arXiv:2303.01457v1 [cs.HC] 2 Mar 2023

作为作曲家、声音设计师和创意媒体制作人之间中介的人工智能

SEBASTIAN LÖBBERS，数字音乐中心，伦敦玛丽女王大学，英国MATHIEU BARTHET，数字音乐中心，伦敦玛丽女王大学，英国GYÖRGY FAZEKAS，数字音乐中心，伦敦玛丽女王大学，英国

为非音乐领域的利益相关者制作素材的音乐专业人士，在早期构思阶段常常面临沟通挑战。表达音乐理念可能很困难，尤其是在缺乏领域专用词汇时。本文立场论文提出运用人工智能促进利益相关者之间的沟通，并加快共识构建过程。其目的并非全部或部分自动化创作过程，而是通过减少用于明确预期成果的时间，为创意留出更多时间。为阐明这一观点，本文探讨了两种交互式音乐系统的应用场景，这些系统基于作者关于手势到声音映射的研究。

CCS 概念：• **应用计算** → **声音与音乐计算**；• **以人为中心的计算** → 用户界面编程；协作与社交计算设备。

A其他关键词和短语：手势到声音的映射、人工智能中介的协作、人际构思

ACM 引用格式：
Sebastian Löbbers、Mathieu Barthet 和 György Fazekas。2023年。作为中介的人工智能：作曲家、声音设计师与创意媒体制作人之间的协作。载于*CHI '23* 研讨会：在人际协作构思中整合人工智能。ACM，纽约，纽约州，美国，4 页。

1 引言

作曲家和声音设计师经常与不具备专业音乐知识的人合作，或为他们工作。例如，作曲家可能会与电影制作人或编舞家合作，创作服务于叙事的原创配乐；声音设计师则可能受雇于某家公司，制作符合其企业形象的音频品牌塑造。在这两种情况下，双方很可能会在早期探索阶段提出并分享想法，以更准确地界定声音或音乐应当表达什么。一个潜在的瓶颈在于双方音乐知识不平衡。人们对声音和音乐可能有强烈的看法和偏好，但当被问及背后的原因时，非音乐人士与音乐家所使用的语义描述各不相同 [9]。与非专业人士相比，音乐家往往依赖专业术语而非日常语言，这些术语对于作曲家或声音设计师来说可能难以理解。专业语言（例如“利底亚调式的乐曲”）虽然能向领域专家表达想法，却令其他利益相关者难以理解。在某些情况下，人们可能更倾向于使用手势或身体动作等非语言沟通形式，例如制作人描述空间音频时 [2]。在实践中，人们会分享能够体现预期想法、并可在构思过程中作为参考的声音和音乐素材，以克服这些困难。然而，这又带来了两个新问题：（1）如何找到这些素材？利益相关者可能会利用自己对现有音乐作品的了解（“内部”音乐目录），但在寻找特定声音时，这种做法就不那么可行了

允许免费制作本作品全部或部分内容的数字或纸质副本，用于个人或课堂用途，前提是副本不得为盈利或商业利益而制作或分发，且副本须附有本声明及首页上的完整引用信息。对于本作品中由 ACM 以外方拥有的部分，其版权必须得到尊重。注明出处后可以摘录。如需以其他方式复制、重新出版、发布到服务器或向列表重新分发，必须事先获得特定许可和/或支付费用。请通过 permissions@acm.org 申请许可。
© 2023年 美国计算机协会。稿件已提交至 ACM

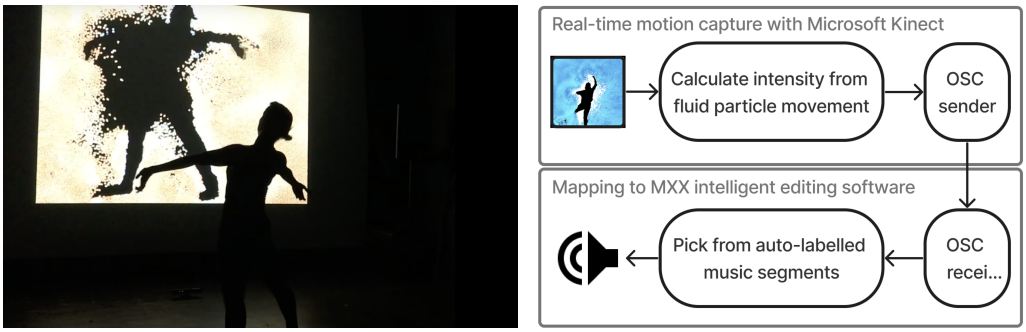
；（2）即使心中已有参考，作曲家或声音设计师创作的素材最终也可能不符合任务要求。如何快速调整想法以用于原型制作，例如让创作的音乐适配电影或编舞，从而减少“白费”的工作?

人工智能（AI）的最新进展为音乐专业人士带来了新的机遇，例如通过文本提示生成音频 [1], 创作特定音乐流派的吉他六线谱 [8], 智能地进行声音可视化 [4] 或从图形草图中检索音频 [3] 以及手势 [10]。然而，人工智能研究往往侧重量化指标的优化，例如音频质量或分类准确率，而非探索其在“现实世界”应用中的作用。本文作者回顾了自己运用人工智能实现手势到声音的映射的工作，并设想了两种有助于人际音乐与声音构思的潜在应用。论文将在下一节简要介绍这些应用，并在第在 人际协作构思中整合人工智能研讨会的背景下，于第3节进行讨论。

2 案例研究

本节介绍两个交互式音乐系统，*Liquid Dance* 和*SketchSynth*，并通过两个假设性案例研究，探讨它们在协作式音乐与声音构思方面的能力。这些案例研究基于第一作者作为职业作曲家和声音设计师的经验。

2.1 Liquid Dance



(a) 视频静帧 (b) 流程图

图 1. *LiquidDance* 概览。演示视频见 https://youtu.be/fYInCkaOZO8。

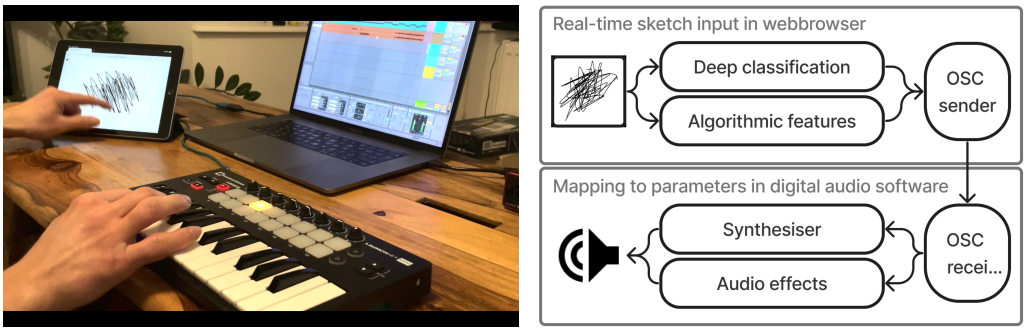
如图1所示，*LiquidDance* 利用动作捕捉和流体粒子模拟来量化舞蹈动作的强度，并自动组合音乐片段，实时生成合适的配乐。该项目与一位编舞家合作开发，旨在为音乐人工智能初创公司 MXX[1] 通过交互式演出展示其智能音乐编辑软件。后来，一个与此无关的舞蹈作曲项目激发了人们将*Liquid Dance* 用于快速测试编舞音乐构想的想法。以下情境将对此进行介绍：

一位编舞家和一位作曲家挑选了几首参考音乐作品，现在想测试它们

[1]https://www.mxx.ai/

与编舞舞蹈表演的契合度。为此，他们计划使用一个系统，让他们能够伴随音乐作品起舞，并自动挑选出与舞蹈动作强度最匹配的片段。通过这种方法，他们希望更好地了解所选音乐如何与编舞相得益彰。

2.2 SketchSynth



（a）视频静帧　　（b）流程图

图 2. *SketchSynth* 概览。演示视频见 https://youtu.be/arSFt3iBAUM。

SketchSynth 基于作者对音乐音色跨模态感知的研究 [5–7]。它运用人工智能和算法方法，根据图形草图输入预测声音，如图2所示。以下场景描述了其在工业声音设计中的一种潜在应用：

一位声音设计师和一位产品经理正在寻找适合融入其产品的声音。为简化这一过程，他们计划使用 *SketchSynth*，让非专业人士也能通过直观的交互方式表达想法并探索声音空间。双方可以直接绘出想法，直至达成共识，之后声音设计师便可进一步详细完善。

3 讨论与结论

在这两项案例研究中，人工智能都充当非语言表达与声音或音乐创作之间的翻译者。尽管人类专家很可能能够根据这些表达创作出合适的内容，但机器经过训练后，可以更快地实时完成这项任务。这让利益相关者能够更快地测试初步想法，将精力集中在人际关系上，并为创作过程留出更多时间；而创作过程可以说最适合由人类专业人士完成。在此背景下，人工智能被严格视为一种工具，预期它能够可靠地完成复杂但定义明确的任务。*Liquid Dance* 采用了一套稳健但非常具体的方案，可能只适用于范围狭窄的项目；*SketchSynth* 则试图寻找通用的草图到声音的映射，但这些映射可能过于宽泛，无法充分细致地传达声音。由于用户需求可能在项目之间甚至项目进行期间发生变化，人工智能工具的适用范围或许也需要随之调整。这不仅引出了人工智能在人际构思过程中如何使用的问题，也引出了如何训练和配置人工智能的问题。一种做法是选择在多样化数据集上训练、能够解决大量项目问题的模型。然而，这可能会让渡过多自主性，导致系统出现不适合当前任务的意外行为。因此，用户可能希望在初始的“简报”阶段调整系统，为即将到来的任务做好准备：

以上述论文中的案例研究为例：对于*SketchSynth*，这可能意味着限制可能生成的声音范围，例如为电动汽车项目限定为机械音效，或为手机游戏限定为数字提示音；对于*LiquidDance*，则可能意味着针对特定舞蹈风格（如芭蕾舞或现代舞）优化界面，并根据特定音乐流派（如古典音乐或电子音乐）调整人工智能驱动的音乐分段。归根结底，用户需要决定希望系统在构思过程中展现多少自主性。总之，近期的人工智能驱动的音乐研究在辅助人际构思方面具有重要的应用价值，例如可充当非语言声音和音乐描述的翻译者。然而，人工智能工具或许需要根据用户需求和项目要求进行调整，以避免出现不可预测或不恰当的行为。

4 致谢

EPSRC 和 AHRC 媒体与艺术技术博士培训中心 (EP/L01632X/1)。

参考文献

[1] Andrea Agostinelli、Timo I Denk、Zalán Borsos、Jesse Engel、Mauro Verzetti、Antoine Caillon、Qingqing Huang、Aren Jansen、Adam Roberts、Marco Tagliasacchi等。2023年。MusicLM：根据文本生成音乐。arXiv 预印本 arXiv:2301.11325（2023年）。[2] Thomas Deacon、Patrick Healey和Mathieu Barthet。2022年。“确实更简洁”：音频制作中的协作过程。计算机支持的协同工作（CSCW）（2022年），1–31。[3] Lars Engeln、Nhat Long Le、Matthew McGinity和Rainer Groh。2021年。基于视觉草图的声音搜索的相似性分析。载于 Audio Mostly 2021 论文集。美国计算机协会，意大利特伦托，101–108。https://doi.org/10.1145/3478384.3478423[4] Leandro Garber、T Ciccola和JC Amusategui。2021年。AudioStellar：一种开源的基于语料库的音乐工具，用于潜在声音结构发现和声音实验。载于国际计算机音乐会议论文集。密歇根大学出版社服务部，智利圣地亚哥，62–67。https://hdl.handle.net/2027/fulcrum.t435gg568[5] Sebastian Löbbers、Mathieu Barthet和György Fazekas。2021年。勾勒声音：一项关于声音与形状关联的探索性研究。载于国际计算机音乐会议论文集。密歇根大学出版社服务部，智利圣地亚哥，299–304。https://hdl.handle.net/2027/fulcrum.t435gg568[6] Sebastian Löbbers和György Fazekas。2022年。看见声音，听见形状：一项评估声音草图的游戏化研究。载于国际计算机音乐会议论文集。密歇根大学出版社服务部，爱尔兰利默里克，174–179。https://hdl.handle.net/2027/fulcrum.nk322g689[7] Sebastian Löbbers和György Fazekas。2023年。SketchSynth：声音合成的跨模态控制。EvoMUSART 预印本（2023年）。https: //sebastianlobbers.com/static/76293e97d37a3d69941f2f1b047cec7e/SketchSynth_跨模态_控制_的_声音_合成.pdf[8] Pedro Sarmento、Adarsh Kumar、Yu-Hua Chen、CJ Carr、Zack Zukowski和Mathieu Barthet。2023年。GTR-CTRL：利用 Transformer 实现吉他音乐生成的乐器与流派条件控制。arXiv 预印本 arXiv:2302.05393（2023年）。[9] Simin Yang、Courtney Nicole Reed、Elaine Chew和Mathieu Barthet。2021年。考察现场音乐表演中的情绪感知一致性。IEEE 情感计算汇刊（2021年）。[10] Michael Zbyszyński、Balandino Di Donato、Federico Ghelli Visi和Atau Tanaka。2021年。手势-音色空间：使用机器学习进行多维特征映射。载于国际计算机音乐多学科研究研讨会论文集。施普林格，日本东京，600–622。https://doi.org/10.1007/978-3-030-70210-6_39

收到日期：2023年2月22日