

引用格式:范欣欣,谢凌云,王鑫.三维声双耳渲染算法的定位感知评价分析[J].中国传媒大学学报(自然科学版),2023,30(03):08-14.
文章编号:1673-4793(2023)03-0008-07

三维声双耳渲染算法的定位感知评价分析

范欣欣¹,谢凌云²,王鑫^{1*}

(1. 中国传媒大学音乐与录音艺术学院,北京 100024;
2. 中国传媒大学信息与通信工程学院,北京 100024)

摘要:随着三维声双耳渲染算法的应用越加广泛,建立有效且完备的双耳渲染算法评价体系成为了重要课题。本文基于之前的研究结果,优化了定位感知的评价方法,提取更细化的声学客观特征,并通过偏最小二乘回归方法建立主客观感知模型,探讨对双耳渲染的定位感知机理。实验结果表明,双耳渲染算法对定位感知有显著的优化作用,尤其适用于影视类信号。1kHz以下的中低频双耳特征和双耳间信号频谱能量分布特征是影响定位主观感知的重要特征。

关键词:三维声;双耳渲染算法;主观评价;客观特征;定位

中图分类号:TN912.2 **文献标识码:**A

Perception evaluation and analysis of location of 3D sound binaural rendering algorithms

FAN Xinxin¹, XIE Lingyun², WANG Xin^{1*}

(1. School of Music and Recording Arts, Communication University of China, Beijing 100024, China;

2. School of Information and Communication Engineering, Communication University of China, Beijing 100024, China)

Abstract: With the increasing application of 3D sound binaural rendering algorithms, establishing an effective and comprehensive evaluation system for binaural rendering algorithms has become an important issue. Based on previous research results, this paper optimized the evaluation method of location perception, refined acoustic objective features extraction, and established a subjective-objective perception model through the partial least squares regression method to explore the location perception mechanism of binaural rendering. The results show that binaural rendering algorithms have a significant optimization effect on localization perception, especially for film and television signals. The binaural features at low and medium frequency below 1 kHz and the energy distribution of interaural signal spectrum are important features affecting subjective perception of localization.

Keywords: three-dimensional sound; binaural rendering algorithm; subjective evaluation; objective feature; location

1 引言

近年来,三维声双耳渲染算法被越来越多的应用于线上音乐会、影视剧、电子游戏等领域,使人们借助

耳机获得更具沉浸式的听觉体验。三维声双耳渲染算法的应用场景广泛,受环境及设备限制小,能实现优良的三维声听感效果,适应主流市场,具备较高的研究价值。随着渲染算法的普及,如何有效的评价三

维声双耳渲染算法,成为了一个重要课题。

目前,已存在三维声评价体系的相关研究成果。Rumsey等人专注于空间声品质的评测,总结前人的评价术语,提及宽度、深度/距离等定位评价术语,并基于场景对术语进行独立声源、乐队、环境和场景的划分,并进行主观实验,探究术语间关系^[1-3]。Lindau等人介绍了评价术语总结方法,通过小组讨论和专家评议等方式,提出评价三维声感知的空间声质量评测目录(Spatial Audio Quality Inventory, SAQI)^[4]。Bagousse等人通过多维尺度分析和自由分类等方法对术语进行整理得到四个评价维度^[5]。Francombe等人设计包含不同重放方式和不同节目源的听音实验,通过文本收集、偏爱度评分及聚类分析等方法整理了用于音频质量评价的术语^[6,7]。

在注重音质的同时,定位听感的大幅度提升是三维声双耳渲染算法的一大亮点,由于耳机重放易出现头中定位、方位混淆等问题,双耳渲染算法在声音定位上的处理仍面临不小的挑战。已有学者开展了对渲染算法的定位感知评价研究。Lindau基于SAQI进行了个性化动态双耳渲染听音实验,使用粉噪信号对比双耳模拟信号和真实声场的差异,发现动态渲染技术会导致定位不准确,距离预估错误,声音外化效果差和声染色的问题^[8]。Simon等人提出了用于评价不同HRTF质量的术语,对非个性化双耳渲染算法进行定性研究,总结了包含外化感、前/后定位、侧方定位等共8个评价术语,以用于描述双耳渲染HRTF的差异^[9]。王晶等人提出了无参考的多维双耳渲染评价方法,评价了算法的总体听感、音质、沉浸感和定位准确度,其中定位准确度仅考虑了水平方向,根据被试对静态及动态简单信号的方向判断正确率进行评价^[10]。

本课题组也在双耳渲染的主客观评测方面做了相应的研究^[11,12]。针对三维声双耳渲染算法的主观评价方面建立了三级评价体系,通过MUSHRA的评价方法,对算法的特点、信号适用性及术语间的关系进行探讨;客观评测方面,提取了声学客观特征,建立客观评测模型。然而,目前构建的评价体系,并未经过严谨合理的甄选,导致被试对三级维度打分结果的相关性较高。此外,基于机器学习方法实现的客观评测更加注重模型的预测性能,重要特征难以从感知层面进行解释。因此,本文在此基础上,优化了实验方案,探究影响定位与距离感知的重要因素。选择具有明确物理含义且在主观感知层面可解释的声学客观特征,进行更为细化的分频处理,使用偏最小二乘回归(Partial Least Squares Regression, PLSR)方法进行主观感知结果与客观参数间的关联性

分析,探究双耳渲染的定位感知机理并建立感知模型。本文研究成果旨在为三维声双耳渲染算法在定位维度上的优化提供参考。

2 主观评价实验

2.1 实验设计

(1) 实验信号的确定

参考前人研究及之前的实验结果,在古典音乐、流行音乐、影视声音等近百条三维声素材中,依照涵盖信号种类广泛、定位表现力强、声音元素丰富且符合真实应用场景等条件,选择了5种素材,如表1所示。三维声素材经6种算法渲染处理,得到30个双耳实验信号。双耳渲染算法如表2所示,渲染算法的选择依据是:第一尽量涵盖市面上现有的双耳渲染算法类型,挑选常见且较成熟的算法;第二由专业听音人员听辨,挑选听感上具有差异的算法。其中算法C是直接线性变换得到的,并未进行双耳渲染处理,理论上不存在音质劣化,纳入到实验中用于与其他渲染算法进行对比和分析。

表1 三维声双耳渲染算法主观评价实验素材

信号类型	时长	信号特点
真人电影	26s	包含对白,定位和距离不同的音效声,交响乐的背景音乐,不同声音空间转换,声音元素丰富
流行音乐	32s	包含常规流行音乐配器,不同乐器定位在不同的空间位置,存在移动声源,定位清晰,定位变化明显
阿卡贝拉	20s	教堂录制的阿卡贝拉,音质好,高中低频不同声部,教堂厅堂感,存在移动声源
音效类电影	30s	丛林环境,鸟叫声,野兽叫声,存在移动声源,定位变化明显
电子音乐	29s	包含合成器音色与人声,高中低频分布均匀,存在明显声源定位变化

表2 三维声双耳渲染算法

序号	算法特点
A	基于HRTF卷积,结合优化音质的后处理
B	基于HRTF卷积,结合提升外化效果的后处理
C	将多声道信号通过加权合并直接线性变换得到立体声信号
D	经过Ambisonic编解码,对Ambisonic信号再计算渲染矩阵函数基于HRTF卷积 ^[13]
E	经过Ambisonic编解码,对Ambisonic信号计算渲染矩阵函数再采用虚拟扬声器方法进行渲染,音色较为明亮 ^[14]
F	基于虚拟半球幅度平移技术 ^[15]

(2) 实验术语的选取

通过梳理三维声评价相关的参考文献,选择定位与距离作为实验术语,解释及两极描述语如表3所示。

表3 三维声双耳渲染算法定位主观评价术语

术语	解释(两极描述)
定位与距离	定位和距离感知的综合评价,从声音定位清晰可辨,声音外化感,声源宽度及距离等方面进行综合评价。(劣-优)

(3) 实验流程

实验使用两套重放系统:5.1.4的三维声重放系统,用于播放三维声参考信号;耳机重放系统,播放实验信号。两系统的重放声压级均为75dB(A)。实验地点及重放系统的校准方法与之前的实验相同^[1],本文不做赘述。

本实验共招募被试20名,年龄在18~28岁之间,均为中国传媒大学的在校学生,专业以音乐及录音艺术为主,均有六年及以上专业音乐学习背景或录音混音经验,能够较好地理解所有评价维度的含义。

实验采用“双盲”重复测量实验方法,使用MUSHRA打分法,以扬声器重放的三维声信号作为参考,令被试比较双耳渲染后的不同实验信号在不同评价术语方面的感知差异,从0到100分进行打分。主观评价实验界面是采用JavaScript开发的基于web的操作界面,如图1所示。本文改进了之前的实验流程,让被试每次实验只针对一个评价术语进行评价,实验信号随机排列。此流程可避免不同评价术语间的干扰,有效节省实验时长,避免被试疲劳,提升实验效率及听辨准确性。为检验被试个人的信度,根据预实验结果,将听感特点不明显的算法D进行重复,即被试单次实验需对7个实验信号就一个评价术语进行打分。每位被试单独进行实验,有一位主试人员全程



图1 定位与距离主观评价实验界面

跟随引导,时长约40分钟,实验期间被试可随意切换参考信号与不同的实验信号,为了避免被试持续听音造成听感疲劳,每听辨20分钟令被试休息10分钟。

2.2 实验数据的有效性检验

实验需要进行被试信度检验,用以筛选有效数据。首先检验被试个人的重复信度,计算被试对算法D的重复信号在各个评价术语上两次打分的标准偏差均值,如果该值小于10,则认为该被试个人信度良好,数据可靠,可作后续处理分析。为了消除不同被试打分的差异性,对经过重复信度检验后的数据进行归一化处理^[6]。随后计算克朗巴哈系数,进行被试间一致性信度检验,剔除一致性差的被试结果,最终所有实验信号的克朗巴哈系数都在0.7以上,表明被试对于定位感知的结果具有较高一致性。经重复信度检验和一致性信度检验后,最终保留被试12人,其中男性4人,平均音乐学习背景为9.5年,标准差2.6,平均录音与混音学习年限3.6年,标准差2.19。

2.3 重复测量方差分析

本文采用实验信号(5水平)×渲染算法(6水平)的两因素重复测量方差分析,对比不同渲染算法的感知差异。所有统计分析都使用双侧检验,且显著性水平为0.05,事后检验采用Bonferroni修正法, η_p^2 来估算效应量。

(1) 主效应分析

不同渲染算法在定位与距离维度上的评价结果如图2所示,采用标准误差表示数据的离散程度。不同渲染算法定位与距离感知存在显著性差异, $F(5, 60) = 5.17$, $p = .001$, $\eta_p^2 = .30$ 。使用成对比较考察算法间差异,结果如表4所示,仅显示有显著性差异结果,其中算法1显著优于算法2,下述表格均如此。在定位与距离感知上,算法A的评分最高,无渲染处理的算法C得分最低,算法A显著优于算法C和D,算法B显著优于D。

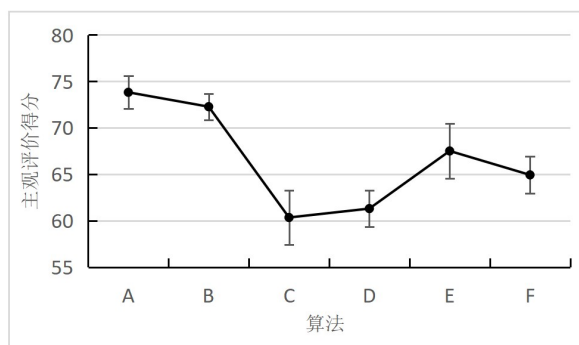


图2 不同渲染算法在定位与距离的主观结果

表4 渲染算法的成对比较结果

算法1	算法2	差值	p值
A	C	13.479	.043
A	D	12.519	.032
B	D	10.970	.001

(2) 实验信号与渲染算法的交互效应分析

实验信号与渲染算法的交互效应结果如图3所示,在定位与距离上, $F(20, 240) = 2.881, p < .001, \eta_p^2 = .19$,存在显著性差异。成对比较结果如表5所示。无渲染处理的算法C在真人电影和音效电影信号上得分均为最低,说明渲染算法在空间定位信息丰富的影视类信号上具有显著的优化作用。算法A和B针对不同信号的得分普遍较高,说明这两个算法的信号适用性较强,可适应不同的渲染场景。此外,在音效电影信号上不同算法的得分差异性较大,此信号是专用于定位空间听感展示的声景素材,能较明显的体现出算法间差异,此类信号可用于不同算法的评测。

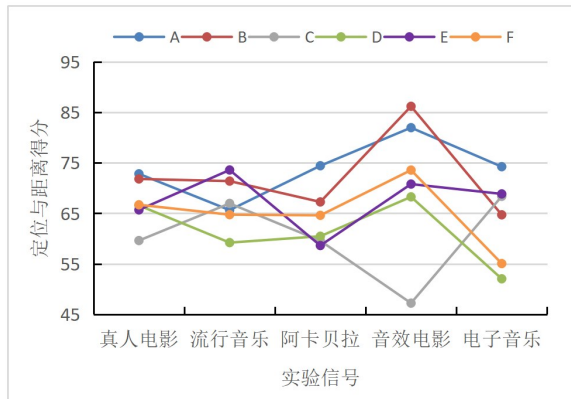


图3 定位与距离的交互效应结果

表5 算法和信号的成对比较结果

信号类型	算法1	算法2	差值	p值
音效电影	A	C	34.765	<.001
	B	C	38.993	<.001
	B	D	17.939	.029
	E	C	23.588	.047
	F	C	26.357	.012
	电子音乐	A	D	22.212
A		F	19.194	.036
E		D	16.819	.044

3 客观评测与感知模型建立

3.1 客观特征的提取与筛选

使用人工头 RS Tech Head001 录制了三维声参考

信号和双耳渲染后的信号,得到30条双耳录音素材,对其提取声学客观特征。文献[12]的结果表明双耳特征及表征中高频频谱能量分布的单耳特征均对定位感知存在较大影响,故提取了双耳特征和单耳时频特征两类特征,所有特征如表6所示,特征的选择依据为:与定位感知相关,且在物理及感知意义上可解释。客观特征的提取流程包括声道分离、分频滤波、分帧、加窗取频谱、特征计算、特征值统计。其中双耳特征主要包含与定位感知相关的信息,包括ILD、ITD和IC,由Two!Ears^[17]工具包提取,根据等效矩形带宽(Equivalent Rectangular Bandwidth, ERB)尺度分42个频带,取25ms帧长;单耳特征主要表征信号的时频特性,由MIR ToolBox^[18]工具包提取,根据倍频程分10个频带,取50ms帧长。均采用帧移50%。选择平均值(记作m)和标准差(记作std)作为统计特征。文献[12]发现差值计算的客观特征模型精度度更好,因此对所有单耳特征进行差值计算。下文对客观特征均采用特征名-统计值-频带数编号的形式进行描述。

表6 客观特征列表

特征名	特征中文名称	是否分频	频带数	统计特征维数	合计维数
ILD	双耳强度差	分频	42	84	522
ITD	双耳时间差	分频	42	84	
IC	耳间一致性	分频	42	84	
RMS	均方根	全频+分频	1+10	22	
ZCR	短时过零率	全频+分频	1+10	22	
Roll Off	频谱滚降	全频+分频	1+10	22	
Flux	频谱通量	全频+分频	1+10	22	
Brightness	明亮度	全频	1	2	
Roughness	粗糙度	全频	1	2	
Regularity	频谱规则度	全频+分频	1+10	22	
Centroid	频谱质心	全频+分频	1+10	22	
Spread	频谱方差	全频+分频	1+10	22	
Skewness	频谱偏态	全频+分频	1+10	22	
Kurtosis	频谱峰度	全频+分频	1+10	22	
Flatness	频谱平滑度	全频+分频	1+10	22	
Entropy	频谱信息熵	全频+分频	1+10	22	
Energy	能量	全频+分频	1+10	22	
Low Energy (Low Rate)	低频能量占比	全频	1	2	

客观特征的筛选流程如图4所示,目的是以最少数目的客观特征精确建立回归模型。首先对客观特征进行归一化处理。然后将每个客观特征与主观评分进行F检验,保留80维影响最显著的客观特征。随后计算保留客观特征之间的相关性,并将相关程度在

0.8以上的组合标记出来,从获得标记数最多的客观特征开始,逐一剔除与其高度相关的特征直至余下特

征间的相关性小于0.8。最终特征数保留在30维以内。

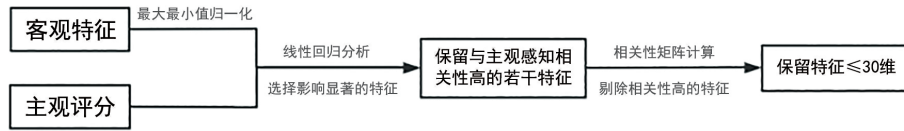


图4 客观特征的筛选流程

3.2 感知模型的建立

本文使用偏最小二乘回归分析(Partial Least Squares Regression, PLSR)建立主客观回归模型。使用 R^2 和 Q^2 描述模型性能, R^2 表示模型的拟合精度,其值越接近1则解释程度越好; Q^2 表示模型的预测能力,其值大于0则具有预测相关性。使用VIP分值评估自变量的重要性,VIP值大于1则为重要变量。

定位与距离的PLSR结果如表7所示,基于 R^2 和 Q^2 结果,保留2个主成分。

表7 定位与距离的 R^2 和 Q^2 值及主成分的方差贡献率

评价术语	R^2	Q^2	主成分数	主成分1	主成分2
定位与距离	0.75	0.46	2	0.56	0.19

图5绘制了定位与距离感知的30个信号的客观特征载荷图,其中横轴表示主成分1,纵轴表示主成分2。图中各点代表该实验信号点主成分得分,各个向量在坐标轴的投影长度代表了各客观特征对于主成分的方差贡献率,红色标记代表不同算法的主观得分。其中主成分1解释了56%的方差贡献率,主要与ITD-sd-5、IC-m-6等中低频双耳特征和Centroid-sd-6、Regularity-m-5等中频段频谱特征有关;主成分2解释了19%的方差贡献率,主要与ITD-m-13、ITD-sd-2等中低频段双耳特征有关。

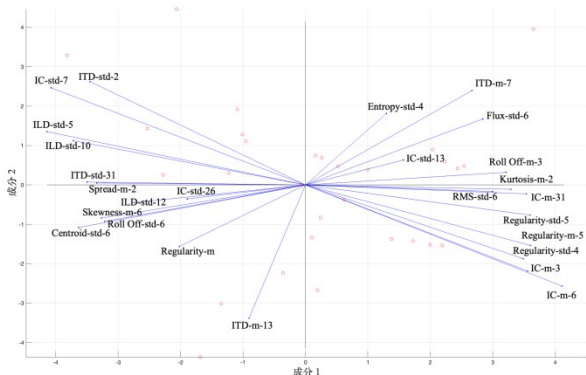


图5 定位与距离感知的客观特征载荷图

为进一步探究哪些特征对定位与距离感知较为重要,计算定位与距离感知PLSR模型中各个特征的VIP值,将结果由高到低横向排列,仅列出VIP值大于1的客观特征如图6所示。由图6可知,中低频段的双耳特征、频谱能量分布及谱结构的时频特征和高频双耳时间差是对定位与距离感知较为重要的特征。

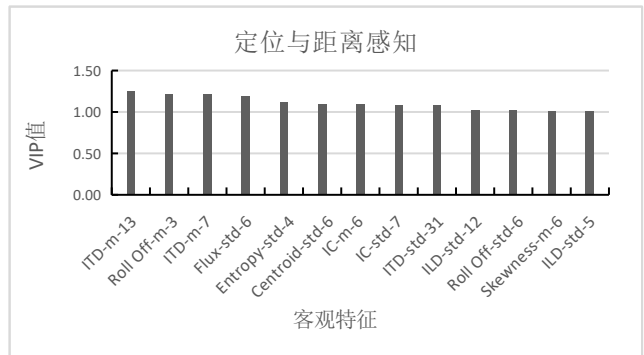


图6 定位与距离感知的客观特征VIP分值图

4 分析与讨论

4.1 双耳渲染算法的对比分析

由主观评价实验可知,6种渲染算法在定位与距离感知上均存在显著差异。算法A和算法B的表现最好,针对不同类型信号具有普适性,两者都是基于HRTF卷积原理实现双耳渲染,可见作为双耳技术的重要基础,HRTF仍是目前实现双耳渲染的最佳方法。算法C的得分最低,该算法是对多声道信号直接线性下混,无额外双耳渲染处理,理论上定位感知效果最差,主观实验结果与理论假设相符,侧面印证了本实验的合理性和有效性,也证实了双耳渲染算法对定位感知存在显著的优化作用,双耳渲染处理存在必要性。算法D普遍较差,算法E整体表现居中,但易受信号种类的影响,性能不稳定,二者均进行了Ambisonic编码处理,再计算滤波矩阵函数实现双耳渲染。虽然Ambisonic的优势在于编解码过程相对独立,可对多声道重放系统灵活适配,且能对声场的空间信息

进行直接处理,实现较为精确的声场重建^[19],但由本次实验结果可知,经过 Ambisonic 的处理并未导致听感上的优化,甚至带来劣化,而且对不同复杂信号的普适性差,此类渲染效果不理想。提出算法 F 的学者使用雷电等干声信号对该算在声像定位的表现进行评测,发现算法 F 的外化感较好,但存在前后混淆和垂直定位偏移等定位准确性问题^[15,20],此结论在本实验中所使用的复杂声音信号中得到进一步印证,算法 F 的定位感知表现差,性能不够稳定,基于幅度平移原理实现的双耳渲染仍存在一定的局限性。

进一步观察渲染算法和实验信号的交互效应结果可发现,相比其他音乐类信号,无渲染处理的算法 C 在真人电影和音效电影两个影视类信号的得分最低,可见双耳渲染处理对空间信息更多、强调定位变化的影视类信号会带来明显的优化作用。此外,在音效电影信号上算法间得分的差异较大,算法 C 显著差于其他算法,可见包含较多移动点声源和丰富声景信息的音效电影类信号更容易体现算法间差异性,适合用作评价不同算法的测试信号。

4.2 双定位与距离维度与特征的关系分析

由 PLSR 结果可以看出,在定位与距离感知上,影响定位与距离感知的重要特征中双耳特征占到一半,这些双耳特征主要处于 200~900Hz,同时具有较高的方差贡献率,可见中低频段的双耳线索对于人耳的定位感知起重要作用。此外,6kHz 左右的双耳时间差对定位亦存在重要影响。处于 100~1kHz 频段的时频特征也占到重要特征的半数,影响较大的参数依次为 RollOff-m-3、Flux-sd-6、Entropy-sd-4 和 Centroid-sd-6 等,即在双耳间 100Hz 左右的频谱滚降差异越大、200~1kHz 左右的信号频谱能量分布变化程度越相近,则越利于定位与距离的感知。此结论与覃龙靖的研究^[12]结果相吻合。倍频程 6(中心频点为 1kHz,频段从 640Hz~1280Hz)的时频特征包含了 4 个,说明这个频段的频谱能量分布对定位感知有重要的影响,即左右声道信号在约 1kHz 的中频能量分布尽量保持一致,会提升定位的感知。

综上,主要分布在 1kHz 以下中低频范围的双耳特征和时频特征与定位与距离的感知密切相关。对双耳渲染算法进行客观评测及优化处理时,可以首先关注信号在中低频的双耳间差异及能量的分布均匀程度。

5 结论与展望

本文针对三维声双耳渲染算法的定位感知设计了主观评价实验,以多声道扬声器重放信号为参考音频,采用 MUSHRA 主观实验方法对不同渲染算法在不同评价维度上进行打分,通过重复测量方差分析和线性回归分析,对比了不同渲染算法的特点和信号适用性,探究定位与距离维度的感知特点。随后,对实验素材进行声学客观特征提取与筛选,通过偏最小二乘回归分析建立主客观回归模型,探讨人耳的定位感知机理。本文的主要结论如下:

- 双耳渲染对定位感知存在显著优化作用,渲染算法在定位感知上存在显著差异。双耳渲染算法对空间信息更多、强调定位变化的影视类信号会带来明显的优化作用。
- 1kHz 以下的中低频双耳特征是影响定位主观感知的重要特征。双耳间信号频谱能量的分布对定位听感的影响也值得关注。

在现有工作的基础上,未来工作包括研究三维声双耳渲染算法的空间感知机理;探讨适用于双耳渲染算法感知评价的实验信号类型;建立双耳渲染算法客观评价模型,简化主观实验流程;探究双耳渲染算法对不同立体声重放设备的适用性问题等。

参考文献(References):

- [1] Berg J, Rumsey F. Verification and correlation of attributes used for describing the spatial quality of reproduced sound [C]//11th AES Convention 9th Int Conference, 2001.
- [2] Berg J, Rumsey F. Validity of selected spatial attributes in the evaluation of 5-channel microphone techniques [C]// AES Convention, 2002.
- [3] Berg J, Rumsey F. Systematic evaluation of perceived spatial quality [C]//AES International Conference: Multichannel Audio, The New Reality, 2003: 184-198.
- [4] Lindau A, Erbes V, Lepa S, et al. A spatial audio quality inventory (SAQI) [J]. Acta Acustica United with Acustica, 2014, 100(5): 984-994.
- [5] Le Bagousse S, Paquier M, Colomes C. Categorization of sound attributes for audio quality assessment—a lexical study [J]. Journal of the Audio Engineering Society, 2014, 62(11): 736-747.
- [6] Francombe J, Brookes T, Mason R. Evaluation of spatial audio reproduction methods (part 1): elicitation of perceptual differences [J]. Journal of the Audio Engineering Society, 2017, 65(3): 198-211.

- [7] Francombe J, Brookes T, Mason R, et al. Evaluation of spatial audio reproduction methods (part 2): analysis of listener preference [J]. *Journal of the Audio Engineering Society*, 2017, 65(3): 212-225.
- [8] Brinkmann F, Lindau A, Vrhovnik M, et al. Assessing the authenticity of individual dynamic binaural synthesis [C]// *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*. 2014, 3: 5.
- [9] Simon L S R, Zacharov N, Katz B F G. Perceptual attributes for the comparison of head-related transfer functions [J]. *The Journal of the Acoustical Society of America*, 2016, 140(5): 3623-3632.
- [10] Wang J, Qian K, Qiu Y, et al. A multi-attribute subjective evaluation method on binaural 3D audio without reference stimulus [J]. *Applied Acoustics*, 2022, 200: 109042.
- [11] 范欣欣, 祁乐融, 杨钊阳, 王鑫. 三维声双耳渲染算法的主观评价分析 [J]. *复旦学报(自然科学版)*, 2022, 61(05): 527-535+545.
- [12] 覃龙靖, 王鑫, 谢凌云. 三维声双耳渲染算法的客观评测与分析 [J/OL]. *复旦学报(自然科学版)*, 2023, 62(1): 53-62.
- [13] Schörkhuber C, Zaunschirm M, Höldrich R. Binaural rendering of ambisonic signals via magnitude least squares [C]// *Proceedings of the DAGA*. 2018, 44: 339-342.
- [14] McCormack, L., Politis, A. SPARTA & COMPASS: real-time implementations of linear and parametric spatial audio reproduction and processing methods [C]// *AES International Conference on Immersive and Interactive Audio*, 2019.
- [15] Lee H, Johnson D, Mironovs M. Virtual hemispherical amplitude panning (VHAP): a method for 3D panning without elevated loudspeakers [C]// *Audio Engineering Society 144th International Convention*, 2018.
- [16] ITU-R. BS.1284-2: General methods for the subjective assessment of sound quality [S/OL]. [2019-01-21]. <https://www.itu.int/rec/R-REC-BS.1284-2-201901-I/en>.
- [17] Raake A. Two! Ears [CP/OL]. (2016) [2023-03-01]. <http://twoears.eu>.
- [18] Lartillot O, Toivainen P, Eerola T. A matlab toolbox for music information retrieval [C]// *Data Analysis, Machine Learning and Applications*, 2008: 261-268.
- [19] 胡瑞敏, 王晓晨, 张茂胜, 李登实, 王松, 高丽, 杨乘, 杨玉红. 三维音频技术综述 [J]. *数据采集与处理*, 2014, 29(05): 661-676.
- [20] Lee H, Mironovs M, Johnson D. Binaural rendering of phantom image elevation using VHAP [C]// *Audio Engineering Society 146th International Convention*, 2019.

编辑: 赵志军, 龙学锋