arXiv:2512.14259v1 [eess.AS] 16 Dec 2025

探究立体声处理的影响——扩展音频质量开放数据集（ODAQ）的一项研究

Sascha Dick[1], Christoph Thompson[2], Chih-Wei Wu[3], Pablo Delgado[1], Phillip A. Williams[3], 和 Matteo Torcoli[1*]

[1]弗劳恩霍夫集成电路研究所 IIS，德国埃尔朗根[2]鲍尔州立大学，美国曼西[3]Netflix, Inc.，美国洛斯加托斯

摘要

本文介绍了一项初步研究，旨在扩展音频质量开放数据集（ODAQ），以探究立体声处理的影响。我们将 ODAQ 中的单声道伪影与左右声道（LR）和中侧（MS）立体声处理相结合，并将其应用于多种刺激材料，包括独奏乐器、典型的宽立体声混音和硬声像混音。我们在不同的呈现情境下开展了听音测试，其中包括对 MS 和 LR 条件进行直接比较，以及不进行直接比较的情形；收集主观数据，以考察单声道伪影之外的因素，同时审视听音测试方法。ODAQ 数据集新增了素材以及来自 16 位专家听众的主观评分。听音测试结果表明，刺激材料的空间特性和呈现情境均产生了显著影响。值得注意的是，LR 与 MS 之间的若干显著差异仅在直接比较时才会出现。研究结果表明，当空间特性一致时，听众主要评估音色劣化；而当音色质量相近时，才会关注立体声声像。额外单声道锚点的评分在不同立体声特性下总体一致，在 MUSHRA 量表上平均为 65 分，进一步印证了听众更重视音色而非空间印象。

1 引言

研究不同信号劣化的感知影响并收集主观基准真值数据，对于开发客观质量指标和优化音频处理方法至关重要。然而，公开可用的数据集十分稀缺，尤其是在立体声和多声道信号处理的影响方面。为解决公开可用数据集匮乏的问题，我们近期推出了音频质量开放数据集（ODAQ） [1, 2] ，其中包含具有不同信号劣化的音频材料，以及相应的主观评分，重点关注单声道伪影和音色变化。

本文介绍了一项研究，旨在通过新增有关立体声处理和双耳听觉对感知质量影响的数据，扩展 ODAQ。文献中已研究空间声像对整体质量感知的影响 [3, 4, 5, 6, 7], ，研究发现整体质量主要受音色质量支配。根据 Rumsey 等人 [3],的研究，回归分析表明，空间质量约占基本音频质量的 30%。

为了客观评估空间质量，近期的研究

*M. Torcoli 在开展这项工作期间任职于弗劳恩霍夫集成电路研究所（Fraunhofer IIS），目前就职于 Amplifon。

已接受，拟在 音频工程学会（AES）第 159 届大会上发表，2025年10月，美国长滩 .

重点关注采用合成数据 [8] 或增强数据 [9]的数据驱动方法。然而，相关数据集通常不公开，这增加了复现和验证这些方法的难度。

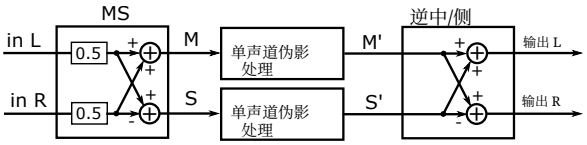
在先前的研究中 [10], ，我们通过主观聆听测试评估了各种孤立信号劣化类型对感知质量的影响。研究重点是单声道信号劣化，并按照图 1a 所示，将相应的处理独立应用于左/右（LR）立体声通道。然而，根据听音者的报告，这随后影响了立体声声像。在音频编码中，中侧（MS）立体声 [11, 12] 被广泛用于对立体声信号进行联合编码，以利用冗余，同时考虑双耳听觉的感知特性。为在 ODAQ 的首次迭代中 [1], 最大限度地减少对立体声声像的改变，对于容易产生空间伪影的伪影类型，我们采用了图 1b 所示的 MS立体声预处理/后处理。

除了实际的信号劣化之外，呈现情境 ——即在一次测试中共同呈现的多种不同条件和质量等级——也会对主观评分和量表使用产生重大影响，正如 Zielinsky 等人 [13]所报告的那样。例如，两个条件之间较小的差异可能只有在直接比较时才会被察觉，而在不同情境中分别呈现时则不会。

为系统研究常用立体声处理技术的影响，同时应对数据稀缺问题，本文基于我们此前的研究，重点在一项统一的听音实验中比较 LR 和 MS 立体声处理方法。本文还探究了呈现情境的影响，即听众是否能够直接比较 LR 和 MS 条件。处理后的材料和主观评分已在线发布于 ODAQ 数据集[1] [2]。



(a) 左右声道（LR）处理



(b) 中侧（MS）处理

图 1：测试信号生成流程中的立体声处理技术

2 方法

2.1 概述

设计与立体声处理相关的听音测试时，一项挑战在于其对空间特性和音色特性的影响，以及这些特性对基本音频质量的相应贡献。作为一种循序渐进的方法，本研究考察了常用立体声处理技术在不同呈现情境和锚点条件下的表现。

2.2 模拟编码伪影

按照 ODAQ [1] 中采用的方法生成了新的测试信号，该方法最初发表于 [10]，通过迫使音频编码器以受控的次优运行模式工作，产生两种相互独立的音频编码伪影。选择了以下两种信号失真，因为它们主要影响音色，并且在我们先前的研究中发现它们容易与立体声编码相互作用：

1. **量化噪声**（QN），通过添加具有频率相关噪声掩蔽比（NMR）的噪声进行模拟；2. **频谱空洞与频谱岛**（SH），由粗量化导致，此处通过以固定且可控的概率随机插入频谱空洞进行模拟。

[1] https://github.com/Fraunhofer-IIS/ODAQ/
[2] https://doi.org/10.5281/zenodo.17162670

方法	参数	质量级别				
		Q1	**Q2**	**Q3**	**Q4**	**Q5**
QN	噪声掩蔽比 [dB]	0	6	12	18	24
SH	空洞概率 [%]	70	50	30	20	10

表 1：单声道伪影和质量等级。

Item	描述	立体声
小提琴	独奏小提琴演奏中速旋律	中心
glock	独奏钟琴演奏清晰的音符	中央
Pop	欢快的流行音乐	wide
RnB	打击乐风格的节奏布鲁斯音乐，配有男声演唱	wide
panDialogM	男性语音 + 氛围音乐	硬声像定位
panDialogF	女性语音 +打击乐风格音乐	硬声像定位

表2：测试刺激列表。

此前的实验中，量化噪声损伤主要用于研究前回声（PE），即量化噪声在瞬态周围的时间扩散。[10, 1] 此处则使用它来评估添加量化噪声对一般信号的影响。这也受到文献中双耳掩蔽级差（BMLD）实验的启发：这些实验采用处于不同耳间相关性条件下的纯音和噪声。[14, 15]

这两种伪影类型都源自频域量化，但在感知上表现为不同的效果。将信号分量粗量化为零，可能会产生随时间波动的频谱空洞或频谱岛，并被感知为额外的信号分量（也称为“啁啾伪影”）；SH 对此进行了模拟。只要量化仍以较粗糙的方式表示原始信号，就会产生额外的量化噪声，QN 对此进行了模拟。这种量化噪声大致遵循信号的时间包络和频谱包络，因此会被感知为类似噪声的信号失真。在之前的实验中，PE 的量化噪声是通过在适当的级别进行实际量化生成的。然而，在低噪声掩蔽比下，这可能会导致类似 SH 的伪影。为确保受测的两种伪影类型明确区分，这里通过向测试信号添加相应级别的随机噪声来生成 QN，而不是进行量化。

上述单声道伪影通过两种不同的策略扩展到立体声信号，即图1所示的低分辨率（LR）和中侧（ MS）处理技术。因此，我们为SH和QN各生成了两种立体声变体，每种变体均包含五个质量等级。质量等级生成所用的详细参数见表1。

2.3 音频材料

测试信号的选择涵盖了三种立体声宽度下广泛的空间特性：位于立体声声像中心的独奏乐器录音；立体声宽度较大的音乐编曲混音；以及经过人工混音、具有硬声像定位的项目。对于每种伪影类型，均选取两个特定项目（即独奏乐器和音乐混音），并针对两种伪影类型分别测试两个硬声像定位项目，因此每种伪影类型包含四个项目，共有六种不同的信号。测试样本的格式为48 kHz / 24 bit立体声，时长为10至16秒。测试项目的详细说明见表2。小提琴、钟琴 和 流行音乐直接取自现有的 ODAQ 数据集；*panDialogF* 和 *panDialogM* 则是利用可用的前景（ FG ）和背景（BG）分轨，对 ODAQ 中对话增强（DE）子集的项目重新混音而成。为本次实验，还在数据集中加入了另一首立体声声像更宽的音乐混音节奏布鲁斯 。

2.4 实验设置

我们采用 MUSHRA [16] 方法开展了一项聆听测试，以评估不同处理方法的主观质量，并探究呈现情境的影响。为考察立体声特性的影响，我们在不同呈现情境下呈现了各项条件，包括有无低分辨率与中/侧之间的直接比较。

分离试次的基本集合共包含两种立体声变体与两种伪影类型的四种组合，每种组合均以五种不同的质量等级呈现。这里将这些试次分别记为 SHLR、SHMS、QNLR 和 QNMS。它们可以直接比较同一种处理类型的所有质量等级，但只能间接比较每种质量等级下的中/侧与低分辨率。

质量级别	Q1		Q2		Q3		Q4		Q5	
立体声处理	低分辨率	中/侧	低分辨率	中/侧	低分辨率	中/侧	低分辨率	中/侧	低分辨率	中/侧
试验										
SHLR	x	-	x	-	x	-	x	-	x	-
QNLR	x	-	x	-	x	-	x	-	x	-
SHMS	-	x	-	x	-	x	-	x	-	x
QNMS	-	x	-	x	-	x	-	x	-	x
SHmix	-	-	-	-	x	x	-	-	x	x
QNmix	-	-	x	x	x	x	-	-	-	-

表 3：测试中选定试验序列的立体声处理方法和质量等级。

较所有质量等级，但只能间接比较每种质量等级下的中/侧与低分辨率。

此外，我们还引入了另一组<style id='1'>混合试验，在同一试验中包含相同质量级别和伪影类型的 MS 与低分辨率项目，以便进行直接比较，从而研究呈现情境的重要性，分别记为$SHmix$ 和$QNmix$。为使待测条件的数量保持合理，我们针对每种伪影选择了两个质量级别，用于直接比较立体声处理方法。相应试验配置的选择也列于表 3 中。

为便于与先前结果进行比较，测试集包含了少量与原始 ODAQ 数据集条件的重叠（流行音乐和钟琴的 SHMS 条件分别对应 [1]中的 SH）。

所有测试均包含 MUSHRA 标准的 3.5 kHz 和 7.0 kHz 低通锚点。为进一步评估空间特性对基本音频质量的贡献，SHmix 和 QNmix 试验中包含了一个单声道锚点。所有试验包含相同数量的待测条件。完整测试由 22 个试验/屏幕组成，每个试验包含七个测试条件和一个隐藏参考。

2.5 测试受试者与聆听环境

这项主观聆听测试在美国鲍尔州立大学（BSU）进行。受试者是在音乐学院音乐媒体制作专业就读的本科生。这些学生都是受过训练的音乐家，除了课程中的技术类课程外，还接受音乐听觉训练。所有听音测试均在适用于音频混音或筛选的声学阻尼听音室中进行。播放音频时，使用了 Beyerdynamic DT770 Pro 250 欧姆封闭式耳机，并搭配专业的 24 位声卡。正式测试开始前，每位听者都参加了一个简短的培训环节，其中包含两个项目，旨在让他们熟悉测试方法、用户界面和测试条件。培训项目不计入正式测试。

测试共由 16 名受试者完成，平均年龄为 20.8 岁（标准差 0.9 岁），平均经验年限为 3.1 年（标准差 1.3 年）。此外，针对主要测试显示值得进一步关注的条件（SHMS 和 SHLR 条件下的硬声像定位项目），还收集了一组规模较小的 8 名专家听众的测试结果。除非另有说明，以下结果均基于 16 名听众组成的主要测试组。较小听众组的结果在图中以灰色 X 标记。

3 结果与讨论

聆听测试结果侧重于不同方面，见图 2 至图 5，其中展示了平均分和自助法计算的 95% 置信区间。为确保所有图中的视觉区分一致，LR 条件用空心标记表示，MS 条件则用实心标记表示。此外，QN 伪影用冷色（蓝绿色）表示，SH 伪影用暖色（橙红色）表示。

3.1 总体结果与质量刻度覆盖范围

图2展示了五个受测质量等级以及分离试次中锚点的整体结果。可以看出，所选质量等级在质量刻度上覆盖良好，且间距相对均匀。然而，QN 的质量等级延伸至更高的

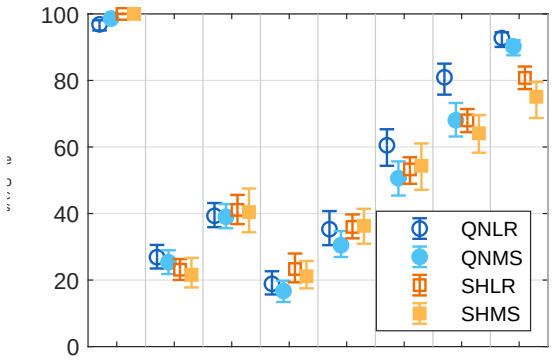


图2：各处理方法和质量级别的总体结果。（均值和95%置信区间，16名参与者）结果表明，质量刻度得到了良好的覆盖。

刻度范围，而 SH 的质量刻度覆盖则更集中在 20 到 80 分之间。采用 NRM、24 dB 条件生成 QN 时所选的最高质量参数，其对应的质量范围高于 10%空洞概率条件下 SH 的最高质量。这表明，未来实验应考虑对 SH 进行旨在提升质量的参数化。

3.2 与此前结果的一致性

为检验一致性，我们纳入了部分与此前在 [2] 中呈现的结果重叠的条件。图3展示了相应项目和质量等级的结果。当前结果的平均评分略高，但总体而言，结果吻合良好，并证实了不同听众群体之间的一致性。

3.3 混合测试：QNmix、SHmix

图 4 展示了 QNmix 和 SHmix 的详细结果，即在相同质量等级下直接比较 LR 与 MS 条件的测试结果，涵盖所有受测单独项目和条件。结果表明，各条件的绝对评分

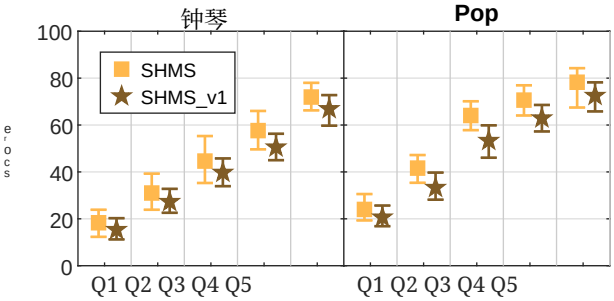


图3：SHMS 与先前测试中的对应条件（标为“SHMS v1”）的结果对比 [2]，显示两者高度一致。

尤其是听者对 LR 或 MS 条件的偏好，在很大程度上取决于具体项目。

3.3.1 低分辨率与中/侧的直接比较

对于硬声像定位项目（*panDialogM*、*panDialogF*），无论是 QN 还是 SH 伪影，LR 条件的评分始终高于 MS 条件。这是意料之中的，因为独立声道的联合编码会导致声道之间出现可听交互和串扰。对于这类项目，在 MS 域中人工生成信号失真会导致声道之间出现可听交互和串扰，其程度超出了常规音频编解码器中更典型的“空间解除掩蔽”特性。

对于独奏乐器项目（QN 的小提琴 和 SH 的钟琴），质量等级的总体评分最低，且 LR 与 MS 的评分相对接近。（对于 SH，SHMS 的平均评分略高，但未达到显著性水平。）结果表明，对于这些纯音调类项目，音调质量失真的影响超过了空间特性方面的任何差异。

对于立体声音乐混音，在两个质量等级的节奏布鲁斯 QNmix 中，以及在较低质量等级（SH30）的流行音乐 QNmix 中，MS 条件的评分显著高于低分辨率。这是意料之中的，因为 MS 编码更能贴合这些宽广但没有明显立体声声像定位的混音的空间属性。然而，对于较高质量等级（SH10）的 SHmix，评分结果正好相反：低分辨率的评分显著高于 MS。鉴

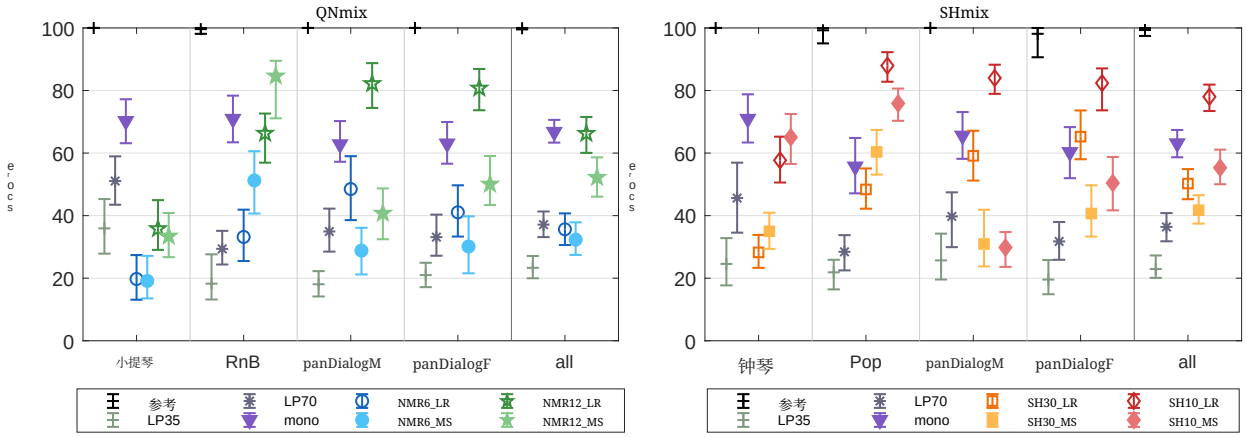


图 4：混合LR/MS测试的详细结果（16 名参与者），显示每个实验和项目的平均分及 95% 置信区间（CI）。结果表明，对LR或MS编码的偏好高度取决于项目特征。此外，即使对于具有硬声像左右立体声特征的项目，单声道锚点的评分也非常高且稳定。

于信号特性，这一结果出乎意料，值得在未来的研究中进一步探讨。一个可能的原因是，在较高质量等级下，左右声道中较为稀疏、彼此独立的频谱空洞所导致的空间波动不太明显，而 M 信号中的频谱空洞会使双耳整体能量损失更大。

3.3.2 单声道锚点条件

实验中加入了单声道下混，以评估整体空间声像的影响，以及单声道信号作为锚点条件的适用性。单声道条件的平均评分为 65 分，即处于 MUSHRA 量表的“良好”范围内。其评分在所有项目中也相当稳定。

单声道锚点的评分高得出人意料，尤其是对于硬声像定位项目而言，因为声源原本鲜明的空间分布完全消失了。可以说，除可能产生轻微的梳状滤波效应外，这种条件下的音色并未受到影响。这表明，总体而言，听众对音色的重视程度远高于对空间属性的重视程度。从这个意义上说，结果与 [3], 的发现一致，该研究指出

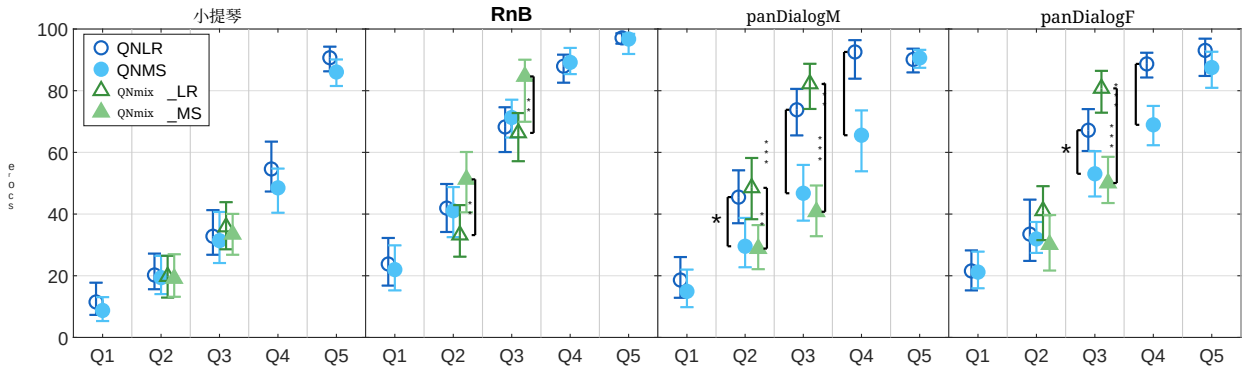
空间质量约占基本音频质量的 30%。不过，有意思的是，这些通过回归分析得出的发现似乎与我们结果中的 MUSHRA 量表约 30% 的比例相当吻合。

此外，对于各项目截然不同的立体声声像特性，单声道条件的评分也相当稳定。因此，它非常适合作为聆听测试中的稳定锚点条件，例如用于未来专注于研究空间属性的实验。

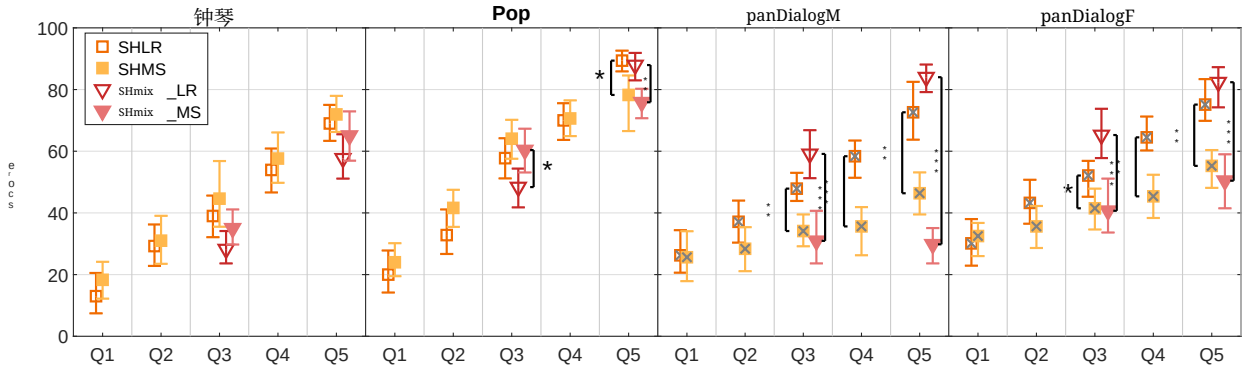
3.4 综合结果

图 5a 和 5b 展示了分离与混合测试变体合并后的各项目详细得分，涵盖所有质量等级以及可用的 LR 和 MS 条件。为聚焦于 LR 与 MS 的比较，同一质量级别和呈现上下文类型下 LR 和 MS 条件之间的显著差异以黑色横线标示，并用星号标记显著性水平。

总体而言，结果表明，在以下试验中，中/侧与低分辨率的评分差异更为明显：



(a) 量化噪声（QN）测试：QNLR、QNMS、QNmix。



(b) 频谱空洞与频谱岛（SH）测试：SHLR、SHMS、SHmix

图 5：按项目和质量级别列出的详细结果，结合了 LR 和 MS 条件的分离比较与直接比较（均值和 95%置信区间，16 名参与者；另外 8 名参与者的结果以灰色 X 标示）。LR 和 MS 条件之间的显著差异以黑色标示（* p¡0.5，** p¡0.1，*** p¡0.001）。结果显示，在若干情况下，只有当听众能够在呈现情境中直接比较两者时，LR 与 MS 之间的差异才达到显著水平，这表明听众主要依据音色质量进行评分，而仅次要地关注空间属性。

听众可以在 QNmix 和 SHmix 试验中直接比较低分辨率与中/侧。对于多个项目，在给定质量级别下，低分辨率与中/侧之间的差异在分离试次中微不足道，但在直接比较中则显著不同。

对于纯器乐项目（小提琴、钟琴），在任何质量级别或呈现情境下，都未发现 MS 与 LR 之间存在显著差异。这表明，对于这些高度调性化的独奏乐器信号，整体质量主要由音色伪影决定。

对于常规的立体声音乐混音（节奏布鲁斯和流行音乐），结果显示，在若干情况下，LR 和中/侧的分离试次得分相近，但在进行直接比较的混合试验中，评级却存在显著差异。（节奏布鲁斯在 Q2 和 Q3 的 QNmix；流行音乐在 Q3 的 SHmix）。这表明，直接比较使听众能够评估空间特性中更细微的差异，而这些差异在其他情况下是无法辨别的。还应注意的是，对具有相似伪影特征的更多质量等级进行评级，可能还受到一种间距偏差的影响

正如 Zielinsky [13]，所报告的那样，即对于能够清晰区分的质量等级，听众倾向于在整个质量刻度上采用更加均匀的间距。

对于硬声像定位项目（panDialogM、panDialogF），无论是分离试次还是相应的混合试验（如结果中有数据），LR 与 MS 之间的显著差异始终存在。对于这些极端的立体声特性，MS 与 LR 之间的差异似乎足够明显，以至于听众即使不进行直接比较也能对其进行评分。这表明，听众能够辨别空间特性上足够大的差异，而不受呈现情境影响。

在 panDialogM 中，SH 的 LR 与 MS 之间差异尤其显著。MS 条件下的整体音质很低，而对于 SHmix，LR 与 MS 之间的差异超过 50 分。一种可能的解释是，对于硬声像定位项目，“M”或“S”信号中的空洞会妨碍 MS 逆向重建，导致左右声道之间出现严重串扰。即使空洞概率较低，也可能产生清晰可闻的杂音和啁啾伪影。这些发现表明，即使空间属性是主导因素，也可能对立体声声像产生非常强烈的影响。不过，这种效应也可以解释为对较安静声道产生了单独而强烈的音色失真。

总之，结果表明，对于不具有极端空间特性的典型项目，听众的评分主要基于音色质量。只有当不同的立体声特性在能够进行直接比较的情境中呈现时，听众才会额外关注空间属性。

4 总结与结论

我们开展了聆听测试，以探究立体声处理、空间特性和呈现情境对主观质量的影响。

主观结果表明，对 LR 或中侧处理的偏好在很大程度上受到测试信号立体声特性的影响：对于立体声声像较窄的录音，中侧处理更受青睐；对于立体声定位明显的项目，LR 更受偏好。对于具有显著音调特性的信号，未发现立体声处理产生显著影响，这表明明显的音色质量变化是主要因素。然而，结果显示，LR 与中侧处理之间的评分差异会受到呈现情境的显著影响；在某些条件下，只有启用直接比较时才会出现差异（例如 SH-mix 和 QN-mix）。研究推测，当空间特性保持一致时，听众主要评估音色劣化；而当音色质量等级相近时，听众则会关注立体声声像失真。因此，在这类情形下，立体声伪影可能会成为“决胜因素”。

纳入的单声道锚点在各项目中的评分始终一致，因此可作为质量刻度的有效锚点。即使对于声像硬性偏置的立体声项目，其评分也相对较高，落在 MUSHRA 量表的“良好”范围内，这一点颇为耐人寻味，但与相关文献中的发现一致：在主观质量评估中，音色差异优先于空间差异。

总之，实验结果有助于深入了解音色和空间印象的评分、空间音频质量聆听测试的设计，并凸显了呈现情境的重要性。此外，将测试信号和主观结果纳入 ODAQ 数据集，标志着 ODAQ 朝立体声和多声道音频扩展迈出了第一步。

5 致谢

我们谨向 Mhd Modar Halimeh、William Wolcott 和 Emanu¨el Habets 致谢，感谢他们在创建 ODAQ 和审阅本文方面给予的宝贵支持。我们也衷心感谢所有参与听音测试的参与者。

参考资料

[1] M. Torcoli、C.-W. Wu 等。《ODAQ：音频质量开放数据集》。载于 *IEEE Int. Conf. Acoustics Speech and Sig. Proc. (ICASSP)*，第 836–340 页，2024 年 4 月。[2] Sascha Dick、Christoph Thompson、Chih-Wei Wu、Matteo Torcoli、Pablo Delgado、Phillip A Williams,

以及 Jinghua Sheng 和 Andreas Silzle。AIR-QA——音频质量开放数据集。载于 *Audio Eng. Soc.*，10 月。*(AES) Conv. 157* 2024

[3] Francis Rumsey、S-lawomir Zieli´nski、Rafael Kassier 和 Søren Bech。空间保真度与音色保真度对劣化多声道音频质量判断的相对重要性。*The Journal of the Acoustical Society of America* 118 2 968 976, 2005 ，（）：–。

[4] Jan-Hendrik Fleßner、Thomas Biberger 和 Stephan D Ewert。对音频质量中单耳与双耳方面的主观和客观评估。*IEEE/ACM Transactions on Audio Speech and* ，，，*Language Processing* 27 7 1112 1125, 2019（）：–。

[5] Yue Qiao、Nick Zacharov 和 Pablo F Hoffmann。利用独立听觉特征映射预测音色、空间及整体音频质量。载于 *Audio Eng. Soc. (AES) Conv. 153*。音频工程学会，2022 年。

[6] Robert Conetta、Tim Brookes、Francis Rumsey、Slawomir Zielinski、Martin Dewhirst、Philip Jackson、Søren Bech、David Meares 和 Sunish George。《空间音频质量感知（第 1 部分）：常见处理过程的影响》。*J. Audio Eng. Soc.*，(): – 。*(JAES)* 62 12 831 846, 2015

[7] Robert Conetta、Tim Brookes、Francis Rumsey、Sla- womir Zielinski、Martin Dewhirst、Philip J ackson、Søren Bech、David Meares 和 Sunish George。《空间音频质量感知（第 2 部分）：线性回归模型》。*J. Audio Eng. Soc. (JAES)*，62(12):847– 860，2015。

[8] Davoud Shariat Panah、Dan Barry、Alessandro Ragano、Jan Skoglund 和 AndrewHines。Binaqual：一种用于双耳音频的全参考客观定位相似度指标，2025。

[9] Yuanming Zheng、Jiaxuan Yao、Xiangyu Deng、Yuhong Yang、Ruiqi Liao、Weiping Tu 和 Cedar Lin。Hapg-saqam：一种由人类听觉感知引导的空间音频质量评估指标。载于 *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1 5, ，第 – 页。2025

[10] S ．Dick, N ．Schinkel-Bielefeld 和 S ．Disch ．孤立音频编码伪影的生成与评估。载于 *Audio Eng. Soc. (AES) Conv. 143*，2017年10月。

[11] J¨urgen Herre、Ernst Eberlein 和 Karlheinz Bran- denburg。《立体声编码的结合》。载于 *Audio Engineering Society Convention 93* ，旧金山，1992年10月。音频工程学会。

[12] J ．D. Johnston 和 A ．J. Ferreira ．和差立体声变换编码。载于 *IEEE Int. Conf. Acoustics Speech and Sig. Proc. (ICASSP)* ，，第 569–572 页，2, 1992 年 3 月。

[13] S. Zielinski、F. Rumsey 和 S. Bech。现代音频质量听音测试中遇到的一些偏差——综述。*J. Audio Eng. Soc. (JAES)*，56(6):427–451，2008 年 6 月。

[14] Jens Blauert。*Spatial hearing: The psychophysics of human sound localization* 。MIT出版社，马萨诸塞州剑桥，1997 年。

[15] D. E. Robinson 和 L. A. Jeffress。改变双耳间噪声相关性对纯音信号可检测性的影响。*The Journal of the Acoustical Society of America* 35 12 1947 1952, 1963 ，（）：–。

[16] ITU-R 建议书 ．BS.1534-3 ．音频系统中等质量级别主观评估方法。国际电信联盟（ITU），无线电通信部门，2015年10月。