混合现实数字音频工作站：迈向混合现实中的协作式数字音频工作站

Torin Hopkins*SolJAMM Research*丹佛，美国
0000-0001-7359-2906

Shih-Yu Ma科罗拉多大学博尔德分校博尔德，美国
0000-0001-6686-0079

Suibi Che-Chuan Weng*ATLAS* 研究所 科罗拉多大学博尔德分校博尔德，美国
chwe1250@colorado.edu

Ming-Yuan Pai新竹县六家高级中学
国立台湾师范大学台北，台湾
maggie93198@gmail.com

Ellen Yi-Luen Do*ATLAS* 研究所 科罗拉多大学博尔德分校博尔德，美国
ellen.do@colorado.edu

Luca Turchet特伦托大学
信息工程与计算机科学系意大利特伦托
luca.turchet@unitn.it

arXiv:2601.00326v1 [cs.HC] 1 Jan 2026

摘要——**数字音频工作站（DAW）是现代音乐制作的核心，但往往会妨碍音乐家的工作流程，将他们束缚在桌前，并阻碍他们与乐器进行自然交互。此外，有效的远程协作仍是一项重大挑战，现有解决方案受网络延迟和异步文件共享的制约。本文探讨混合现实（MR）克服这些障碍的潜力，旨在打造直观的实时远程音乐协作环境。我们采用定性研究和思辨设计技术，以更好地了解：1）演奏者目前如何使用数字音频工作站；2）构想协作式混合现实数字音频工作站的思辨性未来。为促进这一讨论，我们开发并评估了一款设计探针——混合现实数字音频工作站（MR-DAW）的可用性。这是一种混合现实系统，使多个地理分散的用户能够在各自的本地空间中自由移动，同时控制同一个共享的数字音频工作站实例。我们的联网系统使每位远程音乐家都能使用实体脚踏板进行协作循环录制，将熟悉的免提交互方式与共享虚拟会话融为一体。基于对 20 位音乐家的访谈和系统评估，我们分析了当前实践，报告了用户对混合现实系统的体验，并展望了混合现实中音乐协作的未来。研究结果突显了混合现实在实现无拘束音乐交互方面的优势，并对音乐元宇宙中远程协作式数字音频工作站的未来提出了思辨性展望。**

索引词—**音乐元宇宙、音乐扩展现实、数字音频工作站、受限交互**

I. 引言

长期以来，音乐录音一直由音乐家与工程师协作完成。然而，计算和音乐技术的进步 ushered in 了音乐录音的新时代，使音乐家仅使用个人计算机即可轻松录制音乐。这种音乐录音方式已在行业中得到广泛采用，不仅用于创作新音乐，还用于分享音乐创意以及为现场表演准备素材。如今，音乐家将计算机融入音乐工作流，用于上述所有目的及更多场景，包括音乐练习、读谱和远程演奏；这要求他们在创作音乐或手持乐器的同时，能够访问数字资料。

在进行此类活动时，音乐家通常会坐在桌前，一边拿着乐器，一边同时操作计算机。这需要使用鼠标和键盘，或其他音乐设备来控制数字音频工作站（DAWs）——即支持录音及许多其他音乐练习的计算机程序 [1]。对于许多音乐家而言，这种姿势会直接妨碍其工作流程，在他们与乐器之间造成阻碍和距离 [2]。

这种交互范式固有的不便，加上专业级数字音频工作站的计算需求，使协作作曲成为一个困难的过程。参与音乐创作协作的人员往往只能异步进行文件共享，或围绕台式计算机工作，几乎没有空间上的灵活性。尽管联网协作数字音频工作站确实存在 [3]，但它们要么受局域网限制 [3]，，要么就远程数字音频工作站而言[4]，，存在延迟问题、要求轮流操作，或实时文件共享功能非常有限 [5]。

扩展现实涵盖虚拟现实（VR）和混合现实（MR），正迅速成为开展音乐活动（包括协作作曲）的新空间 [6]–[8]。已有研究提出了适用于虚拟现实和混合现实的多种数字音频工作站 [9]–[11]。然而，据我们所知，现有使用传统乐器的系统都聚焦于单用户交互，因为它们并非为多用户环境而设计。这属于音乐元宇宙研究领域的范畴 [12]。

借助混合现实媒介，可以让协作的音乐家在视觉上体验与共享数字音频工作站交互，同时还能让他们在

本项目获得欧盟委员会 EIC Pathfinder Open 计划资助，资助协议编号为 101184379。

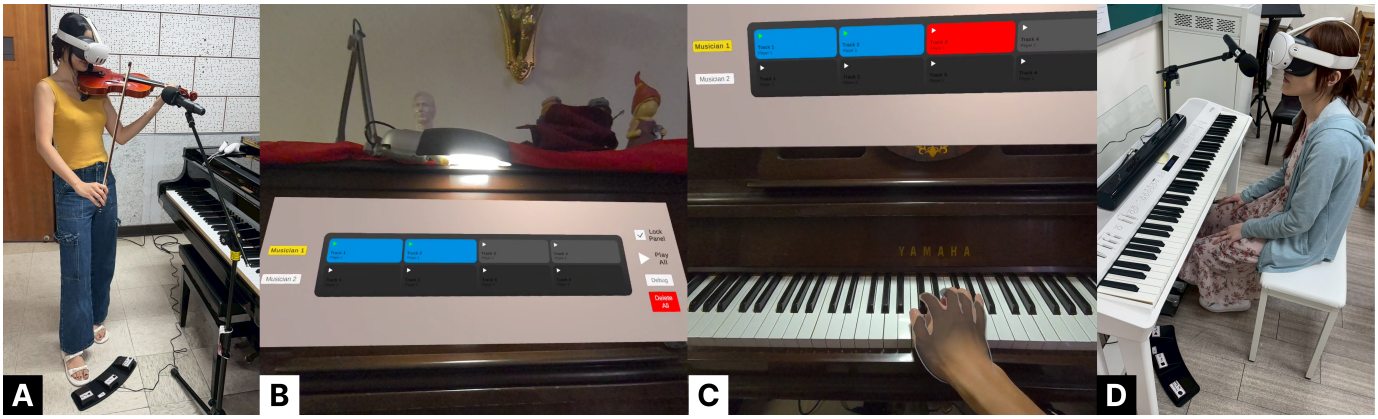

图 1. MRDAW 混合现实音乐协作系统概览。(A) 用户 1 使用脚踏板录音并演奏小提琴。(B) 用户视角下的混合现实画面，空间化用户界面面板及音轨控制叠加在物理环境上。(C) 用户视角下的混合现实画面，展示音轨录制反馈（红色）以及集成于物理空间中的空间用户界面。(D) 用户 2 使用脚踏板录音并演奏钢琴。两位用户均佩戴 Meta Quest 3 头显。

自由、不受束缚地移动。为了配合这种自由，音乐家可以使用外部设备来控制录音过程的各个方面。这些设备包括脚踏板，能够提供许多音乐家熟悉的交互机制，例如吉他手常用的踏板板。这些在共享混合现实环境中为音乐带来的全新交互，使音乐家能够在自己的空间内移动，并以自定义方式组织界面，从而摆脱桌面工作环境的束缚；研究表明，这种环境会妨碍创作过程 [2]。

混合现实系统具有独特优势，能够通过为多位演奏者提供独立界面来控制同一台基于计算机的数字音频工作站，从而克服使用数字音频工作站进行协作的局限。这与当前的模式不同，当前模式需要通过数据密集型文件交换或异步文件共享，同时更新两个或更多系统。借助混合现实同时控制单个数字音频工作站，多位演奏者便能够发送控制消息和近实时音频流——由于采用了网络音乐表演系统，这两项功能已可供音乐家使用，但受地理位置和网络条件的限制 [13]。这使我们得以克服远程协作作曲中的传统障碍，为音乐元宇宙的兴起铺平道路。

本文探讨了与基于计算机的工作流程相比，混合现实在基于数字音频工作站的协作作曲中的优势。首先，我们进行了一系列访谈，20 位参与者在访谈中描述了他们目前如何借助技术进行音乐练习、表演和录制。基于这些认识，我们随后向参与者介绍了我们开发的一款混合现实数字音频工作站，用以展示一种常见的练习和表演技巧：使用脚踏板即时循环录制并叠加音乐。体验该系统后，参与者被要求反思混合现实在协作作曲、演奏和练习中的应用，并推测数字音频工作站在混合现实中的未来应用可能性。

我们的研究为理解混合现实音乐环境中协作的本质提供了新的见解。通过创新的协作循环录制和思辨设计，我们做出了以下贡献：

1) 我们提出了混合现实数字音频工作站，这是一种基于混合现实的协作式数字音频工作站设计探针，使演奏者能够创作并录制新音乐，并与基于计算机的数字音频工作站直接交互；2) 我们介绍了一种协作式联网循环踏板；3) 我们报告了对 20 位音乐家的当前工作流程进行主题分析的结果；4) 我们讨论了主题分析的结果，涉及当前技术的使用，以及音乐元宇宙中协作式数字音频工作站的推测性未来。

II. 系统描述

我们的设计探针——混合现实数字音频工作站（MR-DAW）——面向两位用户设计，但也可以轻松扩展至更多用户（图2）。系统由四个核心组件构成：混合现实用户界面、脚踏板、音频接口，以及基于 Python 的 Ableton Live 控制器。该架构确保虚拟交互、实时音频与专业数字音频工作站功能无缝集成。下文将详细介绍这四个组件。

A. 混合现实用户界面

该混合现实面板使用 Unity 3D 开发，并部署于 Meta Quest 3 头显上，是主要的视觉与交互中枢（见图1C、D）。其中包括：

- **八条音轨：**以可视化方式呈现循环音轨，并实时反馈录音（红色）、播放（绿色）和选择（蓝色）状态；
- **会话控制按钮：**空间用户界面控件，其功能与实体脚踏板相对应，可控制全局播放和单条音轨的切换；
- **调试窗口：**用于系统诊断（网络状态、音频电平、错误）的独立面板。

这一空间化 UI 能实现自然且具身化的交互，让音乐家摆脱传统的桌面布置。

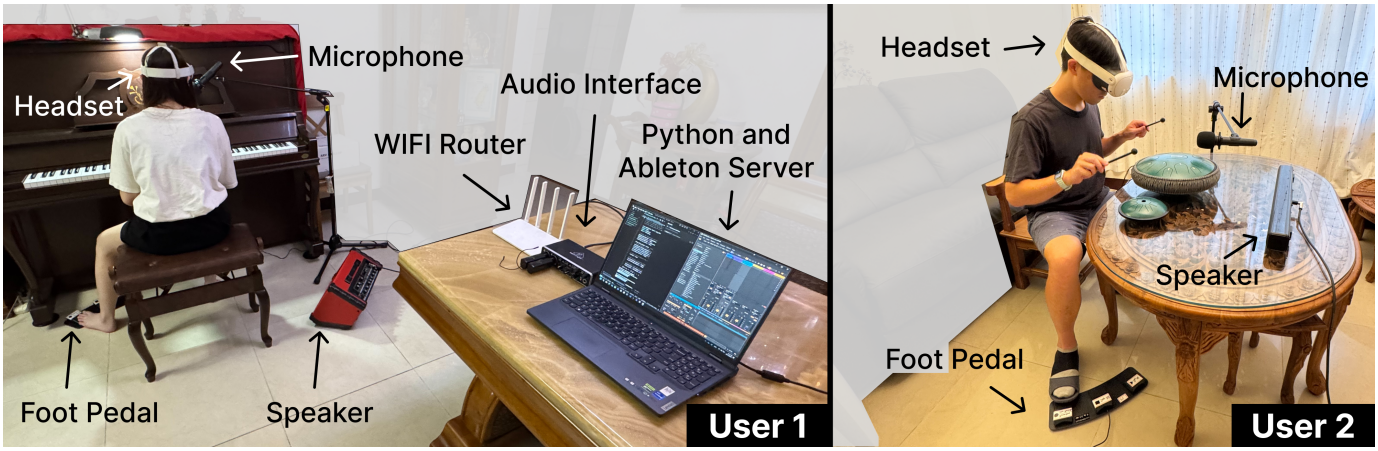

图2. 与混合现实数字音频工作站（MR DAW）交互的两种不同用户设置。“用户 1”（左侧）通过头戴式耳机与虚拟钢琴交互，同时由麦克风采集音频。该设置包括一台显示数字音频工作站软件（Ableton Live 和 Python）的笔记本电脑、一个音频接口、一台 Wi-Fi 路由器、一个扬声器和一个脚踏板。“用户 2”（右侧）使用鼓槌和头戴式耳机演奏虚拟鼓。其设置包括一个麦克风、一个扬声器，以及放在地板上的一个脚踏板。这两种配置展示了在混合现实环境中创作音乐和进行交互的不同方式。

B. 音频接口与通信

一台 Behringer U-Phoria UMC204HD 音频接口负责管理音频输入/输出，采集麦克风输入，并向每位用户提供同步播放。每个工作站都配有一支安装在支架上的麦克风和一台扬声器，分别放置在独立的物理空间中，以模拟远程协作。

C. Ableton Live 控制器

Ableton Live 12 负责协调音乐演奏。自定义 Python 控制器管理用户输入并自动执行 Live 功能，从而衔接混合现实交互与 Live。连接到 Behringer 音频接口的中央笔记本电脑运行 Live。开放声音控制协议（OSC）支持头戴式耳机与中央笔记本电脑（运行 Ableton Live）之间的低延迟网络通信 [14]。Python 脚本与 Live（通过应用程序接口/OSC）和脚踏板通信。每位用户均操作专用脚踏板，以实现触觉控制：

- **音轨录制：**启动和停止单独的循环录音；
- **播放：**全局播放/停止命令。

这一设置确保了精确的同步与控制，使音乐家能够专注于表演。

III. 交互

混合现实数字音频工作站通过其基于循环的环境中的三类主要交互，促进协同音乐创作：循环音轨录制、音轨播放控制和实时通信。这些交互旨在直观易用，并支持音乐家在混合现实环境中实现无缝的协同创作工作流程，同时结合使用手势和可扩展控制器（图 3）。

A. 循环音轨录制

MR-DAW系统中的每位用户最多可以录制四条独立音轨。录音过程通过专用脚踏板启动和结束。脚踏板提供了一种触觉且具身化的控制方式，使音乐家能够专注于乐器或表演。音轨正在录制时，空间混合现实用户界面面板上与之对应的视觉元素会变为红色，提供即时反馈（见图 1D），以表明该音轨正处于录制状态。再次踩下脚踏板后，系统会自动停止录音，对新录制的音频进行量化并循环播放，随后激活下一个可用的空音轨，为后续录音做好准备。

协同同步的一项关键设计决策是，**循环时长由用户 1 录制的第一条音轨决定**。这为整个协作会话确立了基础速度和循环长度。因此，之后任一用户录制的所有音轨都会自动裁剪或延长，以精确匹配这一初始循环时长，确保所有分层声部之间完美的节奏同步。这一功能对于在实时协作即兴演奏场景中维持连贯的音乐结构至关重要。当两位用户都录满各自分配的四条音轨（共八条音轨）后，系统会自动重置到第一条音轨。这一机制使用户能够持续覆盖现有循环或叠加新的录音，从而在固定循环结构内进行迭代创作和动态音乐探索。

B. 音轨播放控制

用户可以全面控制播放，既可使用实体脚踏板，也可使用空间混合现实用户界面面板。脚踏板配有两个专用按钮：一个用于开始播放所有启用的音轨，另一个用于停止所有音轨（见图1A、B）。用户界面面板提供了功能相同且易于使用的切换按钮，作为另一种可视化控制方式。此外，空间用户界面还为每条音轨提供了单独的切换按钮，让音乐家可以有选择地启用或禁用特定的

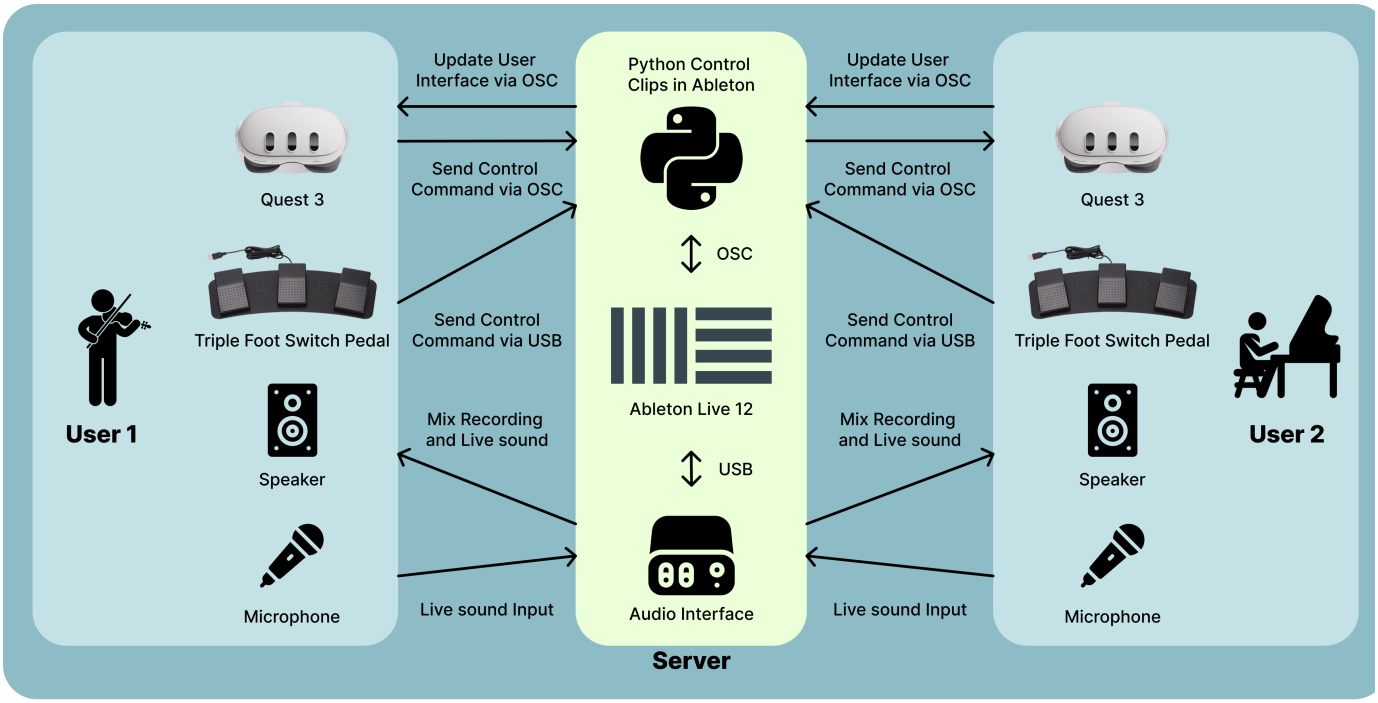

图 3. 实时协作式混合现实音乐环境的系统架构。系统以运行 Ableton Live 12 和 Python 控制脚本的服务器为核心，两者通过开放声音控制（OSC）进行通信。每位用户都配备一台 Quest 3 头显、一个三踏板脚踏开关、一个麦克风和一个扬声器。来自头显（通过 OSC）和脚踏板（通过 USB）的控制命令会发送至服务器上的 Python 脚本。脚本处理这些输入以控制 Ableton Live，并通过 OSC 将视觉反馈发送至用户的头显界面。同时，每位用户麦克风采集的实时音频都会传输至中央音频接口。Ableton Live 对这些实时声音进行处理，并与数字音频混音，最终将混音发送回每位用户的扬声器。

循环。这种精细控制对于协作评审环节中的专注聆听至关重要，使用户能够单独聆听各个部分、尝试不同编曲，并实时优化混音。

C. 实时通信

为了促进真实且自然发生的音乐协作，混合现实数字音频工作站融入了强大的实时通信系统。系统持续监测每位用户的麦克风输入，并将实时音频流直接传送至另一位用户的扬声器。这样便创建了开放的音频通道，使用户能够自然地进行口头交流，也让音乐家即使身处不同的房间，也能听到彼此的实时器乐演奏。这种设置对于自发的音乐交互至关重要，支持用户使用吉他、键盘等实体乐器共同创作，或加入人声，近似还原在同一物理空间即兴合奏的体验。

IV. 评估

A. 评估方案

我们的评估基于对10对参与者进行的半结构化访谈，以及对参与者回答的主题分析。借鉴思辨设计技术 [15]–[17]，我们请参与者完成一个由三部分组成的用户研究流程。具体而言，我们对这10对参与者中的每一对都进行了两次半结构化访谈，均以两人一组的形式进行，并在第一次访谈后使用了设计探针 [18]–[20]。

三个部分详述如下：

1) 访谈 1：首次访谈主要讨论了他们使用数字音频工作站的当前工作流程；2) 系统使用：第二部分包括两人一组使用我们的设计探针（见 C 节）；3) 访谈 2：使用设计探针后，我们讨论了实时音乐协作以及在混合现实中使用数字音频工作站。在这次访谈中，我们请他们设想未来的混合现实数字音频工作站需要整合哪些功能和交互技术。具体而言，这类系统会面临哪些主要挑战，需要哪些交互和功能。

我们对10对参与者的访谈进行了录音和转录，并采用扎根理论指导下的归纳式主题分析进行编码 [21], [22]。作者1和2分别独立对全部20位参与者的数据进行编码和分析，从而能够自然地归纳出主题。随后，我们对这些主题进行了充分讨论，并就数据集中最突出的主题达成共识 [23], [24]。我们使用 MAXQDA[25] 软件跟踪各自独立的编码数据库，以便开展讨论。经过讨论后，我们合并数据库，形成最终主题。以下各节将对这些主题进行概述。

参与者

本研究涉及20位参与者，他们拥有多元的音乐背景，并在音乐创作和录音方面具有专业经验。

人口统计信息：参与者来自不同地区，其中 12 人来自台湾，8 人来自美国。12 位参与者自我认同为女性，8 位为男性。参与者的年龄介于 18 至 54 岁之间（M = 30.22，SD =7.85）。人数最多的年龄组为 25–34 岁（n=16），其次是 18–24 岁（n=2）、35–44 岁（n=1）和 45–54 岁（n=1）。

音乐经验：参与者拥有丰富的音乐经验，平均音乐经验为 20.33 年（中位数 = 23.00）。他们的乐器演奏技能各不相同。钢琴和吉他是最常提及的乐器，参与者通常还会演奏其他乐器，例如小提琴、二胡、贝斯、打击乐器和木管乐器（如萨克斯管、长笛、单簧管）。

技术熟练度：参与者表示，他们熟悉多种音乐技术。常用的数字音频工作站包括 Logic Pro、Ableton Live、GarageBand 和 FL Studio。一些参与者还使用过 Max/MSP 和 ChucK 等创意编程与声音设计环境，以及 Suno 和 Chrome Music Lab 等基于网页的音乐创作工具。该群体普遍具有录音、编辑以及使用交互式音乐界面的经验。

B. 访谈 *1*：当前工具的使用情况

我们首先询问参与者，他们在音乐创作和练习方面的当前工作流程。这让我们更好地了解参与者中的音乐人使用哪些工具来录制、练习音乐，并彼此分享音乐。我们还进一步询问了他们在音乐练习中使用数字音频工作站以及开展协作式共同创作活动的情况。此举旨在促使参与者反思自身的实践，以及如何运用技术促进协作。此外，我们还讨论了鉴于当前可用技术的发展状况，他们使用或愿意使用哪些工具来协作创作。

本节归纳出的主题如下，按重要性不分先后：

*1)*异步文件共享： 20 位参与者中有 18 位提到，当他们与共同创作伙伴不在同一地点时，会采用异步文件共享来进行协作式歌曲创作。例如，参与者 P5 和 P6 表示：“通常我会录好自己的音轨，发给搭档，然后他们之后再加上自己的部分。这不是实时协作——我们依靠后期编辑。”此外，参与者表示，鉴于当前的技术状况，他们很难想象还能以其他方式完成这件事。半数参与者还明确表示，如果能有一款“音乐编辑版 Google 文档”就太好了，但他们并不知道有这样的技术。

2）通过视频通话进行协作：一半的参与者表示，他们会使用视频会议应用程序来共同创作歌曲。P19 表示：“有视频通话工具，但延迟表现不够好。不过，我相信我们还是能完成一些创作的。”我们还观察到，延迟是参与者们主要关心的问题：所有参与者都提到了低延迟对于远程音乐协作的重要性，但只有1人（P20）提到自己曾使用低延迟服务进行音乐协作。不过，P19、P7和P8提到，有时他们觉得需要一些时间独自构思，不与他人分享；而在某些情况下，近实时音乐协作并不理想。P7表示：“只合并最好的部分。”

*3)*手机录音与对便捷性的优先考虑： 20 位参与者中有 8 位表示，在录音过程的初期，他们主要使用手机录制音乐。这些手机录音会分享给乐队成员，或留作个人回忆音乐创意之用。许多人提到，便捷性是最重要的因素之一。例如，P3 表示：“计算机很笨重，手机很方便。”捕捉音乐灵感的瞬间时，便捷和速度至关重要，P15 甚至创造了一个词来描述他们对音乐设备的要求：“应该能舒服地窝在沙发上使用。”

*4)*使用大屏幕进行录音、混音和母带处理： 所有参与者都表示，要对录音进行超出草稿阶段的编辑，就需要使用计算机或平板电脑。相比之下，移动设备屏幕足以用于起草或捕捉初现的音乐创意。P9 是一位来自台湾的专业作曲家，他表示，自己工作的歌曲创作公司要求员工除了使用 iPad 外，还要使用两台计算机和多个屏幕。P9 表示，作曲时至少需要两个屏幕：“我需要两个屏幕，才能更高效地工作。”平板电脑可作为计算机的延伸，用来显示虚拟旋钮和滑块。与此同时，第二台计算机运行不同的操作系统，以提供专用于音色库的资源。

5) 根据计算机摆放位置构建工作流程：20 位参与者中有 17 位表示，录音时会将计算设备直接放在自己面前。这表明计算机、平板电脑或移动设备在录音过程中处于核心位置。主要使用计算机的参与者表示，他们需要根据鼠标和键盘的使用方式调整设备摆放位置。P20 明确表示：“我想乐器最好放在左边……这样我就能同时用右手操控鼠标点击。”他们表示，这种摆放位置不太舒服，但考虑到技术需求，这是最合适的选择。使用平板电脑或手机的参与者则表示，他们需要用专用乐谱架、桌子或其他平面来放置设备。”这些额外的平面也因此成为工作流程的一部分，对高效录制音乐至关重要。

6) 传统乐谱记谱法的使用： 一半的参与者还表示，他们在很大程度上依赖传统乐谱记谱法。有些人会先用传统记谱法创作音乐，之后再将特定文件类型转移到数字音频工作站；另一些人则几乎完全使用传统记谱法进行练习和创作。5 名参与者在初始阶段完全不使用软件

进行创作，而是完全依靠手写记谱法来构思想法。他们表示，这样做是为了提高速度并进行草稿记录。例如，P11 表示：“我可能会先把它写下来，等它更完整之后，再进行录音。”还有几位参与者表示，他们担心如果不及时记下想法，就会忘记：“我会忘记。我需要先把它写下来，这样之后才能编排。”

C. 关于混合现实数字音频工作站设计探针的思考

在研究的这一阶段，参与者两人一组，在混合现实中使用我们的设计探针进行协作。参与者通过踩下脚踏板上的“开始录音”和“停止录音”按钮来录制音频。按下“停止录音”按钮后，录制的片段便会循环播放，模拟音乐家常用循环踏板的功能，让他们能够即兴创作独奏作品。两位参与者都能够录音，并为集体创作作出贡献。

录音过程中，界面会显示当前存在的音轨，参与者可以随时停止或启动音轨（见图 1C）。这使双方能够同时调整作品，并通过手势和踩踏脚踏板来录制和呈现创作过程，从而营造出共同创作的体验。

显示音轨的用户界面面板可以调整大小并在环境中移动，让用户能够自由地定制和重新排列个人数字音频工作站配置，同时操作相同的相对虚拟对象。使用手和脚与虚拟对象交互时，参与者通过 Ableton 内置的应用程序接口操作本地版本的 Ableton Live12。因此，两位参与者共同在 Ableton 中创作，并在混合现实中对作品进行渲染和交互。

参与者分处不同房间，以体现网络协作创作的过程。尽管为了测试而采用直接有线连接，以尽可能降低延迟并简化系统，但考虑到当前网络音频系统具备近实时音频能力，在一定距离范围内（约1000公里以内），这些过程是可以实现的。

*1)*设计探针与当前录音技术的可用性比较： 参与者使用 MR-DAW 系统后，我们请他们完成一份简短的 NASA-TLX 可用性问卷，旨在了解 MRDAW 系统的可用性与其当前工作流程及技术使用情况相比如何。这有助于我们了解参与者如何看待 MR-DAW 的功能，以及它是否能提升可用性（见图4）。

V. NASA-TLX 工作负荷评估

为评估参与者对虚拟现实数字音频工作站系统的主观工作负荷，我们在参与者使用该系统和当前工作流程后，向他们发放了 NASA-TLX 问卷。本分析旨在量化虚拟现实数字音频工作站系统相较于现有

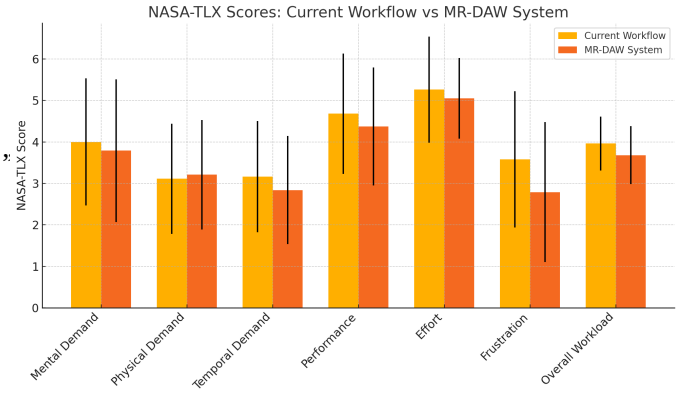

图4. 当前工作流程与 MR-DAW 系统的 NASA-TLX 子量表得分比较。每个柱形表示一个子量表的平均得分，误差线表示标准差。得分越低表示工作负荷越低。

技术和实践的可用性。本分析的结果见图4。

结果显示，使用 MR-DAW系统时，认知工作负荷呈下降趋势。MR-DAW系统的平均总体工作负荷得分较低（M =3.68, SD = 0.70），低于当前工作流程（M =3.96, SD = 0.65）。虽然这一差异在统计学上并不显著， t(18) = 1.74, p = 0.10，但得分持续偏低这一模式表明，可用性有所提升。

考察各个子量表后发现，工作负荷降幅最明显（尽管在统计学上并不显著）的领域是挫折感 （平均值从 3.58 降至 2.79）和时间需求 （从 3.16 降至 2.84）。这表明，参与者可能认为 MR-DAW系统不那么令人烦躁，任务节奏也没有那么紧迫。总的来说，虽然这些发现尚无定论，但表明 MR-DAW系统有望通过降低任务中感知到的努力程度和挫折感，带来切实的可用性益处。

A. MR-DAW的未来展望

第二次访谈要求参与者讨论他们使用该系统的体验。我们讨论了哪些系统组件会引发问题、令人愉悦，以及哪些组件对于未来系统来说是理想的。随后，我们开始讨论未来混合现实数字音频工作站的设计构想。我们询问了参与者认为需要克服的关键挑战、应当保留或扩展的组件，以及未来系统中值得加入的功能。我们还讨论了未来系统的可能性及其用例，概述了新的交互将带来哪些功能，以及它们可能如何改变当前工作流程。

本节归纳出的主题如下：

1) 混合现实中的空间控制增强： 20 位参与者中有 16 位明确提到了，与传统数字音频工作站相比，混合现实数字音频工作站所带来的空间自由度更高。他们提到，能够完全按照自己的意愿布置录音室，并且

随时重新配置，会带来自由感。P20 提到，这会改变他们在录音室中的工作方式，因为他们可以使用混合现实，“以一种在计算机上绝不可能实现的方式，因为你只有一组输入和输出。”；而 P15 则表示：“我在家工作时总觉得特别压抑，就像笔记本电脑摆在那里，我怎么都觉得不自在。我觉得虚拟现实的优势在于，它就像一整台戴在头上的计算机。”

P19 和 P20 是从事现场演出和录音的音乐人，他们提到，可以将大部分硬件转化为虚拟对象和面板，从而节省录音室空间：“在我们的车库里，空间很快就会不够用，对吧？因为这边放着计算机，那边有一个巨大的音频接口，还得管理一大堆接入其中的线缆。很快就会变得一团糟。如果能移除其中几个组件，直接在虚拟现实中使用它们，就能提升移动性、模块化程度和可定制性。”他们还进一步讨论了在这种空间中协作的好处：“如果你有一长串待办事项，而且信任与你合作的人，就可以把其中一半的工作交给对方处理。这种方式在计算机上根本无法实现，因为你只有一组输入和输出接口。”

P19 还提到，他们可以在混合现实中调整屏幕的角度。正如初次访谈的研究结果所述，这通常需要额外配备专用硬件。他表示：“你可以把虚拟用户界面面板放到一些位置，而实体屏幕或设备根本不可能放在那里，但这些位置对录音来说确实非常方便。”

此外，参与者提到，与传统数字音频工作站相比，在混合现实中能够更好地安排音轨和效果：“我可以把所有正在处理的音轨都放在这边。我把效果总线放在这边。我还可以把整体混音器放在这边。而且，如果我正在处理某条特定音轨，我实际上可以把这条音轨放到另一个空间中。

*2)*传感器的替代控制用途：所有参与者都认为，混合现实的运用为控制数字音频工作站开辟了新方法。这些方法包括头部运动追踪、眼动追踪、脑机接口、手部追踪、脚踏控制器和语音命令。其中一些控制机制无需借助混合现实也能实现；不过，参与者提到，得益于较少的空间限制，混合现实为替代控制器提供了比其他设备更大的灵活性。例如，由于虚拟内容可以向各个方向渲染，几位参与者提到可以像指挥家一样启动音轨。P20表示：“你可以调高歌曲中某些部分的音量，就像面前有一支管弦乐团一样。P20还提出：“你能想象把所有不同的音轨都安排在空间中吗？这样就能表现力度。我想，如果大家身处同一个虚拟现实空间，就没有理由不能指挥和带领远在千里之外的人。”

3) 各类乐器在交互方面的特殊性：我们的参与者演奏各种各样的乐器，我们很快发现，有些乐器在混合现实中的交互难度远高于其他乐器。例如，演奏小提琴的参与者必须用颈部固定乐器，同时还要承受虚拟现实头显带来的额外重量，因此很容易感到疲劳。所有吉他手还提到，他们很难看清吉他的琴颈，这对演奏产生了负面影响。

相比之下，歌手可以充分灵活地使用双手与虚拟内容互动。此外，对于不依赖视觉输入的萨克斯演奏者而言，虚拟现实头显造成的视觉畸变并不会对他们产生干扰。

同样，脚踏板的使用也给某些乐器带来了独特的障碍。例如，大提琴演奏时，乐器通常在演奏者拉弓的过程中由其双腿或膝盖固定；而使用脚踏板时，抬腿踩下脚踏板会导致乐器失去稳定。不过，一位参与者建议，脚踏板可以由演奏者进行定制和控制。例如，在循环录制的情况下，20 位参与者中有 10 位建议提供通过语音为初始循环启动倒计时的功能。这样就能释放脚踏板上的可用功能；一位参与者建议，这些功能可以在混合现实中通过用户界面面板实时重新分配和定制。

4) 音频路由方面的考量：20 位参与者中有 6 位明确提到，需要将用于交流和音乐的音频通道分开。P17 和 P18 建议，在与另一位演奏者协作录音时，可以切换音频通道，这样他们就能分别交谈和聆听录音。他们举例说：“我想找的类比有点像分组讨论室。他们并没有离开，但在音频方面，他们只听自己想听的内容，就像独奏一样。”

5) 希望在混合现实中看到面部渲染效果：与音频路由方面的考量类似，20 位参与者中有 8 位提到，希望看到协作者的面部表情。尽管这在混合现实领域仍是一项普遍存在的挑战，但这凸显了面部表情和视觉呈现在混合现实中的重要性。几位参与者提到，只要启用一个显示对方的屏幕，混合现实就能轻松实现这一功能，就像混合现实中的视频通话。P17 甚至表示，如果没有这一功能，会感到自己仿佛孤身一人：“是不是会觉得自己独自待在这里？”

6) 音乐记谱与创意记录的整合：鉴于有相当数量的参与者表示他们会定期使用传统乐谱记谱法，因此许多人建议将其纳入数字音频工作站，这并不令人意外（20人中有8人）。“传统数字音频工作站通常不包含传统乐谱软件，乐谱软件一般都是独立运行的。这就要求使用乐谱软件的音乐家同时运行这两种软件，并持续在程序之间交换文件。P17将这一过程描述为“笨拙”。20人中有6人建议在混合现实中采用自动转录软件，这同样是音乐技术领域持续面临的技术挑战；不过，使用混合现实设备输入传统音乐记谱法的过程也并非简单直接，而采用

自动转录软件可能会大有裨益。

除了乐谱软件，几位参与者还提到，可以通过录制视频或音频来捕捉音乐创意，记录最初的创作灵感。正如首次访谈中所述，音乐家们认为共同打磨音乐创意非常重要，而在混合现实中这样做，就能以多种方式捕捉这些创意。

7) 创造力提升： 6 位参与者提到，我们的系统主要通过协作互动以及引入新颖的外部视角来激发创造力。用户发现，与他人共同创作这一最初的行动，可能会带来意想不到且有趣的音乐方向。P11 指出："我一开始演奏的和弦进行，和我原本想象的听起来完全不同。听听别人如何理解同一段和弦进行，以及他们会在其中加入什么，真是很有意思。"这凸显了系统如何促进想法的动态交流：一个人的贡献会成为激发另一个人灵感的催化剂。正如一位参与者所说，这种相互促进的过程有助于克服个人创作的局限："听到别人做了什么，你突然就会想到自己可以加入什么。这就像两个人之间迸发出创造力。"系统所具备的即时反馈是开启新颖创作路径的有力工具。P5 表示："也许你一开始没有这种创造力，但听到别人的即时反馈后，就会产生新的创意。"此外，系统中即便是简单的视觉提示，也能成为强有力的起点，正如另一位用户所说："我们之前看到的那个树形元素，真的让我们有了创作的动力。"归根结底，这种协作环境不仅能激发新想法，也能带来强烈的成就感。

VI. 讨论

访谈让我们深入了解了音乐家如何在各种音乐实践中使用设备，以及音乐家对未来音乐技术的需求。我们意外地发现，大家对混合现实中的数字音频工作站的未来充满期待，所有参与者都提到，混合现实将在许多方面简化工作流程。现代音乐工作室的形象几乎总是工程师面前摆着一块巨大的屏幕；而在专业音乐创作和编曲工作室中，许多从业者会在工作流程中使用各种屏幕和交互设备。

根据我们的发现，业余音乐家似乎更看重便利性，也受到空间限制。例如，许多专业音乐家在家中设有录音室；然而，这些录音室占用的空间惊人——通常是一整个房间，甚至一整层——业余音乐家难以负担。因此，他们更倾向于使用计算机或平板电脑，但这也带来了自身的交互限制（屏幕尺寸、手指操作、单一的输入和输出机制等）。未来的混合现实数字音频工作站可以优雅地解决这些问题，为业余音乐家提供不受空间限制、便于携带且可定制的音乐工作室。

此外，通过运用混合现实进行协作作曲，并纳入前文提到的几项功能改进，协作可以得到极大提升。首先，需要特别考虑音频路由。有人提醒我们，如果一边仔细聆听，一边又在播放录音，音频可能会变得混乱。我们认为，这需要进一步研究，以更好地了解如何与他人一边演奏和录音，一边进行交流。此外，能够看到协作者也有助于双方沟通。我们还从以往的研究中了解到 [8]‚[26], [27], 网络化视觉信息对延迟的容忍度更高，因此，加入视频可能不会给音乐协作带来问题。

不过，正如近期 [28]所强调的，要实现混合现实中的低延迟协作式数字音频工作站这一愿景，仍需解决一些关键技术问题。这些问题包括如何在元宇宙系统中有效集成网络音乐表演系统，以及如何改进硬件和软件技术，以更高效地采集、处理和传输音乐信号（包括音频和 MIDI）。

因此，我们呼吁专注于音乐元宇宙的研究人员考虑测试并构建混合现实数字音频工作站的组件；这项工作注定会遇到许多意想不到的障碍。例如，许多音乐家描述了如何使用替代控制器来控制音乐元素。我们认为这是一个值得深入探索的领域：站立并在演奏时使用混合现实，可以解放双手和双脚，并自然地将传感器、眼动追踪和前置摄像头引入头部。这些功能以及更多可能性，必将为控制音乐元素开辟新途径，以便录制音乐并与他人分享新作品。

尽管我们的比较可用性结果并未显示设计探针（混合现实数字音频工作站）与用户当前工作流程之间存在显著差异，但混合现实数字音频工作站的可用性呈现出更高的趋势。我们认为，这是因为用户有时间适应并构建自己的工作流程，而我们让他们学习并使用一个未经优化的新系统进行测试。该探针的目的仅仅是展示概念验证，并引发参与者之间的思考与讨论。为此，我们认为，未来经过优化且可定制的混合现实数字音频工作站界面，将比当前工作流程更具可用性。

此外，我们认为混合现实技术有许多机会可以促进各种音乐活动。例如，学习新音乐时，音乐家通常会在书桌前读音乐记谱或观看视频 [2],，而乐器则以别扭的姿势放在他们面前。混合现实数字音频工作站能让许多音乐家采用自然的演奏姿势，并可随心所欲地将音乐记谱或视频流放置在任意位置，同时继续进行音乐练习。遗憾的是，仍有许多技术问题亟待解决，例如从某些注视角度观看时会出现视觉畸变，参与者提到的吉他琴颈变形便是一个例子。此外，还需要考虑如何呈现虚拟对象，使演奏者能够

清楚地同时看见自己的手以及虚拟和真实物体——这是参与者提出的一个严重问题，也是混合现实特有的问题。

值得注意的是，我们的研究存在一些局限性。首先，我们在评估环节中仅纳入了10对音乐家。增加参与者人数或许能提高研究结果的普遍性。

其次，我们在不同房间开展用户研究，以模拟远程环境，但两台控制器向 Ableton 发送的控制消息均通过本地主机上的开放声音控制协议发送至同一台机器。因此，我们没有测试能够远距离处理控制消息的基础设施，限制了我们评估系统如何应对哪怕极小延迟、以及参与者会如何反应的能力。我们认为，鉴于我们采用的方法，这是构建近实时系统可能面临的最大障碍。为了让系统顺畅运行，必须优化在 Ableton 中触发操作的控制消息。这项工作超出了当前研究的范围，因为我们认为，提供无延迟体验是让参与者能够设想这种推测性技术的首要任务。后续研究将进一步揭示混合现实中联网数字音频工作站的能力与局限性。因此，我们还建议开展一项研究，比较此类系统中具身组件的作用，例如将其他 MIDI设备和控制器纳入研究，并与虚拟元素进行比较。

七、结论

本文探讨了混合现实克服数字音频工作站使用障碍的潜力，旨在打造直观的环境，以支持实时远程音乐协作。这些障碍主要包括：数字音频工作站往往会妨碍音乐家的工作流程，将他们束缚在桌前，阻碍他们与乐器进行自然交互；此外，有效的远程协作仍是一项重大挑战，现有解决方案也受到网络延迟和异步文件共享的制约。

我们的研究采用推测性设计方法，探究音乐元宇宙中协作式混合现实数字音频工作站的益处与潜在问题。为促进讨论，我们提出混合现实数字音频工作站，这是一种混合现实系统，可让身处不同地理位置的多位用户在各自的本地空间自由移动，同时控制同一个共享的数字音频工作站实例。这样的联网系统使每位远程音乐家都能使用实体脚踏板进行协作循环录制，将熟悉的免手操作交互与共享虚拟会话相结合。

我们通过半结构化访谈和用户研究评估了该系统，研究涉及 10 对音乐家。我们分析了当前的实践，报告了用户与我们提出的混合现实系统交互时的体验，并探讨了混合现实中音乐协作的未来。研究结果突显了混合现实在实现不受束缚的音乐交互方面的优势，并对音乐元宇宙中远程协作式数字音频工作站的未来进行了展望。

致谢

我们感谢由欧盟委员会的 EIC Pathfinder Open 计划资助的 MUSMET 项目提供支持（资助协议编号：101184379）。然而，文中表达的观点和意见仅代表作者本人，并不一定反映欧盟或欧洲创新理事会的观点。欧盟和欧洲创新理事会均不对此承担责任。

参考文献

[1] M. Marrington，《创作型歌手手册》中的“使用数字音频工作站进行创作”，第77–89页，2017年。[2] J. P. M. Avila、C. Greenhalgh、A. Hazzard、S. Benford 和 A. Cham- berlain，“受限交互：音乐家准备演出的研究”，载于*2019*年计算机系统中的人为因素*CHI*会议论文集，2019年，第1–13页。[3] F. Goltz，“Ableton Link——一种同步音乐软件的技术”，载于*Linux*音频会议，2018年，第39–42页。[4] A. Barate、L. A. Ludovico、G. Presti 等，“面向年轻学习者的协作式数字音频工作站”，载于第*14*届国际计算机辅助教育会议论文集，2022年，第458–464页。[5] S. Stickland、R. Athauda 和 N. Scott，“可扩展实时在线数字音频工作站协作框架的设计与评估”，《音频工程学会期刊》，第69卷，第6期，第410–431页，2021年。[6] L. Turchet、R. Hamilton 和 A. C. amci，“扩展现实中的音乐”，*IEEE Access*，第9卷，第15 810–15832页，2021年。[7] T. Hopkins、S. C. C. Weng、R. Vanukuru、S. Novack、C. Tobin、E. Wenzel、A. Banic、M. Gross 和 E. Y.-L. Do，“XR Jam：在扩展现实中利用人工智能进行音乐联网的设计考量”，载于*2024*年*IEEE*人工智能、扩展现实与虚拟现实国际会议（*AIxVR*）。IEEE，2024年，第97–101页。[8] T. Hopkins、S. C. C. Weng、R. Vanukuru、E. A. Wenzel、A. Banic、M. D. Gross 和 E. Y.-L. Do，“AR鼓圈：增强现实中的实时协作式击鼓”，《虚拟现实前沿》，第3卷，2022年。[9] L. Turchet、M. Carraro 和 M. Tomasetti，“FreesoundVR：利用在线声音库在虚拟现实中进行声景创作”，《虚拟现实》，第27卷，第2期，第903–915页，2023年。[10] A. Cassorla、G. Kearney、A. Hunt、H. Riaz、M. Stiles 和 D. Murphy，“利用智能手机进行基于数字音频工作站的空间音频创作的增强现实”，载于音频工程学会第*148*届年会。音频工程学会，2020年。[11] L. Men 和 N. Bryan-Kinns，“Lemo：支持虚拟现实中的协作式音乐创作”，载于*IEEE*虚拟环境声音交互研讨会，2018年。[12] L. Turchet，“音乐元宇宙：愿景、机遇与挑战”，《个人与普适计算》，第1–17页，2023年。[13] C. Rottondi、C. Chafe、C. Allocchio 和 A. Sarti，“网络化音乐表演技术概述”，*IEEE Access*，第4卷，第8823–8843页，2016年。[14] D. Jones，“AbletonOSC：Ableton Live 的统一控制应用程序接口”，载于音乐表达新界面国际会议论文集，*Miguel Ortiz* 和 *Adnan Marquez-Borbon*（编）。墨西哥墨西哥城，文章，第60卷，第5期，2023年。[15] J. Auger，“思辨设计：构建思辨”，数字创意，第24卷，第1期，第11–35页，2013年。[16] A. Dunne 和 F. Raby，《思辨万物》，作者新增序言：设计、虚构与社会梦想。麻省理工学院出版社，2024年。[17] G. Ouzounian、C. Haworth 和 P. Bennett，“思辨设计：迈向社会音乐”，2017年。[18] J. Wallace、J. McCarthy、P. C. Wright 和 P. Olivier，“让设计探针发挥作用”，载于计算机系统中的人为因素*SIGCHI*会议论文集，2013年，第3441–3450页。[19] K. Tahıro˘glu、T. Magnusson、A. Parkinson、I. Garrelfs 和 A. Tanaka，“作为探针的数字乐器：计算如何改变乐器的存在方式”，《有组织的声音》，第25卷，第1期，第64–74页，2020年。

[20] K. E. Fisher、K. Yefimova 和 A. P. Bishop，“将设计思维和文化探针应用于移民青年的经历：揭示视觉媒体和音乐在信息通信技术导航中的作用”，载于*2016*年计算机系统中的人为因素*CHI*会议扩展摘要论文集，2016年，第859–871页。[21] O. Badreddin，“扎根理论在软件工程中的应用：主题回顾与分析”，《软件工程进展》，第2013卷，第1期，第468021页，2013年。[22] T. Hopkins、A. Jude、G. Phillips 和 E. Y.-L. Do，“Virtual AI Jam：面向人机协同音乐即兴创作的人工智能驱动虚拟音乐家”，*AIMC 2023*，2023年8月29日，https://aimc2023.pubpub.org/pub/9064kmvw。[23] G. Terry、N. Hayfield、V. Clarke、V. Braun 等，“主题分析”，《*SAGE*心理学定性研究手册》，第2卷，第17–37期，第25页，2017年。[24] T. Hopkins、R. Sun、S. C. C. Weng、S.-Y. Ma、J. Crum、L. Hirshfield 和 E. Y.-L. Do，“Brain Jam：用于与人工智能驱动的虚拟音乐家开展击鼓协作的神经信号感知自适应系统”，《计算机科学前沿》，第7卷，第1570249页。[25] U. Kuckartz 和 S. R¨adiker，《使用*MAXQDA*分析定性数据》。施普林格，2019年。[26] T. Hopkins、S. Weng、R. Vanukuru、E. Wenzel、A. Banic 和 E. Do，“多晚才算太晚？网络延迟对增强现实远程音乐协作期间视听感知的影响”，载于*2022*年*IEEE*虚拟现实与三维用户界面会议摘要与研讨会。IEEE，2022年，第686–687页。[27] T. Hopkins、S. C.-C. Weng、R. Vanukuru、E. Wenzel、A. Banic、M. D. Gross 和 E. Y.-L. Do，“研究网络延迟对增强现实音乐任务中视听感知的影响”，载于*2022*年*IEEE*混合现实与增强现实国际研讨会（*ISMAR*），2022年，第26–34页。[28] A. Boem、M. Tomasetti 和 L. Turchet，“音乐元宇宙中的音频技术：问题与挑战”，《音频工程学会期刊》，第73卷，第94–114页，2025年。