应用声学
第 37 卷 第 5 期
2018 年 9 月
Journal of Applied Acoustics
Vol. 37, No.5
September, 2018

纪念应崇福院士诞辰100周年

基于听觉感知特性的双耳音频处理技术*

李军锋[1†] 徐华兴[2] 夏日升[1] 颜永红[1]

(1 中国科学院声学研究所 北京 100190)

(2 郑州大学电气工程学院 郑州 450001)

摘要 自20世纪30年代引入立体声以来，人类对逼真的听觉体验一直进行着孜孜不倦的追求。双耳音频处理技术基于人耳听觉感知特性，利用计算机和数字信号处理等技术在听者双耳鼓膜处模拟出与真实场景相同的声压，以期给人以“身临其境”的体验，一直是音频信号处理领域的重要研究内容，特别是近两年随着虚拟现实等应用的蓬勃发展，得到更多关注。该文主要围绕双耳音频处理技术中所涉及的关键环节：双耳录音、双耳合成、耳机重放、扬声器重放、头跟踪等领域，以及相关典型应用场景进行较为系统的介绍，最后给出总结与展望。

关键词 三维音频，双耳技术，耳机重放，扬声器重放

中图法分类号： O429 **文献标识码：** A **文章编号：** 1000-310X(2018)05-0706-11

DOI: 10.11684/j.issn.1000-310X.2018.05.015

Binaural audio technologies based on human auditory perception

LI Junfeng[1] XU Huaxing[2] XIA Risheng[1] YAN Yonghong[1]

(1 *Institute of Acoustics, Chinese Academy of Sciences, Beijing* 100190, *China*)

(2 *School of Electrical Engineering, Zhengzhou University, Zhengzhou* 450001, *China*)

Abstract Since the introduction of stereophonic sound in 1930s, human beings have been pursuing more authentic auditory experience. Binaural technology tries to simulate the same sound pressure at listener's eardrums as that in a real sound scene by using the signal processing techniques, which is expected to provide listeners an immersed sense. It has been an important research topic in the field of audio signal processing, especially with the rapid development of virtual reality in the last two years. This paper mainly focuses on the key issues in binaural technology: binaural recording, binaural synthesis, headphone-based binaural audio playback, loudspeaker-based binaural audio playback, head movement tracking and the associated typical applications. Finally, the summary and prospect are outlined.

Key words 3D Audio, Binaural technology, Headphone playback, Loudspeaker playback

2018-05-24 收稿; 2018-07-13 定稿

*国家重点研发计划项目 (2017YFB1002803), 国家自然科学基金项目 (11722437, 11674352)

作者简介: 李军锋 (1979-), 男, 河南安阳人, 博士, 研究员, 研究方向: 语音/音频信号处理。

†通讯作者 E-mail: lijunfeng@hccl.ioa.ac.cn

1 引言

声音是我们日常交流、传递信息和互相通信必不可少的一部分。人类的听觉系统对声音的感知不仅包括响度、音调和音色等主观属性，还包含声音的空间属性等[1]。基于人耳听觉感知特性的双耳音频处理技术利用信号处理、计算机等技术手段尽可能地在听者双耳鼓膜处模拟出与真实声源场景相同的声压，使听者感知到空间中特定位置的虚拟声像[2]。双耳音频技术使得声音具有强烈的空间感、包围感和沉浸感，其在军事航空[3]、虚拟/增强现实[4]、通信多媒体娱乐[5]及科学研究[6]等领域都有着重要应用。

将人类的听音过程看成声源-媒介-接收(Source-Medium-Receiver)过程，自然环境下的听音过程和双耳音频重放中的听音过程存在着很大不同。自然场景下的声源信号不仅包含各个声源的方位信息，也包括周围的环境信息，因此双耳技术首先需要利用双耳录音或者合成虚拟出包含特定声源空间信息的双耳信号。在利用耳机对双耳信号进行重放时，由于不满足自然传输条件，需要对耳机传输函数(Headphone transfer function, HpTF)进行均衡，同时也会存在头内定位、方向混淆等问题。当利用扬声器对双耳信号进行重放时，由于扬声器到听者的双耳之间存在串声(Crosstalk)，实际音频重放系统中需要引入额外的串声消除(Crosstalk cancellation)对双耳信号进行预处理。为了更好地模拟真实自然听觉环境，当听者移动时还需要利用头跟踪技术实时更新相应的声学参数。

本文主要针对双耳音频处理系统所涉及的几个关键技术：双耳录音、双耳音频合成、双耳音频耳机重放、双耳音频扬声器重放、头跟踪等进行较为系统的介绍，并介绍双耳音频技术在不同领域的典型应用，最后给出总结和展望。

2 双耳录音

双耳录音(Binaural recording)，也称人工头录音，是一种与普通立体声拾音不同的录音方式。利用特定材料制作的人工头可模拟出头部、耳廓等生理结构对空间声波的散射和反射过程，通过放置在人工头耳道入口或者耳道内的传声器进行测量或拾拾可获得包含空间声场信息的双耳声信号。早期许多声学先驱(如Steinhauser Thompson和Lord Rayleigh等[7])即认为人类的双耳掌握着人耳听觉系统关于空间中声音的方向、距离等感知的主要信息。1881年，法国的发明家Ader实现了第一个双耳声音传输系统的雏形[8]。利用两根电话线采集巴黎歌剧院现场声音传输给2000 m外的听者，Ader发现利用两个接收端聆听即可带来很好的听觉体验。尽管这项命名为Théatrophone技术，但由于其在当时高额的花费，未能得到广泛应用。

为了更好地重放真实声场，可以利用两个间距18 cm (人头直径的平均尺寸)的麦克风采集空间中的两点声压。基于此想法，众多研究者进行了不同尝试，比较著名的有Harvey 等[9]的双耳助听器(Binaural hearing aid)和Doolittle[10]的双耳广播系统(Binaural broadcasting)。1927年，Bartlett[11]申请人工头(Artificial head)专利，开始研究麦克风置于耳廓部位的录音效果。1933年，AT&T实验室制造的“Oscar”亮相芝加哥世博会，引起极大关注[12]。随着人工头材料和模型的不断发展，不同类型的人工头相继不断出现[13]，其可更加准确地拾取空间声场信息。近几年也出现了相对便携的简化双耳录音设备，如3Dio公司的自由空间双耳麦克风。

3 双耳音频合成

双耳音频信号获取的最直接方式是利用人工头录制包含相应声源空间方位及环境信息的双耳信号，但其具有费时、耗力等问题。实际应用较多的是基于头相关传输函数的双耳信号虚拟合成方法。虚拟的声学场景主要包含的信息有声源空间方位(方向和距离)信息和周围房间环境信息。

3.1 方向信息模拟

自由场下，利用消声室录制的“干”信号卷积相应方向的左右耳的头相关冲激响应(Head-related impulse response, HRIR)即可得到虚拟位置的双耳信号。然而实际中常见的音频信号大都是声源和环境信息混合在一起的多通道立体声信号。直接做法是将不同通道信号与相应方向的左右耳HRIR进行卷积求和来虚拟某一方向声源场景[14]。然而由于立体声信号通常包含多个声源信息，直接卷积求和会产生不确定的虚拟方向，缺少深度感，同时声像

较窄[15−16]。常用的解决方法是利用声音场景分解(Sound scene decomposition)，从多通道信号中分离出各自的主声音信号(Primary sound)和背景声音信号(Ambient sound)，再利用不同方位的HRIR卷积各自的主声音信号，与背景信号相加合成双耳信号[17]。

双耳音频合成需要用到听者HRIR数据，而HRIR和听者头部生理结构、尺寸密切相关，是具有明显个性化特征的物理量。理想情况下，虚拟声源方向所用HRIR应与实际听者相匹配。常用的方法有两种：直接测量[18−19]或者根据听者头部模型理论计算。直接测量方法利用特定设备对听者进行空间中不同角度测量，其通常耗时费力。相应地也有研究根据互易原理[20]或者其他快速测量算法[21]，但现阶段对每个听者进行测量还不现实。理论计算通常利用光学设备如激光、CT或核磁共振成像的方法获得真人或人工头生理外形的计算机图形，然后利用边界元(Boundary element method, BEM)、有限元(Finite element method, FEM)等计算听者HRIR数据。数值计算方法的一个很大挑战是需要较为复杂的计算模型，特别是求解HRIR的高频信息，计算量较大，主要用于实验研究。

针对HRIR严格理论计算或测量较为复杂、不现实问题，实际中可利用个性化HRIR算法一定程度改善重放声像的性能，相应地主要有基于生理参数匹配和测听实验反馈调节两大类。HRIR在频域或者空间域，可以表示一系列基函数的权重之和。因此个性化HRIR信息通常包含在基函数的权重中，其可基于生理参数利用近似线性回归拟合[22]。实际中可利用测量获得听者相应生理参数，进而拟合出具有个性化特征的权重获得相应HRIR。此外，根据实际听者测量的某些生理参数与实际数据库中不同测量听者生理参数进行误差对比，也可选择数据库中误差较小的HRIR数据[23]。听者选择个性化HRIR的基本思路是从公共数据库中选择HRIR合成双耳信号，听者利用一系列测听实验，根据重放声像的定位性能，选择合适的HRIR直至获得满意的效果。类似地也可调节HRIR 中基函数分解(如PCA分解)的权重系数来获得满意效果[24]。随着获取HRIR数据相对容易以及数据的增多，近年来也有基于机器学习算法的HRIR个性化研究[25]。

不管是实际测量或者个性化HRIR数据，通常都是空间中不同水平角和仰角离散分布，为完整虚拟空间中各个位置的声像实际中需要对HRIR数据进行插值[26]。主要分为局部插值算法，即HRIR根据周围相邻方向测量插值计算得到如双线性插值[27]、比值插值[28]等；全局插值算法，即HRIR根据所有测量方向数据利用合适基函数展开对系数插值如基于球谐级数(Spherical harmonics)的插值[29]、PCA插值[30]等。此外，HRIR数据通常阶数较长，实际中为降低双耳信号与HRIR卷积运算量，特别是针对合成多声源情形，相应地也需要对HRIR进行低阶建模。Li等[31]将HRIR分解成最小相位部分和纯延时，对最小相位部分进行FIR建模，可将几百阶HRIR降低到几十阶，声学上可重现同样的效果，大大降低计算量，利于系统实时应用。

3.2 距离信息模拟

虚拟出接近真实环境的声像距离信息是双耳音频合成中的另一重要组成部分，然而实际中存在很大困难。首先，人耳在实际环境中对声源距离的感知相对方向感知更加的不灵敏[32]，声源距离感知也与听者认知水平(如声源熟悉度)相关，且对于近距离声源感知较远，远距离声源感知较近[33]。此外由于耳机重放时缺乏个性化信息、均衡等会引起头内定位现象。实际中利用个性化HRIR、HpTF均衡、加入混响等能一定程度上提高声像头部外化效果，但仍旧不能确保准确的距离感知[17]。直达声混响比(Direct-to-reverberation energy ratio, DRR)是一个决定距离感知定位的关键因素，但其与房间特性密切相关，实际中需要精细的反复调整。

值得说明的是，双耳音频信号中方位信息完整还包括仰角(即高度)方向，而现有方位信息模拟主要关注的是水平面，也有文献利用双耳房间脉冲响应(Binaural room impulse response, BRIR)中的早期反射来合成高度信息[34]。

3.3 声场信息模拟

为真实重现空间中三维声像，虚拟双耳信号中的声场环境信息是必不可少的，特别是对于室内声场的模拟。此外，双耳信号中包含环境信息还可一定程度上减少或消除耳机重放中的头内定位问题和实现重放声像的距离控制。环境信息虚拟最直接的方法是利用双耳房间脉冲响应BRIR代替HRIR与声源信号进行卷积[32]。同样，实际中BRIR也需

要一定测量且阶数相对 HRIR 更长，与特定房间、听者以及听者和声源的绝对位置相关，更为复杂。将 BRIR 看成 HRIR 与房间冲击响 (Room impulse response, RIR) 的卷积，实际中一定程度上可在听觉上模拟等效的空间环境声效果。因此 BRIR 的模拟转换成房间冲击响应 RIR 的模拟。完整的 RIR 主要包含直达声、早期反射和后期混响 [35]；主要模拟算法可以分为三大类：基于物理特性 (Physics-based modeling) 的建模、基于感知特性 (Perception-based modeling) 的建模和两者混合的建模方法。

基于物理特性的 RIR 建模主要模拟声源在空气中的传播和来自四周边界引起反射的物理机制。从基本原理划分，包括基于几何声学的建模和基于波动声学的建模两大类 [2]。基于几何声学的房间建模思路是构造声场空间的计算表示然后由此导出声音的传播路径。基于波动声学的房间响应是建模构造声源的传输声线路径，其遵循相应的波动方程，然后在虚拟空间中追踪其路径，最后利用数学模型逼近声源模式、空气吸收、边界反射、衍射等特性。相应的主要方法有声线跟踪法 (Ray-tracing method, RTM)、镜像源法 (Image source method, ISM) 和波束跟踪法 (Beam tracing method, BTM)。

基于感知特性建模中早期反射可看成是衰减和延时的直达声，利用具有稀疏间隔分布的 FIR 滤波器表征，其延时和衰减参数通常基于经验确定。后期混响建模早期常用梳状滤波器建模，其能产生一个时移和衰减的直达声，缺点是会出现额外的叠加音调感觉且由于频率响应不为常数会产生频谱染色。而后提出利用全通滤波器的改进算法，频率响应更加平滑，同时其延时与频率相关，一定程度上减少主观听觉上的频谱染色问题。Schroeder[36] 结合梳状滤波器和全通滤波器提出经典 Schroeder 混响算法，该算法成为现代混响算法的基石。Moorer[37] 为了模拟空气的高频衰减特性，对梳状滤波器引入一阶低通滤波器，通过精细调整延时和衰减参数，相对 Schroeder 混响算法带来更自然平滑的混响性能。对 Schroeder 混响算法进行更多关键改进，由 Gerzon[38] 提出，然后有多位研究者改进 (特别是 Jot 等 [39]) 的反馈延时网络算法 (Feedback delay network, FDN) 是如今较为广泛使用的后期混响模拟算法。FDN 算法包含一个多通道延时回路和一个个反馈延时网络，其系统为酉矩阵保证输入输出信号能量相等。反馈矩阵调整每一个反馈路径上的延时，可以看成 Schroeder 混响算法中级联梳状滤波器的推广。通过调整反馈矩阵中非零反馈系数和不等延时长度，可产生更高的混响密度。Jot 等提出了一系列 FDN 算法设计思想，可以较为独立控制不同频带内的混响时间，模拟出高质量的人工混响效果。实际中反馈延时网络的选取至关重要，相关实验和研究表明酉矩阵，如三角矩阵、Householder 矩阵 [39]、Hadamard 矩阵 [40]，可以产生较好的混响模拟效果。

基于物理特性混响模拟不利于实时系统运用，而基于感知特性混响模拟提高了计算速度，但不能很好反映待模拟环境声学特性。综合考虑两种混响模拟的优点，实际中也常采用结合物理特性和感知特性的混合混响模拟方法。Rindel[41] 利用 ISM 建模早期反射而后期混响采用 RTM 实现。此外，Murphy 等 [42] 提出从实际测量 RIR 中直接截取一较短 FIR 响应产生早期混响，后期混响利用 FDN 模拟。但由不同方法模拟的早期反射和后期混响之间的平滑过渡是混合混响模拟算法需要解决的一个重要问题。为解决这个问题，徐华兴等 [43] 和 Xia 等 [44] 提出一种基于物理特性和感知特性的混合混响模拟方法，利用 ISM 建模产生的早期 RIR 卷积得到早期反射，而后期混响利用 FDN 实现。进一步又利用一参数化预测模型估计 ISM 建模产生的早期反射的能量衰减曲面 (Energy decay relief, EDR)，相应地实时自动计算 FDN 参数。所提出算法不仅保证了早期反射到后期混响在时 - 频域的平滑过渡，且一定程度上模拟的后期混响能反映待模拟环境的声学特性。

4　双耳音频信号的耳机重放

录制或合成的双耳信号利用耳机重放时由于耳机不平直的传递函数会破坏双耳感知信息，因此需要相应的均衡。此外由于非个性化 HRTF 影响以及缺乏动态定位因素等也会引起头内定位、前后混淆等问题。

4.1　耳机均衡

通常耳机传递函数包括耳机换能器 (Transducer) 响应和耳机与听者双耳耦合 (Coupling) 响应。由于 HpTF 幅频特性不平直，需要均衡。基本做

法是在听者封闭耳道或鼓膜处测量HpTF，然后双耳录制信号与HpTF进行解卷积(Deconvolved)。这种直接均衡方式称为非耦合均衡模式(Non decoupled mode of equalization)[45]。HpTF由于包含耳机到听者之间的传输响应，因此也具有个性化特征同时与测量位置密切相关。研究表明个性化HpTF均衡和个性化HRIR同样重要，在低频HpTF变化较小，但是在高频其偏差可达10 dB，实际中针对某一个具体测量位置的HpTF均衡可能带来比不均衡更差的效果[46]。实际中对多次测量的HpTF取平均进行均衡，一定程度上可减少均衡效果对位置的依赖性。为减少HpTF均衡效果对听者的高度依赖性，Sunder等[47]提出一种Type-2均衡算法，其利用前方投射耳机(Frontal projection headphone)结构模型，解除耳机与听者双耳耦合关系，只均衡发射端(Eimitter)引起的畸变，保留个性化的耳廓信息。第二种常用的均衡算法是解耦和均衡算法(Decoupled equalization technique)，其采用一个参考声场(Reference sound field, REF)(自由场、扩散场或者一个参考响应)进行均衡[45]。实际中如果录制环境的参考声场已知，其相对非解耦和均衡方法效果更加自然。

针对均衡滤波器具体设计，早期有FIR、IIR滤波器均衡设计。由于人耳对低频段声音更为敏感，Härmä等[48]将线性频率转换成弯折(Warp)频率域进行均衡滤波器设计，优化低频段性能。进一步地，Karjalainen等[49]又提出Kautz滤波器实现更为复杂的频率分辨率映射。但这些滤波器设计都是基于数学意义上误差最小，并不一定听觉最优。Fang等[50]引入考虑人耳听觉特性的相关加权函数，对LMS算法求取均衡滤波器系数的步长加权，在人耳比较敏感的低频段采用较小步长，相反在人耳不太敏感的高频段采用较大步长，合理利用LMS算法收敛速度与稳态误差之间的矛盾，获得低频段均衡误差相对高频段较小的均衡效果。

4.2 方向畸变

耳机重放双耳信号也常常会出现声像方向的畸变，如前后镜像方向的声像混乱(Front-back confusion)和仰角畸变(Elevation error)等。在各种声源方向定位因素中，耳廓等带来的高频谱因素和头动引起的动态定位因素对区分前后镜像方向和中垂面方向的声源定位有着重要作用。而耳廓等引起的高频谱因素极具个性化特征，因此在HRIR与HpTF的测量、合成、均衡以及处理每一环节的误差都可能引起耳机重放的声像畸变[2]。不同的声像实验研究表明个性化的HRIR 双耳声像合成和HpTF均衡，一定程度上可减少耳机重放声像畸变[51]。实际中针对非个性化的HRIR，也有研究者通过修正HRIR频谱一定程度提高性能。Zhang等[52]夸大前后方位HRIR的幅度谱差异，使原有频谱峰值基本保持不变，但谷值部分加深。类似的Park等[53]利用不同的权重函数实现。Lee等[54]则利用不同权重函数，提升前半球方向的HRIR增益，衰减后半球方向的HRIR增益，来减少前后混乱。Tan等[55]则是利用不同频带内衰减不同的增益来减少前后混淆。

4.3 头内定位

耳机重放时存在的另一个问题是重放时虚拟声像主要集中听者头内，这种不自然的听觉现象统称为头内定位(Inside-the-head Localization, IHL)[56]。研究表明，头内定位主要是由声重放在双耳处产生的错误空间信息导致，如非个性化的人工头捡拾信号或者非个性化的HRIR合成以及耳机传输响应不平直等导致的双耳声压畸变。因此个性化的HRIR 和HpTF均衡处理可一定程度上提高头外声像效果[57]。在自然环境中，除了直达声外，增加环境反射声对产生头外声像也很重要，相应地在双耳合成中增加室内空间信息减少头中定位。Xia等[58]利用混合混响建模构建BRIR用于双耳重放，主观实验证实其可提高声像外在化和真实感。

5 双耳音频信号的扬声器重放

相对耳机重放，双耳信号利用扬声器重放可很大程度上避免头内定位问题，同时针对一些特定场景如多人会议通话及家庭影院等，运用扬声器更适宜方便。但扬声器与听者双耳之间存在的串声问题极大地干扰听者对空间三维声像的感知，需要增加相应的串声消除系统对双耳信号预处理。串声消除问题的提出可追溯到1961年，CBS实验室的Bauer[59]在研究录制的立体声双耳信号利用扬声器重放时，分析听者双耳接受到信号丢失空间信息，因为存在扬声器到听者双耳之间的串声。随后1962年Atal等[60]在其专利中具体实现了相应的串声消

除系统，串声消除问题有了雏形。20 世纪 90 年代，考虑听者头部等外在影响，Bauck 等[61] 将串声消除推广到多扬声器多听者情形研究了一般化的串声消除系统。大致来说，串声消除问题的研究可以分为两大类：(1) 从算法实现上，主要关注串声消除滤波器的具体数学求解，尽可能地降低数值误差，国内外研究者分别提出了时、频域不同串声消除滤波器设计算法；(2) 从系统层面，分析串声消除系统对外在干扰如听者头动、外在误差等鲁棒性能，寻找较优的扬声器与听者布置。

5.1　串声消除滤波器求解

串声消除直接实现是对频域传输矩阵 (扬声器到听者双耳之间传输函数组成的矩阵) 直接求逆，但由于在某些频率点传输矩阵可能是奇异的，直接求逆会出现所谓病态问题 (Ill-conditioned problem)，造成某些频率点对声源信号有较大的提升，引起频谱染色和扬声器重放时动态范围损失 (Dynamic range loss)。Kirkeby 等[62] 基于正则化原理提出了频率规整化算法。理想的频率规整化参数应该是频域相关的，Liew 等[63] 考虑频域人耳的听觉掩蔽效应，引入与掩蔽阈值相关的规整化参数。针对频谱染色主要由串消滤波器求逆某些频点峰值引起，但常数的规整化因子，将频谱单峰值变成双峰值，特别是在低频引入滚降 (Roll-off) 特性。Choueiri[64] 具体深入分析引起频谱染色的原因，实际中对串消滤波器峰值设定一给定的阈值，将串消滤波器划分为不同的频带规整化求解来减少扬声器串消的频谱染色。频率规整化参数的引入不仅限制了串消滤波器的频域峰值也减少了串消滤波器的时域阶数[65]，但也造成串消滤波器出现非因果“瑕疵”(Artfacts)。Masiero 等[66] 利用维纳-霍夫分解 (Wiener-Hopf decomposition) 将传输矩阵行列式分解为因果稳定部分 (Causal stable parts) 和非因果稳定部分 (Anti-causal stable parts)，其中非因果稳定的频域与传输矩阵之积的时域解利用窗函数加窗截取其因果部分，求解串消矩阵的全局最小相位规整化解。

串消矩阵的频域直接求解由于可以利用 FFT，其实现相对高效易于实时应用，但也存在圆周卷积效应同时不能严格保证因果性需要增加额外的延时。而时域求解虽然计算复杂度较高但更加准确，常用于离线计算串消滤波器系数。频率规整化的串消形式也可以转换成时域求解[67]。基于时域维纳滤波的固有因果稳定特性，Kim 等[68] 提出了时域的解卷积算法。利用 HRTF 的共极点建模，Wang 等[69] 提出了基于 CAPZ 的共极点串声消除算法，寻求计算复杂度和串消性能的折中。Warp 滤波器由于其接近对数频率尺度更好地符合人耳的听觉特性，Kirkeby 等[70] 将传输矩阵函数 HRTF 利用 Warp FIR 建模，模拟人耳的非线性特性提高串消滤波器的低频性能。同样基于 Warp 变换，Jeong 等[71] 通过求解线性域的串消滤波器系数然后利用 Warp IIR 滤波器建模逼近，提升低频性能。基于 Warp 变换方法从线性域转换到非线性域，对 HRTF 的高频部分进行了平滑，使其更符合人耳听觉特性带来低频性能的提升，但增加了计算量一定程度上牺牲了高频性能。

传统的串声消除系统设计算法针对听者人头某一固定位置而设计，而当听者微小移动 (如 75∼100 mm)，期望的三维声像可能崩塌[72]。针对听者头部微小移动，Ward 等[73] 提出联合最小均方误差设计的概念，以人头为中心在一定区域内利用最小误差准则求解串声消除滤波器系数，寻求串声消除性能和头动鲁棒性之间的平衡。基于类似思想，Huang 等[74] 考虑头部转动的多点位置的串消滤波器设计，Wang 等[75] 在 Ward 基础上，针对不同偏移位置设置不同的权重采用加权的多点串消滤波器设计。进一步地又提出在频域，根据多个位置的频率规整化参数求解串消通道分离度，选取多点中最优的规整化参数[76]。Bai 等[77] 在听者双耳控制点附近增加额外的控制点，分别表示为明区 (Illuminated zone) 和暗区 (Shadow zone)，来提高串消系统对听者偏移的鲁棒性。考虑到实际应用中听者头部转动或者移动具有随机性，Xu 等[78] 引入随机矩阵建模听者头部移动提出基于统计逼近的鲁棒串声消除算法，分析表明其可提高对听者头部微小移动的鲁棒性。

5.2　串声消除系统鲁棒性分析

实际应用中，各种误差的存在 (如不匹配的 HRIR 等) 对串消系统性能的影响可看成串消系统鲁棒性问题，可用传输矩阵条件数 (Condition number) 表征。而串消系统的最佳听音区域 (Sweet

spot) 则说明人头转动或移动时串消系统鲁棒性能。将两者结合，为提高串消系统的总体鲁棒性能，以串消系统条件数为基础或者实验中理论分析最佳听音区域大小确定扬声器听者位置和数目寻找鲁棒的扬声器与听者布置，相应的研究主要分为两大类：线性扬声器阵列分布和环形扬声器阵列分布。

早期，Ward 等 [79] 将扬声器到人耳的传输函数利用简单自由场模拟 (其后又用一个简化结构模型验证)，定义一简单代价函数模拟人头前后左右移动对串声消除性能的影响，得出在频率 600 Hz 以上扬声器之间距离分布的鲁棒性与频率成反比。针对声源的不同频率，不同扬声器应按不同间隔摆放，这可以看成是线性扬声器阵列 (Linear loudspeaker array) 思想的雏形。在此基础上，Ward 等 [80] 又利用传输矩阵的条件数，进一步来论证串声消除的鲁棒性，提出一个简化的针对频率范围 400 Hz∼5 kHz 的分频带线性三扬声器串声消除系统。Yang 等 [81] 对三扬声器整体重放的传输矩阵利用条件数进行鲁棒性分析，论证得出三扬声器系统可提高鲁棒性。Zheng 等 [82] 将传输矩阵的简化自由场传输函数用考虑人头散射的钢球模型代替，分析系统不同频率的条件数，综合考虑串声消除系统鲁棒性和频率范围提出线性扬声器阵列最优分布 (Linear optimal source distribution)。在此基础上设计串消滤波器时考虑多点控制，分频时低频以系统条件数为主要参考指标，高频时考虑传输函数相位畸变，提出八扬声器的线性鲁棒系统 [83]。

与 Ward 等 [79] 几乎同时，南开普敦大学的 Kirkeby 等 [84] 利用自由场模型模拟扬声器到人耳传输过程理论分析串声消除，推导出扬声器与人头中心夹角和串声消除鲁棒频率的“环”频率 (Ring frequency, RF) 关系，提出双扬声器角度 10° 的立体声偶极子 (Stereo dipole) 系统。进一步，Takeuchi 等 [85] 利用奇异值分解、系统条件数同时综合考虑系统频率鲁棒性范围和扬声器重放时的动态范围损失等因素提出了一个扬声器与人头中心夹角随频率而变化的环形最优声源分布扬声器系统 (Optimal source distribution, OSD)，理论上解决了扬声器重放对于不同频段的鲁棒性问题。具体实施时可利用分布在空间 ±90°，±16°，±3° 六扬声器系统来重放声源信号的不同频带范围 (低于 450 Hz，450∼3500 Hz，3500 Hz 以上)，在此基础上，Takeuchi 等 [86] 又分析提出了扩展的三通道 OSD 系统。

一般的串声消除均是考虑单一听者，理论上，扬声器双耳重放设计对应的理想串声消除系统，多个听者都能感受到三维音效，但实际应用仍较为困难，对应的研究还较少。早期，Kim 等 [87] 利用自由场传输函数基于条件数鲁棒性分析，对多扬声器多听者串声消除系统 (主要是 4 扬声器 4 听者) 进行了初步的理论分析和研究。进一步地，Masiero 等 [88] 考虑声波衍射将听者人头用钢球模拟 (Rigid sphere) 基于条件数理论分析了线性扬声器分布和圆形扬声器分布的 4 扬声器 2 听者串消系统，得出实际中不同频率扬声器鲁棒性分布间隔不同且相差较大，实现较为困难。

6 听者头部运动跟踪

实际自然的听觉环境，听者头部会移动或者转动，其会改变声源到听者的传输路径和听者生理结构对声波的反射、散射特性等。因此真实情况下，需要实时跟踪检测听者头部运动，动态调整双耳重放系统相应参数。常用的头跟踪设备有电磁跟踪系统或摄像头 [89]，激光扫描仪 (Laser scanner)[90] 或光学传感器 [91] 等。

真实倾听环境中，声源变化和听者运动是同步且连续的，理想情况下，虚拟的双耳重放系统应能反映出这种动态信息，是线性时变系统。实际中应该考虑和解决的相关问题有 [2]

(1) 虚拟声学场景的刷新率 (Update rate) 问题。刷新率表示动态双耳重放系统单位时间内的刷新次数，次数越高，听觉效果越接近真实的声源倾听情况，相应的需要计算量也越大。实际应用中，受软硬件能力的限制，虚拟的双耳重放系统的刷新率应该在基于心理实验结果的基础上折中选择。早期的 Sandvad 等 [92] 心理声学实验表明，刷新率应高于 10 Hz，否则会显著影响倾听者定位判断速度。

(2) 虚拟声学场景刷新的听觉连续性问题，即从上一场景的信号输出平滑地过渡到新场景信号的输出 (Crossfading)。常用的方法有输出过渡 (Output crossfading) 和参数过渡 (Parameter crossfading)[93]。

(3) 头跟踪系统引起的时间滞后问题。当听者移动时，由双耳重放系统软硬件结构所决定的多种

因素如头跟踪装置的响应时间、不同模块之间的通信和相关算法处理等都会引起时间滞后。不同的心理声学实验研究了滞后时间对听觉的影响，一方面关注对虚拟定位的影响，包括准确性和进行定位所需时间；另一方面关注听觉上的可察觉性。一些研究表明，滞后时间达到150 ms[94]，甚至500 ms[95]对虚拟声源定位的影响较少；但其他一些研究表明滞后时间93 ms将引起明显定位错误且会降低定位判断速度。此外，实际中滞后时间的影响还与所用信号的长度、类型相关。Brungart等[96]研究表明，对于大多数虚拟听觉应用，少于60 ms的滞后时间可以接受，而对于一些要求苛刻的虚拟听觉重放，则需要滞后时间少于30 ms。Stitt等[97]最近研究表明，相对单个声源场景，多声源场景的延时听觉感知阈值更高(高10 ms)。

7　双耳技术的典型应用

虚拟现实系统的目标是尽可能地逼真模拟一个特定虚拟环境，而听觉作为人们感知外界信息的重要形式，利用双耳技术重放，与其他视觉、触觉等信息相结合能显著增加沉浸感，如汽车驾驶的虚拟训练[98]。类似的方法也可用于军事航空等特殊环境，如飞行员训练[99]，当引入三维音频双耳技术时飞行员可以更高效地执行任务、战场模拟或军事训练[100]。增强现实听觉信息可以作为自然听觉信息的增强或补充，双耳技术应用到增强现实可用于各种场景的展示，如各种展览会、娱乐场所展示[101]等。

利用多媒体环境，可实现双耳技术在消费娱乐领域的各种应用，包括游戏、声像节目的制作播放，以及一些插件使得浏览器中也可提供双耳三维音频体验等。目前个人计算机和移动手机等电子产品不断普及，手机、Pad等移动设备声音重放通常采用两侧一对小型扬声器重放，将虚拟听觉技术运用到这类产品中可以显著改善重放效果。目前Cecchi等[102]已有相关研究在移动Pad产品上实现双耳扬声器重放。汽车也渐渐成为现代人娱乐消费的一个重要方向，由于汽车相对特殊的环境，利用双耳扬声器重放技术，可以在实际较狭窄的听觉环境获得更广阔的听觉体验，双耳扬声器重放系统应用到汽车系统的研究也渐渐成为一个研究热点[103]。语音通信如电视电话会议的一个良好前景是通信双方能够自由交谈如同在同一个房间，利用双耳技术将采集的参会者语音以及包含空间位置等特定声学环境信息虚拟给听者，可显著提高可懂度和真实感，特别是针对多个参与者情形[104]。

8　总结与展望

随着2016年虚拟现实“元年”的到来，人类视听体验已不满足于二维平面。而对于听觉，期望在空间中逼真的重现三维声场，给人以“身(声)临其境”的体验将会是不断的追求。相对3D视频，现阶段3D音频技术仍有所差距，而相对其他3D音频技术如WFS、Ambsonics等，双耳音频技术有其自身优点，特别随着手机等移动端设备的普及。本文针对双耳音频技术所涉及的各个主要环节及其相关研究进行了概述。双耳音频合成中个性HRIR获取，耳机重放中的降低头内定位和减少方向混淆，扬声器重放中多听者鲁棒串消系统设计等都将是未来双耳技术走向实质性应用中需要解决的关键问题。

总之，双耳音频处理技术因其广泛的应用前景将持续占据音频信号处理领域重要研究位置。随着研究的不断深入，以及其他相关领域的不断发展，相信将会有越来越多的双耳重放技术问题得到解决，从而在实际中得到更为广泛应用。

参　考　文　献

[1] Blauert J. Spatial hearing: the psychophysics of human sound localization[M]. Cambridge: MIT Press, 1997.

[2] 谢菠荪. 头相关传输函数与虚拟听觉 [M]. 北京: 国防工业出版社, 2008.

[3] 吕嶷, 潘皓, 李锋, 等. 三维音频技术在航空领域的应用与展望 [J]. 电讯技术, 2015, 55(11): 1304–1310.
Lyu Yi, Pan Hao, Li Feng, et al. Application and trends of 3D audio in aviation[J]. Telecommunication Engineering, 2015, 55(11): 1304–1310.

[4] Hong J Y, He J, Lam B, et al. Spatial audio for soundscape design: recording and reproduction[J]. Applied Sciences, 2017, 7(6): 627.

[5] Huang Y, Chen J, Benesty J. Immersive audio schemes[J]. IEEE Signal Processing Magazine, 2011, 28(1): 20–32.

[6] Blauert J, Lehnert H, Sahrhage J, et al. An interactive virtual-environment generator for psychoacoustic research. I: architecture and implementation[J]. Acta Acustica united with Acustica, 2000, 86(1): 94–102.

[7] Paul S. Binaural recording technology: a historical review and possible future developments[J]. Acta Acustica united with Acustica, 2009, 95(5): 767–788.

[8] Hertz B F. 100 Years with Stereo—the beginning[C]//Audio Engineering Society Convention 68. Audio Engineering Society, 1981.

[9] Harvey F, Sivian L J. Binaural telephone system: US, 1624486[P]. 1927-04-12.

[10] Doolittle F M. Radiotelephony: US, 1513973[P]. 1924-11-04.

[11] Bartlett J W. Method and means for the ventriloquial production of sound: US, 1855149[P]. 1932-04-19.

[12] Hammer K, Snow W. Binaural transmission system at academy of music in philadelphia[R]. Memorandum MM-3950, Bell Laboratories, 1932.

[13] Vorländer M. Past, present and future of dummy heads[C]. Proceedings of the Acustica, Guimarães, Portugal, 2004: 13–17.

[14] Garas J. Adaptive 3D sound systems[M]. Dordrecht, Netherlands: Kluwer Academic, 2000.

[15] Goodwin M M, Jot J M. Binaural 3-D audio rendering based on spatial audio scene coding[C]//Audio Engineering Society Convention 123. Audio Engineering Society, 2007.

[16] Breebaart J, Schuijers E. Phantom materialization: a novel method to enhance stereo audio reproduction on headphones[J]. IEEE Transactions on Audio Speech & Language Processing, 2008, 16(8): 1503–1511.

[17] Sunder K, He J, Tan E L, et al. Natural sound rendering for headphones: integration of signal processing techniques[J]. IEEE Signal Processing Magazine, 2015, 32(2): 100–113.

[18] Gardner W G, Martin K D. HRTF measurements of a KEMAR[J]. J. Acoust. Soc. Am., 1995, 97(6): 3907–3908.

[19] Algazi V R, Duda R O, Thompson D M, et al. The cipic HRTF database[C]//Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the. IEEE, 2001: 99–102.

[20] Zotkin D N, Duraiswami R, Grassi E, et al. Fast head-related transfer function measurement via reciprocity[J]. J. Acoust. Soc. Am., 2006, 120(4): 2202–2215.

[21] Enzner G. 3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering[C]//Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on. IEEE, 2009: 325–328.

[22] Nishino T, Inoue N, Takeda K, et al. Estimation of HRTFs on the horizontal plane using physical features[J]. Appl. Acoust., 2007, 68(8): 897–908.

[23] Zotkin D Y N, Hwang J, Duraiswaini R, et al. HRTF personalization using anthropometric measurements[C]//Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on. IEEE, 2003: 157–160.

[24] Holzl J. A global model for HRTF individualization by adjustment of principal component weights[D]. Graz: Graz University of Technology, 2014.

[25] Fayek H, van der Maaten L, Romigh G, et al. On data-driven approaches to head-related-transfer function personalization[C]//Audio Engineering Society Convention 143. Audio Engineering Society, 2017.

[26] Zhang W, Samarasinghe P N, Chen H, et al. Surround by sound: a review of spatial audio recording and reproduction[J]. Appl. Acoust., 2017, 7(5): 532.

[27] Queiroz M, de Sousa G H M. Efficient binaural rendering of moving sound sources using HRTF interpolation[J]. Journal of New Music Research, 2011, 40(3): 239–252.

[28] Freeland F P, Biscainho L W P, Diniz P S R. Efficient HRTF interpolation in 3D moving sound[C]//Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio. Audio Engineering Society, 2002.

[29] Andreopoulou A, Begault D R, Katz B F G. Inter-laboratory round robin HRTF measurement comparison[J]. IEEE Journal of Selected Topics in Signal Processing, 2015, 9(5): 895–906.

[30] Kistler D J, Wightman F L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction[J]. J. Acoust. Soc. Am., 1992, 91(3): 1637–1647.

[31] Li J, Zhang J, Sakamoto S, et al. An efficient finite-impulse-response filter model of head-related impulse response[C]//Proceedings of Meetings on Acoustics ICA2013. ASA, 2013, 19(1): 050173.

[32] Begault D R, Trejo L J. 3-D sound for virtual reality and multimedia[M]. US: Academic Press, 2000.

[33] 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.

[34] Karapetyan A, Fleischmann F, Plogsties J. Elevation control in binaural rendering[C]//Audio Engineering Society Convention 140. Audio Engineering Society, 2016.

[35] Valimaki V, Parker J D, Savioja L, et al. Fifty years of artificial reverberation[J]. IEEE Trans. Audio Speech Lang. Process., 2012, 20(5): 1421–1448.

[36] Schroeder M R. Natural sounding artificial reverberation[J]. J. Audio Eng. Soc., 1962, 10(3): 219–223.

[37] Moorer J A. About this reverberation business[J]. Computer music journal, 1979, 3(2): 13–28.

[38] Gerzon M A. Unitary (energy-preserving) multichannel networks with feedback[J]. Electronics Letters, 1976, 12(11): 278–279.

[39] Jot J M, Chaigne A. Digital delay networks for designing artificial reverberators[C]//Audio Engineering Society Convention 90. Audio Engineering Society, 1991.

[40] Stautner J, Puckette M. Designing multi-channel reverberators[J]. Computer Music Journal, 1982, 6(1): 52–65.

[41] Rindel J H. Computer simulation techniques for acoustical design of rooms[J]. Acoustics Australia, 1995, 23: 81–86.

[42] Murphy D, Stewart R. A hybrid artificial reverberation algorithm[C]//Audio Engineering Society Convention 122.

Audio Engineering Society, 2007.

[43] 徐华兴, 夏日升, 李军锋, 等. 一种基于物理特性和感知特性的混响模拟方法 [J]. 中国科学: 信息科学, 2015, 45(6): 817–826.
Xu Huaxing, Xia Risheng, Li Junfeng, et al. A hybrid physically-and perceptually-based approach for reverberation simulation[J]. Scientia Sinica Informationis, 2015, 45(6): 817–826.

[44] Xia R, Li J, Primavera A, et al. A hybrid approach for reverberation simulation[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2015, 98(10): 2101–2108.

[45] Larcher V, Jot J M, Vandernoot G. Equalization methods in binaural technology[C]//Audio Engineering Society Convention 105. Audio Engineering Society, 1998.

[46] Kulkarni A, Colburn H S. Variability in the characterization of the headphone transfer-function[J]. J. Acoust. Soc. Am., 2000, 107(2): 1071–1074.

[47] Sunder K, Tan E L, Gan W S. Individualization of binaural synthesis using frontal projection headphones[J]. J. Audio Eng. Soc., 2013, 61(12): 989–1000.

[48] Härmä A, Karjalainen M, Savioja L, et al. Frequency-warped signal processing for audio applications[J]. J. Audio Eng. Soc., 2000, 48(11): 1011–1031.

[49] Karjalainen M, Paatero T. Equalization of loudspeaker and room responses using Kautz filters: direct least squares design[J]. EURASIP Journal on Applied Signal Processing, 2007, 2007(1): 185–185.

[50] Fang Q, Xu H, Xia R, et al. Equalization of sound reproduction system based on the human perception characteristics[C]//Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2015 International Conference on. IEEE, 2015: 380–383.

[51] Martin R L, McAnally K I, Senova M A. Free-field equivalent localization of virtual audio[J]. J. Audio Eng. Soc., 2001, 49(1/2): 14–22.

[52] Zhang M, Tan K C, Er M H. Three-dimensional sound synthesis based on head-related transfer functions[J]. J. Audio Eng. Soc., 1998, 46(10): 836–844.

[53] Park M, Choi S, Kim S, et al. Improvement of front-back sound localization characteristics in headphone-based 3D sound generation[C]//Advanced Communication Technology, 2005, ICACT 2005. The 7th International Conference on. IEEE, 2005, 1: 273–276.

[54] Lee S, Kim L H, Sung K M. Head related transfer function refinement using directional weighting function[C]//Audio Engineering Society Convention 115. Audio Engineering Society, 2003.

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

[56] Møller H, Hammershøi D, Jensen C B, et al. Transfer characteristics of headphones measured on human ears[J]. J. Audio Eng. Soc., 1995, 43(4): 203–217.

[57] Wightman F L, Kistler D J. Headphone simulation of free-field listening. II: psychophysical validation[J]. J. Acoust. Soc. Am., 1989, 85(2): 868–878.

[58] Xia R, Li J, Xu C, et al. A sound image externalization approach for headphone reproduction by simulating binaural room impulse responses[J]. Chinese Journal of Electronics, 2014, 23(3): 527–532.

[59] Bauer B B. Stereophonic earphones and binaural loudspeakers[J]. J. Audio Eng. Soc., 1961, 9(2): 148–151.

[60] Atal B S, Schroeder M R. Apparent sound source translator: US, 3236949[P]. 1966-02-22.

[61] Bauck J, Cooper D H. Generalized transaural stereo and applications[J]. J. Audio Eng. Soc., 1996, 44(9): 683–705.

[62] Kirkeby O, Nelson P A, Hamada H, et al. Fast deconvolution of multichannel systems using regularization[J]. IEEE Trans. Audio Speech Lang. Process., 1998, 6(2): 189–194.

[63] Liew Y H, Yang J, Tan S E, et al. Power improvement in crosstalk cancellation using psychoacoustic frequency masking[C]//Audio Engineering Society Convention 109. Audio Engineering Society, 2000.

[64] Choueiri E Y. Optimal crosstalk cancellation for binaural audio with two loudspeakers[D]. New Jersey: Princeton University, 2008: 28.

[65] Kirkeby O, Rubak P, Farina A. Analysis of ill-conditioning of multi-channel deconvolution problems[C]//Applications of signal processing to Audio and Acoustics, 1999 IEEE workshop on. IEEE, 1999: 155–158.

[66] Masiero B, Vorlander M. A framework for the calculation of dynamic crosstalk cancellation filters[J]. IEEE/ACM Trans. Audio Speech Lang. Process., 2014, 22(9): 1345–1354.

[67] Kirkeby O, Nelson P A. Digital filter design for virtual source imaging systems[C]//Audio Engineering Society Convention 104. Audio Engineering Society, 1998.

[68] Kim S M, Wang S. A Wiener filter approach to the binaural reproduction of stereo sound[J]. J. Acoust. Soc. Am., 2003, 114(6): 3179–3188.

[69] Wang L, Yin F, Chen Z. A stereo crosstalk cancellation system based on the common-acoustical pole/zero model[J]. EURASIP Journal on Advances in Signal Processing, 2010, 2010(1): 719.

[70] Kirkeby O, Rubak P, Johansen L G, et al. Implementation of cross-talk cancellation networks using warped FIR filters[C]//Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction. Audio Engineering Society, 1999.

[71] Jeong J, Kim J T, Lee J, et al. Design and implementation of IIR crosstalk cancellation filters approximating frequency warping[C]//Audio Engineering Society Convention 118. Audio Engineering Society, 2005.

[72] Bai M R, Lee C C. Comprehensive analysis of loudspeaker span effects on crosstalk cancellation in spatial sound reproduction[C]//Audio Engineering Society Convention 120. Audio Engineering Society, 2006.

[73] Ward D B. Joint least squares optimization for robust acoustic crosstalk cancellation[J]. IEEE Trans. Audio Speech Lang. Process., 2000, 8(2): 211–215.

[74] Huang C R, Hsieh S F. Robust 3-D crosstalk canceller design[C]//Multimedia and Expo, 2007 IEEE International Conference on. IEEE, 2007: 1882–1885.

[75] Wang J, Ye Q, Zheng C, et al. A robust algorithm for binaural audio reproduction using loudspeakers[C]//Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on. IEEE, 2010, 1: 318–321.

[76] Wang J, Ye Q, Wu X. A robust frequency domain crosstalk cancellation algorithm[C]//Computer Application and System Modeling (ICCASM), 2010 International Conference on. IEEE, 2010, 4: V4-568–V4-571.

[77] Bai M R, Tung C W, Lee C C. Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm[J]. J. Acoust. Soc. Am., 2005, 117(5): 2802–2813.

[78] Xu H, Wang Q, Xia R, et al. A stochastic approximation method with enhanced robustness for crosstalk cancellation[J]. Chinese Journal of Electronics, 2017, 26(6): 1269–1275.

[79] Ward D B, Elko G W. Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation[J]. IEEE Signal Process. Lett., 1999, 6(5): 106–108.

[80] Ward D B, Elko G W. A new robust system for 3D audio using loudspeakers[C]//Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on. IEEE, 2000, 2: II781–II784.

[81] Yang J, Gan W S, Tan S E. Improved sound separation using three loudspeakers[J]. Acoustics Research Letters Online, 2003, 4(2): 47–52.

[82] Zheng J, Lu J, Qiu X. Linear optimal source distribution mapping for binaural sound reproduction[C]//Inter-Noise and Noise-Con Congress and Conference Proceedings. Institute of Noise Control Engineering, 2014, 249(7): 939–946.

[83] Zheng J, Zhu T, Lu J, et al. A linear robust binaural sound reproduction system with optimal source distribution strategy[J]. J. Audio Eng. Soc., 2015, 63(9): 725–735.

[84] Kirkeby O, Nelson P A, Hamada H. The "stereo dipole": a virtual source imaging system using two closely spaced loudspeakers[J]. J. Audio Eng. Soc., 1998, 46(5): 387–395.

[85] Takeuchi T, Nelson P A. Optimal source distribution for binaural synthesis over loudspeakers[J]. J. Acoust. Soc. Am., 2002, 112(6): 2786–2797.

[86] Takeuchi T, Nelson P A. Extension of the optimal source distribution for binaural sound reproduction[J]. Acta Acustica united with Acustica, 2008, 94(6): 981–987.

[87] Kim Y, Deille O, Nelson P A. Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners[J]. J. Sound Vib., 2006, 297(1/2): 251–266.

[88] Masiero B, Qiu X. Two listeners crosstalk cancellation system modelled by four point sources and two rigid spheres[J]. Acta Acustica united with Acustica, 2009, 95(2): 379–385.

[89] Song M S, Zhang C, Florencio D, et al. An interactive 3-D audio system with loudspeakers[J]. IEEE Transactions on Multimedia, 2011, 13(5): 844–855.

[90] Georgiou P G, Mouchtaris A, Roumeliotis S I, et al. Immersive sound rendering using laser-based tracking[C]//Audio Engineering Society Convention 109. Audio Engineering Society, 2000.

[91] Kim S, Kong D, Jang S. Adaptive virtual surround sound rendering system for an arbitrary listening position[J]. J. Audio Eng. Soc., 2008, 56(4): 243–254.

[92] Sandvad J. Dynamic aspects of auditory virtual environments[C]//Audio Engineering Society Convention 100. Audio Engineering Society, 1996.

[93] Wenzel E M, Miller J D, Abel J S. Sound lab: a real-time, software-based system for the study of spatial hearing[C]//Audio Engineering Society Convention 108. Audio Engineering Society, 2000.

[94] Bronkhorst A W. Localization of real and virtual sound sources[J]. J. Acoust. Soc. Am., 1995, 98(5): 2542–2553.

[95] Wenzel E M. Effect of increasing system latency on localization of virtual sounds[C]//Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction. Audio Engineering Society, 1999.

[96] Brungart D, Kordik A J, Simpson B D. Effects of head tracker latency in virtual audio displays[J]. J. Audio Eng. Soc., 2006, 54(1/2): 32–44.

[97] Stitt P, Hendrickx E, Messonnier J C, et al. The influence of head tracking latency on binaural rendering in simple and complex sound scenes[C]//Audio Engineering Society Convention 140. Audio Engineering Society, 2016.

[98] Krebber W, Gierlich H W, Genuit K. Auditory virtual environments: basics and applications for interactive simulations[J]. Signal Processing, 2000, 80(11): 2307–2322.

[99] Simpson B D, Brungart D S, Gilkey R H, et al. Spatial audio displays for improving safety and enhancing situation awareness in general aviation environments[R]. Wright State Univ. Dayton Oh Dept. of Psychology, 2005.

[100] Jones D L, Stanney K M, Foaud H. An optimized spatial audio system for virtual training simulations: design and evaluation[C]. Georgia Institute of Technology, 2005.

[101] Jin C, Kan A, Lin D, et al. 3DApe: a real-time 3D audio playback engine[C]//Audio Engineering Society Convention 118. Audio Engineering Society, 2005.

[102] Cecchi S, Virgulti M, Primavera A, et al. Investigation on audio algorithms architecture for stereo portable devices[J]. J. Audio Eng. Soc., 2016, 64(1/2): 75–88.

[103] Bai M R, Hong J R. Signal processing implementation and comparison of automotive spatial sound rendering strategies[J]. EURASIP Journal on Audio, Speech, and Music Processing, 2009, 2009(1): 876297.

[104] Zhang C, Cai Q, Chou P A, et al. Viewport: a distributed, immersive teleconferencing system with infrared dot pattern[J]. IEEE Multi Media, 2013, 20(1): 17–27.