engineering Society Convention. San Francisco, 2010. Paper 8248
57 Schuller G. Fillter Banks and Audio Coding. Cham: Springer, 2020
58 Lecomte P, Gauthier P A, Langrenne C, et al. Cancellation of room reflections over an extended area using Ambisonics. J Acoust Soc Am, 2018, 143: 811–828
59 Marston D. Assessment of stereo to surround upmixers for broadcasting. In: Proceedings of 130th Audio Engineering Society Convention. London, 2011. Paper 8448
60 Herre J, Hilpert J, Kuntz A, et al. MPEG-H 3D audio—The new standard for coding of immersive spatial audio. IEEE J Sel Top Signal Process, 2015, 9: 770–779
61 ITU-R Report BS.1116-3. Methods for the subjective assessment of small impairments in audio systems. In: Proceedings of International Telecommunication Union. Geneva, 2015
62 ITU-R Report BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems. In: Proceedings of International Telecommunication Union. Geneva, 2015
63 Søndergaardand P L, Majdak P. The Auditory modeling toolbox. In: Blauert J, ed. The Technology of Binaural Listening. Berlin, Heidelberg: Springer-Verlag, 2013
64 Gong M, Xiao Z, Qu T S, et al. Measurement and analysis of near-field hear-related transfer function (in Chinese). Appl Acoust, 2007, 26: 326–334 [龚玫, 肖峥, 曲天书, 等. 近场头相关传输函数的测量与分析. 应用声学, 2007, 26: 326–334]
65 Li J F, Xu H X, Xia R S, et al. Binaural audio technologies based on human auditory perception (in Chinese). Appl Acoust, 2018, 37: 706–716 [李军锋, 徐华兴, 夏日升, 等. 基于听觉感知特性的双耳音频处理技术. 应用声学, 2018, 37: 706–716]
66 Han Z R, Wu M, Zhu Q X, et al. Two-dimensional multizone sound field reproduction using a wave-domain method. J Acoust Soc Am, 2018, 144: EL185–EL190
67 Xie X F. A mathematical analysis of three dimensional surrounding sound filed (in Chinese). Acta Acoust, 1988, 13: 321–328 [谢兴甫. 三维立体声场的(数学)分析. 声学学报, 1988, 13: 321–328]
68 Xie B S, Mai H M, Rao D, et al. Analysis of and experiments on vertical summing localization of multichannel sound reproduction with amplitude panning. J Audio Eng Soc, 2019, 67: 382–399
69 Liu Y, Xie B S. Analysis using binaural auditory model and experiment on the timbre of Ambisonics recording and reproduction (in Chinese). Acta Acoust, 2015, 40: 717–729 [刘阳, 谢菠荪. Ambisonics声捡拾与重放音色的双耳听觉模型分析与实验. 声学学报, 2015, 40: 717–729]
70 Xie B S, Zhong X L, Rao D, et al. Head-related transfer function database and its analyses. Sci China Ser G-Phys Mech Astron, 2007, 50: 267–280
71 Yu G, Wu R, Liu Y, et al. Near-field head-related transfer-function measurement and database of human subjects. J Acoust Soc Am, 2018, 143: EL194–EL198
72 Zhong X L, Xie B S. Maximal azimuthal resolution needed in measurements of head-related transfer functions. J Acoust Soc Am, 2009, 125: 2209–2220
73 Xie B S. Recovery of individual head-related transfer functions from a small set of measurements. J Acoust Soc Am, 2012, 132: 282–294
74 Xie B S, Zhang T T. The audibility of spectral detail of head-related transfer functions at high frequency. Acta Acust Acust, 2010, 96: 328–339
75 Xie B S, Zhong X L, He N N. Typical data and cluster analysis on head-related transfer functions from Chinese subjects. Appl Acoust, 2015, 94: 1–13
76 Zhang C Y, Xie B S. Platform for dynamic virtual auditory environment real-time rendering system. Chin Sci Bull, 2013, 58: 316–327
77 谢菠荪, 张林山, 管善群, 等. 一种5.1通路虚拟环绕声信号处理方法. 中国专利, ZL 0037495.0, 2014-04-02

78 张承云, 谢菠荪. 5.1通路环绕声的耳机动态虚拟重放方法及其实现装置. 中国专利, ZL 0387541.4, 2016-10-05
79 谢菠荪, 张承云, 刘路路. 一种多通路三维空间环绕声的虚拟重放方法. 中国专利, ZL 1297263.8, 2020-8-18
80 刘路路, 谢菠荪. 一种扬声器虚拟声重放的音色均衡方法. 中国专利, ZL 0310956.7, 2021-7-20
81 Hirvonen T, Robinson C Q. Extended bass management methods for cost-efficient immersive audio reproduction in digital cinema. In: Proceedings of 140th Audio Engineering Society Convention. Paris, 2016. Paper 9595
82 Vorländer M. Auralization, Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality. Berlin: Springer-Verlag, 2008
83 Simon L S R, Dillier N, Wüthrich H. Comparison of 3D audio reproduction methods using hearing devices. J Audio Eng Soc, 2020, 68: 899–909

Principle of spatial sound and related physical and auditory problems

XIE BoSun*

Acoustic Lab, School of Physics and Optoeletronics, South China University of Technology, Guangzhou 510641, China

Spatial sound aims to record, transmit, and reproduce the spatial information of the sound field and recreate the desired spatial auditory events or perceptions. It is a frontier field in acoustics and has undergone rapid advancement recently. It has also been widely used in scientific research, engineering technology, and cultural entertainment. The physical analysis of the sound field and the spatial auditory perception mechanism are two foundations of spatial sound. Signal processing is a technical approach for implementing spatial sound. Therefore, spatial sound handles fundamental and technical problems of sound field, hearing, and practical implementation in the crossfield of acoustics (physics), auditory perceptions, and signal processing. The principle, development, key scientific and technical problems, progress, and representative applications of spatial sound are reviewed in this article. Moreover, the technical and application prospects of spatial sound are addressed.

spatial sound, spatial hearing, sound reproduction

PACS: 43.60.+d, 43.55.+p, 43.66.+y

doi: 10.1360/SSPMA-2021-0273