arXiv:2304.03407v2 [cs.HC] 8 Sep 2023

音频工程协会
大会论文
发表于第154届大会，2023年5月13日
至15日，埃斯波，赫尔辛基，芬兰

本文作为完整稿件经过同行评审，并获准在本次大会上发表。本文可在音频工程协会电子图书馆（*http://www.aes.org/e-lib*）查阅，版权所有。未经《音频工程学会期刊》直接许可，不得复制本文或其任何部分。

# 在音乐混音工作流程中采用人工智能技术：一项调查

Soumya Sai Vanka[1], Maryam Safi[2], Jean-Baptiste Rolland[2],和 György Fazekas[1]

[1]伦敦玛丽女王大学，英国伦敦[2]*Steinberg Media Technologies GmbH*，德国汉堡
通信请联系 Soumya Sai Vanka (s.s.vanka@qmul.ac.uk)

**摘要**
人工智能（AI）技术融入音乐产业，正在显著改变音乐创作、制作和混音的方式。本研究考察人工智能在混音工作流程中的现状及不同用户群体对其的采用情况。通过半结构化访谈、基于问卷的研究以及对网络论坛的分析，本研究确认了三类用户群体：业余爱好者、半专业人士和专业人士。研究结果表明，人工智能混音工具能够简化流程，并为业余爱好者提供尚可的效果；半专业人士则寻求精确的控制和定制选项，而专业人士除了需要控制和定制选项外，还希望获得辅助性和协作技术。本研究提出了面向不同用户群体设计有效人工智能混音工具的策略，并指出了未来的发展方向。

## 1 引言

近年来，人工智能辅助音乐制作领域发展显著，涌现出越来越多旨在自动化制作流程各个环节的工具和技术。然而，尽管这些工具具有潜在优势，许多音乐制作人和工程师仍怀疑它们能否达到人类专家的水平。这种怀疑源于多种因素，包括担心被取代 [1, 2], 怀疑机器能否像人类一样胜任主观性任务 [3, 4, 5], 以及由于不了解模型输出背后的推理过程而不信任人工智能的建议 [6, 7]。这些原因构成了用户在工作流程中采用此类工具的一大障碍。

在本研究中，我们考察了使用人工智能技术的不同用户群体的态度和认知混音工作流程中的技术，并描述各类用户的使用场景，以及他们对人工智能工具的期望。为此，我们对专业音频工程师（专业人士）进行了深入访谈，随后向专业人士和半专业人士发放问卷。我们还对多个互联网论坛上的讨论进行了定性评估，以更全面地了解更广泛用户群体（包括初学者和业余爱好者）对这些工具的情感倾向。我们相信，本研究的发现将推动开发适用于混音工作流程的基于人工智能的工具，专门满足各用户群体的需求和具体要求，并符合他们的期望。

## 2 背景

### 2.1 音乐混音

音乐混音是将多条音频轨道融合在一起，形成连贯的最终作品，从而引发情感共鸣并讲述故事的过程 [8]。混音工程师的工作是处理原始音频录音，通过技术和创意上的调整，使其符合艺术家的构想，并制作出精致、连贯的最终混音。这包括突出关键元素、平衡乐器声部以及塑造独特音色等任务。除了增益分级，还会使用均衡、声像定位、压缩和调制效果等技术来塑造声音 [9]。混音完成后，作品会交由母带工程师进一步处理，之后即可进行商业发行。

### 2.2 音乐制作的民主化

在数字技术出现之前，录制和制作音乐需要昂贵设备以及专业技能，这些条件大多数音乐人都难以获得。这意味着，拥有经济资源和基础设施支持的人，更有机会使用专业录音室并发行自己的音乐。然而，随着个人电脑、互联网和数字音频工作站（数字音频工作站）的出现，录制、制作和发行音乐的成本显著降低，使其能够被更广泛的用户群体使用，包括音乐人 [10]。传统成本和技能壁垒的消除，使更多音乐人/艺术家/制作人得以进入该行业，为这些工具创造了更加多元的用户群体。下一节将介绍现有文献中确定的三类用户 [11, 12, 13]。

#### 2.2.1 业余爱好者

在混音工程领域，业余者是指刚接触这一领域，可能没有接受过正规培训或积累过混音经验的人。他们可能是业余爱好者，也可能是刚刚起步、正在学习音乐混音基础知识的音乐人 [14]。他们通常技能较为生疏，也不会通过音乐混音获得报酬。如果他们只是偶尔运用混音技能，可能会长期留在这一类别中。

#### 2.2.2 半专业人士

社会学家罗伯特·斯特宾斯创造了“现代业余者”一词，用来指那些有技能、专注且充满热情，但不同于专业人士和业余爱好者的人 [15]。近来关于业余主义的讨论已认识到社会变迁和互联网对业余主义这一概念的影响，并引入了“Pro-Am”一词，描述那些认真投入某一领域、按照专业标准开展工作，却得不到专业人士所享有的同等基础设施支持的人 [16]。Pro-Am 被视为介于成熟的专业人士与传统业余者之间的一个连续谱 [17]。他们可能热衷于创作和混音音乐，并希望提升技能，但未必直接以混音音乐为生，也可能不会将其作为自己在该领域的主要重心。

#### 2.2.3 专业人士

专业混音工程师是指在音乐混音方面技艺高超，并通过提供服务获得报酬的个人。他们可能接受过正规教育，或通过在传统工作室环境中进行正式学徒培训或实习获得默会知识 [13]。他们在音乐产业中拥有良好的人脉，并获得基础设施方面的支持，能够负担设备和设施。他们还被期望不断提升和更新自己的知识与技能，以跟上其所在领域的最新发展。他们被视为混音方面的专家，其表现标准高于业余爱好者。

### 2.3 用于混音工作流程的智能工具

1975年，Dan Dugan在一次音频工程协会会议上介绍了一种自动混音系统，该系统专注于麦克风混音，以减少反馈 [18]。此后，人们广泛研究了如何利用基于知识、心理声学和基于感知的方法来构建辅助混音工具和自动混音系统 [19, 20]。随着机器学习和人工智能在计算机视觉及其他领域取得成功，这些技术也被应用于音频领域，包括为音乐混音构建智能工具。这些工具涵盖了从接收分轨并提供混音 [21, 22, 23, 24, 25, 26, 27, 28, 29]的混音系统，到能够模拟其模拟／数字对应物并通过分析所提供的音频直接进行处理的基于人工智能的插件

此前也曾提出过采用基于规则的处理方式的插件 [30, 31, 32, 33, 34, 35, 36,37, 38, 39]。

智能工具或基于人工智能的工具旨在利用人工智能领域最先进的统计软件技术，为混音工作流程中的各种任务提供解决方案。开发用于音乐混音的智能插件和工具，旨在自动化混音过程中的技术性及非创意性任务，从而减少混制一首歌曲所需的时间和技能 [41]。此外，此类工具旨在提高混音过程的便捷性和效率，让用户能将更多时间用于音乐制作的创意环节 [19, 42]。

近年来，市场上已推出多种用于音乐混音工作流程的商业化人工智能工具。为便于讨论，我们将适用于混音工作流程的智能工具分为两类。

- **自动混音**：自动混音可定义为任何能够根据分轨或原始音频创建混音的系统。换言之，自动混音是一个更广泛的术语，涵盖了利用人工智能自动化整个混音过程。RoEx[2] 和 iZotope Neutrone[1] 分别提供自动混音服务和工具。

- **辅助混音工具**：辅助混音工具运用人工智能技术来辅助混音工作流程。这些工具包括基于人工智能的插件和处理器，例如智能均衡器、智能混响、智能人声推子等。多家插件公司均提供此类工具，包括 iZotope1、Sonible3、Focusrite4 等。

然而，正如 [41] 中所述，这些工具能否在市场上取得成功取决于三个因素：

1. **生成结果的精确度与质量**：基于人工智能模型生成的音频必须具备高质量，才能供专业人士用于其工作流程。然而，当前的人工智能模型仍常常难以生成高质量的结果，因此专业工作流程对采用这些模型存在抵触。

2. **无缝集成到现有工作流程中**：专业人士通常会经过多年反复调整，在实践中形成最适合提升效率和生产力的工作流程。为了更好地接受这些工具，将其无缝集成到各自现有的工作流程中至关重要 [43]。

3. **有助于建立信任的交互模型**：人们对基于人工智能的工具最大的顾虑，源于它们被视为缺乏可解释性的“黑箱” [44, 45]。对于许多专业音频工程师而言，无法通过调整来实现理想结果，往往也是一大弊端 [46, 47]。

在本研究中，我们旨在通过识别用户群体并分析其使用案例、期望和情感倾向，了解基于人工智能的工具在音乐混音工作流程中的应用情况。我们通过半结构化访谈、基于问卷的研究以及互联网论坛分析收集数据，并进行定性分析。本文呈现并分析研究结果，随后讨论基于人工智能的混音工具设计未来可能的发展方向。最后，研究总结了研究发现并提出相关的未来发展方向。

## 3 方法论

为了兼顾混音技能水平各异的个人所持有的不同观点和见解，我们采用了三种数据收集方式，即结构化、半结构化和非结构化方式，顺序不限 [48]。文献 [49] 已详细阐述了每种方法的优势。这些方法在所得信息方面在一定程度上互为补充。

### 3.1 对专业人士进行半结构化访谈

专家知识可用于构建世界模型 [50]。从解释性建构主义的视角来看，这种观点或许过于简单，但对于开展访谈、收集信息而言却很有用。首先，我们对业内专业的混音工程师进行了半结构化访谈。为避免选择性这一对研究有效性至关重要的问题，我们进行了多次访谈（五次） [51]。由于本研究阶段的目标对象是专家，他们的人数远少于非专家用户，

[2]https://www.roexaudio.com/[3] https://www.sonible.com/[4] https://focusrite.com/en

而且难以联系，因此我们根据第 2.2 节所述标准，从熟人中亲自确定了五位专家参与者，并邀请他们参加一对一访谈 [52]。所有专家工程师从事混音工作均已超过五年，且其技能被视为达到高级水平。他们为自己和其他艺术家制作音乐混音。这些混音工程师接受过正规培训，其工作有报酬，作品也以传统唱片发行形式及音乐流媒体平台上的形式发表。开展访谈的主要目的是为初步构建理论提供助力。访谈重点了解参与的工程师在混音工作流程中对人工智能技术的使用情况、期望和情感倾向。访谈通过在线视频通话进行。通话内容均被录制并转录。我们对访谈记录进行了主题分析，以归纳主题，详见第 3.3 节。

### 3.2 基于问卷的专业人士和半专业人士结构化研究

结构化且标准化的提问方式能够使不同参与者的回答具有可比性 [49]。我们的目标是基于从访谈中形成的主题，获取可量化的数据，并在更大范围的混音工程师群体中检验其一致性。因此，我们依据第2.2节所述的标准，对受邀的专业及专业业余混音工程师开展了一项基于问卷的结构化研究。该研究涉及 22 名参与者，这些工程师认为自己的混音技能处于高级至专家水平，并且拥有五年以上的经验。问卷由多项选择题和自由形式的长答题组成。

#### 3.2.1 来自互联网论坛的人工智能混音与母带处理数据

此外，为拓展研究结果的视野，我们从 Reddit、Quora 和 Twitter 等多个互联网论坛的讨论中抓取了文本数据，这些讨论专门聚焦于用于混音和母带处理流程的基于人工智能的工具 [53, 48]。我们收集了近六个不同讨论串的数据，涉及所有讨论串中约三十名参与者。由于我们寻找的是这一小众领域的近期讨论，因此只能找到数量有限的讨论串。此外，在许多这些讨论中，参与者的身份资料并不清楚。我们假定参与者要么是业余爱好者，要么是半专业人士。

### 3.3 数据分析

不同研究收集的数据兼具定性和定量性质。因此，我们采用混合方法，根据定性和定量研究方法对数据进行评估。半结构式访谈经过录音并转录。我们运用扎根理论分析方法，对转录文本进行了定性主题分析 [54, 55] ，包括对数据进行编码、定制编码系统，以及以迭代和归纳的方式构建理论。我们还采用类似的归纳式、迭代式和扎根式方法，对从问卷和论坛中收集的所有定性数据进行了主题分析。定性分析使用 MAXQDA[5] 软件完成。MAXQDA 专为学术、科研和商业机构中的计算机辅助定性与混合方法数据、文本及多媒体分析而设计。我们首先生成了一系列“代码”，用于识别人工智能混音工具用户的不同子群体。经过多轮迭代，我们归纳出他们使用这些工具的场景以及对这些工具的情感倾向。问卷中的多项选择题采用数值和统计方法进行了可视化处理，在许多情况下，这些结果对定性研究发现起到了补充和强化作用。研究结果将在以下各节中讨论。

## 4 结果

如图1所示，我们的研究证实，基于人工智能的混音工具与三类不同用户群体均相关：业余人士、半专业人士和专业人士。我们发现，每个用户群体对智能工具都有独特的需求，并在其工作流程中将这些工具用于各种目的。我们将三项研究的结果放在一起讨论。我们引用了半结构式访谈和问卷中的节选，标记为 ’I’，以及互联网论坛中的节选，标记为 ’F’。

### 4.1 业余爱好者

我们观察到，许多自称业余爱好者的论坛参与者主要是创作和编曲的音乐人。他们表示自己对音乐混音了解有限

[5] https://www.maxqda.com/

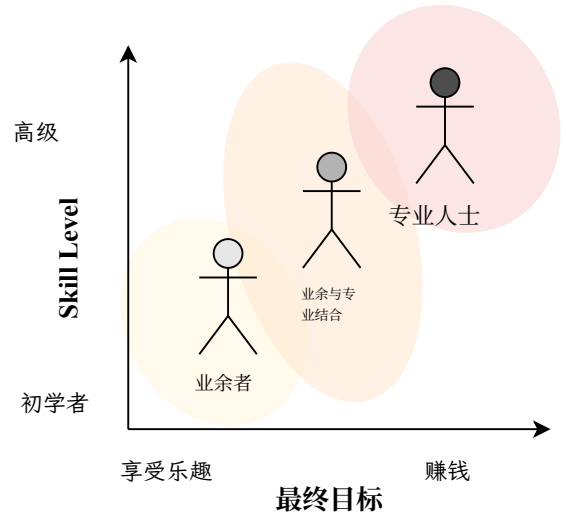


**图1：**智能工具的用户类别

，他们的主要目标是与全世界分享自己的音乐。

“阻碍我将录音作品发布到世界各地的最大障碍，总是混音过程。这很乏味……给我一个可以按下的按钮，让我能够自动完成大部分工作。”

几位业余爱好者指出，混音是他们发布音乐时遇到的最大障碍。他们也承认，只要混音效果足够好，他们就满意了。

“……给初学者做出一个还不错的混音。我并不要求听起来百分之百完美，但只要他们能帮上忙，对我来说就足够了。”

全自动混音工具对这类人更有帮助，因为他们通常没有时间或资源去投资昂贵设备，或聘请专业工程师。这些工具可以为音乐爱好者提供一种经济实惠的解决方案，他们的主要目标是创作和分享音乐。

“对我来说，把混音或母带处理做到九成听得过去，就足以发行音乐，也能让我专注于自己喜欢的事情。”[F]

音乐制作行业的业余爱好者将人工智能驱动的工具视为节省时间和精力的方式，认为无需付出太多努力，也不必深入了解混音的技术方面，就能制作出不错的混音。

由于对相关主题了解有限，他们期待这些系统具有高度自主性，几乎就像“一键施法”的神奇按钮。他们对这些人工智能系统制作出的混音质量也抱有较低的期望。

“我喜欢录制自己的歌曲，但我实在弄不懂许多与混音有关的概念，一碰到均衡器、压缩等内容就很容易摸不着头脑。我基本上算是个业余者；我尊重混音这门艺术，但做到这一步时，制作音乐的乐趣就消失了。”[F]

总的来说，我们观察到业余爱好者正积极拥抱这项新兴技术。

#### 4.2 半专业人士

我们的观察证实，半专业人士或半专业用户的技术技能水平高于业余爱好者，但其经验和专业知识可能仍不及专业人士。此外，他们通常不会通过这项技艺赚钱，而是出于兴趣创作和混音音乐。他们以与业余爱好者相似的方式使用智能混音工具，但也将其作为提升技能、努力成为专业混音工程师的途径。半专业人士还会使用智能混音工具，以便在混音中快速实现特定的声音或风格。

“它可以用作学习基础知识的工具，然后结合参考法以及反复试验，成为学习工具箱中的又一种工具。人工智能用于打基础，其他方法则用来制作更精彩的混音。”[F]

鉴于自身较高的专业水平，半专业人士很清楚这项技术的局限性，但也愿意尽可能充分地发挥现有智能工具的潜力。他们对这些工具的未来持谨慎乐观态度。

“这些人工智能混音工具都不会替你完成音乐混音。它们只会对你可以采取哪些操作来更好地处理音频提出一些基本建议。你自己仍然必须承担大部分决策和工作。”[F]

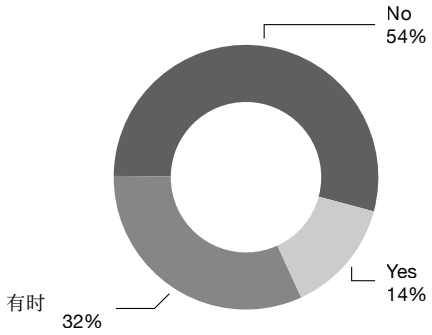


**图 2：**半专业人士和专业人士对在混音工作流程中使用人工智能工具的反馈。

### 4.3 专业人士

谈到该领域的专家，专业音频工程师对在混音工作流程中使用人工智能持不同看法。一些专业人士使用智能混音工具来改善工作流程，并节省处理重复性任务的时间。

“在混音唱片或制作唱片时，如果有智能工具能替我们完成一些枯燥的任务，我们该如何最有效地利用时间？”[I]

他们可能会使用这些工具辅助混音过程，但最终仍依靠自己的专业知识和经验来创作精良的最终作品。他们认为，这些工具能为混音提供一个起点和方向。专业人士也可能用这些工具尝试新的音色和风格，或是在混音中快速实现某种特定的音色或风格。

“如果运气好，你也许能在混音时想到五个绝妙的创意，而系统可能会给你带来一百个好点子，供你挑选。”[I]

“我用过 *Neutron* 的掩蔽功能来提升清晰度，但这只是制作过程中可能涉及的成千上万种操作中的一种。”[I]

他们期望这些工具具有高度的准确性、精确性，并且能够根据自身需求进行定制。他们当然支持辅助性和协同创作技术，这些技术能够实现协作：机器提出建议，工程师随后可以根据自己的喜好调整并微调设置。

“……你知道，将你的混音与参考歌曲中的情况进行比较，并建议低频过多、高频过多，这个那个也太多了。”[我]

“机器也可以进行处理，然后把控制权交给我，让我调整所有内容。这比从头开始要有帮助得多。”[I]

我们的调查发现，专业人士目前正在使用基于人工智能的智能工具执行滤波、峰值检测、音高检测、母带处理、均衡和声音增强等任务。

然而，正如图 2 所示，我们研究中的许多专业音频工程师表示，他们不会在混音工作流程中使用智能人工智能工具。这些专家认为，人工智能工具无法完全取代混音过程中所需的人为参与和创造力。他们认为，这些工具通过计算方法生成结果，却无法理解有意进行音乐混音所需的“感受”和“情感”。他们认为，由人工智能控制的混音可能会导致最终成品失去控制力和精准度。

“作为业余爱好，在家里这样做没问题。但在制作环境中的专业层面，这行不通，因为客户和公司的修改要求很高，而且对细节的要求非常精准。这不是为了做得更快，而是为了做到精准。”[F]

另一位专家指出，这些工具在公众舆论中往往受到质疑，因为它们提供了一种“按下按钮就能创造奇迹的解决方案”，来处理原本需要投入大量时间和精力才能掌握的技能；这可能会让许多专业人士多年学习技艺所付出的时间和努力显得毫无价值。专家群体中仍普遍认为，传统混音方法更胜一筹，而通过反复试验来学习才是掌握混音艺术的最佳途径。

总之，一些专业人士看到了人工智能改善其工作流程的潜力，但归根结底，这取决于个人偏好和工作流程。图 3 展示了专业业余人士和专业人士所报告的、在混音工作流程中对不同类别人工智能工具的使用偏好。我们看到，相比于直接对给定录音应用音频效果的工具，人们更能接受自动混音。这可能是因为这些自动音频效果工具通常无法理解给定分轨中的上下文

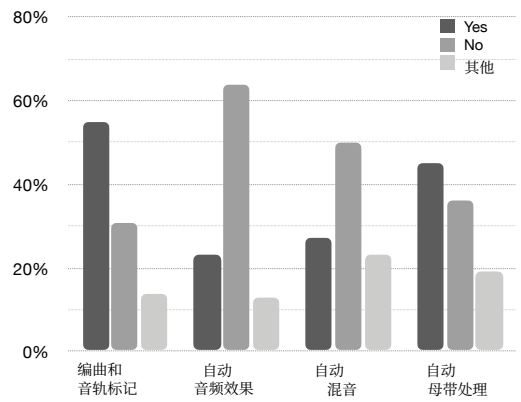


**图3：** 半专业人士和专业人士报告的混音工作流程中对不同人工智能工具使用的偏好

，混音的性质非常通用。我们还观察到，人们对标记工具和自动母带处理工具的接受度更高。许多专业人士确认，他们通常会使用自动母带处理工具，在将混音草稿分享给客户之前，调整其整体动态和响度。他们还解释说，在混音之前将多轨音频加载到数字音频工作站中，对分轨进行标记和颜色编码，都是重复性工作，而且非常耗时。

## 5 讨论

“混音无疑是一种艺术形式，但人工智能有很多方面可以帮助我们，这也是不可避免的下一步。”[F]

大多数业余爱好者都希望与全世界分享自己的音乐。然而，成功商业发行的瓶颈往往在于混音。提升混音技能或聘请专业音频工程师通常花费不菲。这类用户可能会使用自动混音服务来制作他们认为足以用于商业发行的混音。业余爱好者也可能会使用具有纠正功能、支持智能编辑的辅助混音工具 [56]。这类用户对精细控制和定制的兴趣较低。

另一方面，半专业人士对最终成果的质量有更明确的目标，也更为挑剔；他们可能为自己或他人制作并混音音乐。半专业人士涵盖了业余爱好者与专业人士之间的所有用户。由于混音并非他们的主要收入来源，因此他们更倾向于探索和学习。这类用户会在工作流程中使用基于人工智能的工具；不过，由于他们的批判性聆听技能较强，他们会寻找精准度更高、能够提供更多控制权的工具。他们将人工智能混音工具用作起点或获取灵感，但需要对最终混音拥有更多控制权。

最后，对专业人士而言，混音本身就是一种艺术形式。他们愿意花更多钱购买自己认为有用、且能帮助他们加快技术性和重复性任务处理速度的工具。不过，他们需要大量的控制和定制选项。表 2 列出了专业人士认为能够简化其混音工作流程的各类智能工具及其任务。还应指出，与能够提供一定控制的工具相比，专家对全自动混音工具的热情较低。这可能是因为自动化混音常被视为控制有限的“黑箱”。通过访谈，我们发现大多数专业人士都愿意在工作流程中使用辅助技术和协作技术。

“.....但我认为，计算机系统与人类互动的地方，才是真正有趣之处。”[我]

我们在表1中汇总了不同用户群体使用智能工具进行混音的用例和期望。

### 5.1 混音智能技术的未来

本研究的关键结论是，每个用户群体都有不同的目标和需求，因此他们可能需要能够满足其特定目标和需求的工具。人工智能在混音工作流程中的未来发展和应用取决于工具的设计目标。为了让更广泛的用户接受这些工具，重要的是以满足第2.3 节中提到的三个因素的方式来构建这些工具。在此，我们希望简要概述一些可能的方向。

#### 5.1.1 控制与自动化的平衡

在人机交互（HCI） [57, 58]领域，控制与自动化的适当程度一直备受争议。人工智能系统常被视为黑箱，给出的结果往往让用户缺乏控制权，也难以理解。我们的研究结果表明，不同用户群体对自动化和控制的需求程度各不相同。为此，我们提出设计能够提供

| 智能<br>工具\用户 | 业余爱好者 | 半专业人士 | 专业人士 |
|---|---|---|---|
| 使用案例 | 制作出不错的混音 | 学习与探索<br>tool | 自动化重复且耗时的工作。<br>耗时任务 |
| | 作为学习工具 | 寻找一个起点 | 共同创作与辅助 |
| | | 自动化技术任务 | 寻找起点/方向<br>用于混音 |
| | | 创造力与灵感 | 创造力与灵感 |
| 期望 | 更自主，控制较少<br>控制 | 更高级，控制更多 | 高度先进且涵盖广泛范围<br>的控制选项 |
| | 经济高效的 | 准确且精确 | 准确且精确 |
| | 易于使用 | 辅助式 | 辅助性 |
| | | 成本效益高 | 轻松融入当前工作流程-<br>flow |
| | | 轻松集成到当前<br>工作流程 | 上下文感知 |
| 情感倾向 | 积极 | 谨慎乐观 | 混音 |

**表1：** 混音工作流程中不同人工智能技术用户类别的使用案例、期望和情感倾向比较

不同程度控制与自动化的人工智能工具，并根据用户的专业水平进行调整。这可以通过让同一款插件提供不同模式来实现，包括完全自动化、有条件自动化、部分自动化、辅助模式和无自动化 [41]。这样，用户便可根据可用时间和所需精度，让同一款插件/工具胜任多种用途。已有研究沿着这一方向开展，旨在设计可解释且能让最终用户拥有更多控制权的基于人工智能的混音系统 [27]。

#### 5.1.2 无缝集成

专业混音工程师通常拥有一套成熟的工作流程和熟悉的工具，帮助他们高效地开展工作。因此，我们需要构建格式和配置与这些用户熟悉的内容相似的工具和技术。这意味着要将研究成果转化为用户熟悉、易于获取和操作，并能顺利融入现有工作流程的产品。例如，可以将经过训练、能够执行自动均衡的模型整合为插件，方便用户在数字音频工作站中加载。这里提到的呼吁】【。用于创建开发者工具，以更快速、更轻松地将研究成果转化为技术。这可以通过使用 Neutone[6], JUCE[7], CMajor[8],iPlug2[9], 和 RTNeural[10] 等框架和软件开发工具包（SDK）来实现。这样一来，研究人员就能更快地从创作者那里获得声反馈，从而不断迭代，创建出更好的工具，并更清晰地了解用户需求。另一种潜在方法是让人工智能学会控制数字音频工作站。这意味着，给定项目会话中的分轨后，系统能够预测各分轨的增益电平、声像值以及带参数的音频效果链，从而完成混音。这可能有助于业余爱好者和预算有限的音乐人改善其制作成果，而专业人士则可以利用它节省时间，创建一个之后还能进一步完善的样板混音。这也为重新建模和重新设计现有数字音频工作站打开了大门

[6]https://neutone.space/[7]https://juce.com/[8]https://cmajor.dev/[9]https://iplug2.github.io/[10]https://github.com/jatinchowdhury18/RTNeural

| 阶段 | 类别 | 任务 |
|---|---|---|
| **混音前** | 混音准备******* | 为音轨添加标签并进行颜色编码*** |
| | | 分组和编曲音轨 |
| | | 导入会话数据 |
| | | 设置发送通道、AUX 和总线** |
| | 编辑 | 裁剪静音部分 |
| | | 识别并修复咔嗒声和爆音 |
| | | 识别并修复相位问题 |
| | | 检查单声道兼容性 |
| **混音期间** | 电平平衡 | 增益分级** |
| | | 设置粗略的推子混音*** |
| | 频谱校正 | 识别并修复掩蔽问题** |
| | | 自动高通滤波器，用于处理低频内容的内容 |
| | | 自动均衡和均衡匹配*** |
| | 沟通 | 及时获取艺人对时间戳的反馈 |
| | 协同创造性** | 为混音提出想法** |
| | | 音频效果处理建议 |

**表 2：**混音工程师希望用于简化其工作流程的智能功能列表；对该工具的需求频率以星号（*）的数量表示

以便利用消费级笔记本电脑和专业混音控制台中可用嵌入式硬件不断增强的性能 [59]。

#### 5.1.3 上下文感知系统

我们发现，几位半专业人士和专业人士都承认，基于人工智能的工具十分通用，对特殊情况的处理效果不佳。他们还表示，这些工具往往只分析给定的音轨，却不了解混音中的上下文，从而降低了人们对这些工具的信任和信心。

“人工智能无法有意义地理解你的歌曲需要什么，也无法领会它的氛围。它只能提供泛泛的结果。”[F]

这为开发具备上下文感知能力的系统提供了可能性，这些系统经过更多样化的音乐和音频数据训练，能够很好地适用于各种风格、流派以及通常是该领域创新标志的细分特征。

还可以通过允许用户使用文本 [61], 音频 [29], 和语义描述符 [62, 63, 64]查询这些人工智能系统，让它们更充分地理解混音构想。近期研究展示了基于文本、旋律和语义令牌的音乐生成方面的先进表现。我们认为，可以进一步探索这些系统，并针对音频工程应用进行微调。

具备情境感知能力的工具凭借其带来的制作便利，有望为音乐产业吸引新人才、新流派并促进多元化。业余爱好者或许能更轻松地创作和分享音乐，从而有机会凭借自己的技艺获得认可。这些工具也可能帮助半专业人士提升技能，并练习专业混音。总体而言，这些工具有望改进混音工作流程，提高生产效率和创造力。

## 6 总结

在混音工作流程中使用人工智能是一个复杂且细致的话题。有人认为它可能带来颠覆性变革，也有人对其能否取代混音过程中所需的人类参与和创造力持更怀疑的态度。在本研究中，我们确定了人工智能混音工作流程工具的三类用户。基于从他们那里获得的认识，我们分析并报告了各类用户的情感倾向和期望。我们认为，这些信息有助于为后续开发和应用人工智能工具奠定基础，使其能够简化各类用户的工作。

总而言之，我们相信，本研究所提供的用户特定洞见将帮助研究人员构建更易被音频工程界接受并融入其中的人工智能混音技术。除此之外，人工智能工具还可以变得更加精准、灵活，并能够适应不同的音乐流派和风格，同时处理范围更广的输入音频。另一种可能性是开发能够理解并响应用户创作意图的人工智能，从而在音乐制作过程中实现人类与机器之间更加直观、高效的协作。人工智能与音乐领域正在不断发展，而未来的可能性仅受当前技术水平和研究人员想象力的限制。

## 7 致谢

我们谨向 AES 欧洲大会的评审人员致以诚挚的谢意，感谢他们为我们的工作提供了宝贵的声反馈。对于所有参与访谈和研究并与我们分享宝贵知识的混音工程师及参与者所作出的贡献，我们深表感谢。同时，我们也感谢 Steinberg 的研发团队在整个项目期间始终给予我们的支持，并提供坦诚的声反馈。

此外，我们感谢同事 Christian Steinmetz、Gary Bromham 和 AngelikiMourgela 的宝贵贡献。他们围绕这一主题进行了富有见地的讨论，并提供了关键的声反馈，助力完善了我们的研究。

我们也感谢英国研究与创新署提供的财政支持， [资助编号 EP/S022694/1]。这项资助对推动我们的研究并使我们能够为该领域作出贡献发挥了重要作用。

## 参考文献

[1] Campolo, A., Sanfilippo,M., Whittaker, M., and Crawford, K.,《*AI Now 2017* 报告》, 纽约大学 AI Now 研究所 , 2017.

[2] Sturm, B.L., Iglesias, M., Ben-Tal, O., Miron, M., and Gómez, E., “人工智能与音乐：版权法和工程实践中的开放问题,” 载于《*Arts*》，第 8 卷，第 115 页，MDPI，2019。

[3] Jordanous, A., “评估评估：计算创造力研究中的进展与实践,” 《计算创造力：自主创造系统的哲学与工程》，第 211–236 页，2019。[4] Ritchie, G., “创造性系统的评估,”《计算创造力：自主创造系统的哲学与工程》，第 159–194 页，2019。[5] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J., “构建像人一样学习和思考的机器,” 《行为与脑科学》，40，第 e253 页，2017。[6] Ribera, M. and Lapedriza, A., “我们能否提供更好的解释？一项以用户为中心的可解释人工智能方案。”载于《*IUI* 研讨会》，第 2327 卷，第 38 页，2019。[7] Roy, Q., Zhang, F., and Vogel, D., “自动化的准确性固然重要，但较高的可控性或许更重要,” 载于 《*2019* 年 *CHI* 计算机系统中的人为因素会议论文集》，第 1–8 页，2019。[8] Miller, M.,《音乐混音》，企鹅出版社，2016。[9] Izhaki, R.,《音频混音：概念、实践与工具》，劳特利奇出版社，2017。

[10] Hracs, B. J., “转型中的创意产业：数字驱动型独立音乐制作的崛起”，*Growth and Change*，43(3)，第442– 461页，2012年。

[11] Sandler, M., De Roure, D., Benford, S., and Page, K., “用于音乐制作与消费链各环节新体验的语义网技术”，载于*2019 International Work-shop on Multilayer MusicRepresentation andProcessing(MMRP)*，第49–55页，IEEE，2019年。

[12] McGrath, S., Chamberlain, A., 和 Benford, S., “共同创作音乐：对业余及半职业 grime 音乐制作的探索”，载于《*Audio Mostly 2016* 会议论文集》，第186–193页，2016年。

[13] Bromham, G., “学术实践如何塑造混音技艺？”载于音乐混音，第265–276页，劳特利奇出版社，2016年。[14] Hoare, M., Benford, S., Jones, R. 和 Milic-Frayling, N., “走出边缘：网络时代的业余音乐人”，载于*SIGCHI*人机交互系统中的人为因素会议论文集，第1295–1304页，2014年。[15] Stebbins, R. A., “业余者：两种社会学定义”，《太平洋社会学评论》，20(4)，第582–606页，1977年。[16] Prior, N., “重新审视新业余者：流行音乐、数字技术与文化生产的命运”，载于劳特利奇文化社会学手册，第340–348页，劳特利奇出版社，2018年。[17] Leadbeater, C. 和 Miller, P., 业余专家革命：爱好者如何改变我们的社会与经济，伦敦 Demos，2004年。[18] Dugan, D., “自动麦克风混音”，载于音频工程协会第*51*届大会，音频工程协会，1975年。[19] De Man, B., Reiss, J. D. 和 Stables, R., “自动混音十年”，载于第三届智能音乐制作研讨会，2017年。[20] Steinmetz, C. J., Vanka, S. S., Martínez Ramírez, M. A. 和 Bromham, G., 用于自动混音的深度学习，ISMIR，2022年。[21] Tom, A., Reiss, J. D. 和 Depalle, P., “一种基于消除掩蔽和最佳声像摆位实践的立体声多轨空间化自动混音系统”，载于音频工程协会第*146*届大会，音频工程协会，2019年。

[22] Moffat, D. 和 Sandler, M., “鼓组多轨增益混音的机器学习”，载于音频工程协会第*147*届大会，2019年。

[23] Martínez-Ramírez, M. A., Stoller, D., 和 Mof- fat, D., “利用 Wave-U-Net 实现智能鼓组混音的深度学习方法”，《音频工程学会期刊》，2021年。

[24] Steinmetz, C. J., Pons, J., Pascual, S. 和 Serrà, J.，“使用可微分神经音频效果混音台进行自动多轨音频混音”，载于*ICASSP*，IEEE，2021。[25] Colonel, J. T. 和 Reiss, J.，“通过可微分数字信号处理对录音混音进行逆向工程”，《美国声学学会期刊》，150(1)，第 608–619 页，2021。[26] De Man, B. 和 Reiss, J. D.，“知识工程驱动的自主混音系统”，载于第 *135* 届音频工程协会大会，音频工程协会，2013。[27] Steinmetz, C. J.，《在波形域中使用神经音频效果进行混音学习》，硕士论文，庞培法布拉大学，2020，doi:10. 5281/zenodo.4091203。[28] Martínez-Ramírez, M. A., Liao, W.-H., Fabbro, G., Uhlich, S., Nagashima, C. 和 Mitsufuji, Y.，“利用深度学习和域外数据实现自动音乐混音”，载于*ISMIR*，2022。[29] Koo, J., Martínez-Ramírez, M. A., Liao, W.-H., Uhlich, S., Lee, K. 和 Mitsufuji, Y.，“音乐混音风格迁移：一种用于解耦音频效果的对比学习方法”，*arXiv* 预印本 *arXiv:2211.02247*，2022。[30] Martínez-Ramírez, M. A. 和 Reiss, J. D.，“使用卷积神经网络进行端到端均衡”，载于第 *21* 届国际数字音频效果会议（*DAFx-18*），2018。[31] Sheng, D. 和 Fazekas, G.，“一种用于动态范围压缩器智能控制的特征学习孪生模型”，载于 *2019* 年国际神经网络联合会议（*IJCNN*），第 1–8 页，IEEE，2019。[32] Singh, S., Bromham, G., Sheng, D. 和 Fazekas, G.，“动态

范围压缩器的智能控制方法：一项用户研究”，《音频工程学会期刊》，69(7/8)，第 576–585 页，2021，doi:10.17743/jaes.2021.0028。

[33] Ma, Z., De Man, B., Pestana, P. D., Black, D. A., and Reiss,J. D.,“智能多轨动态范围压缩”，音频工程协会期刊，63(6)，第 412–426 页，2015 年。

[34] Steinmetz, C. J.and Reiss, J. D.,“用于实时建模模拟动态范围压缩的高效神经网络”，载于音频工程协会第 *151* 届大会，音频工程协会，2022 年。

[35] Kuznetsov, B., Parker, J. D. 和 Esqueda, F.，“面向机器学习应用的可微 IIR 滤波器”，载于国际数字音频效果会议（*eDAFx-20*）论文集，第 297–303 页，2020 年。[36] Colonel, J. T., Steinmetz, C. J., Michelen, M. 和 Reiss, J. D.，“通过抽样随机多项式，利用深度学习直接设计双二阶滤波器级联”，载于*2022* 年 *IEEE* 声学、语音与信号处理国际会议（*ICASSP*），第 3104–3108 页，IEEE，2022 年。[37]Nercessian, S.，“使用可微双二阶滤波器进行神经参数均衡器匹配”，载于国际数字音频效果会议（*eDAFx-20*）论文集，第 265–272 页，2020 年。[38] Colonel, J. T., Comunità, M. 和 Reiss, J.，“利用可微波形整形器对无记忆失真效果进行逆向工程”，载于第 *153* 届音频工程协会大会，音频工程协会，2022 年。[39] Lee, S., Choi, H.-S. 和 Lee, K.，“可微人工混响”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 30 卷，第 2541–2556 页，2022 年。[40] Wilmering, T., Fazekas, G. 和 Sandler, M. B.，“用于控制多轨音频自适应音效的高级语义元数据”，发表于音频工程协会第 *133* 届大会，美国旧金山，2012 年。[41] Tsiros, A. 和 Palladini, A.，“面向人工智能辅助音乐制作的以人为本设计框架”，载于国际会议论文集音乐表现新界面会议，第 399–404 页，2020 年。

[42] De Man, B.,Stables, R., and Reiss,J. D., 智能音乐制作，劳特利奇出版社，2019 年。

[43]Long, D., Padiyath, A., Teachey, A., and Magerko, B.,“协作、创造力与具身性在人工智能学习体验中的作用”，载于创造力与认知，第 1–10 页，2021 年。

[44] Linardatos, P., Papastefanopoulos, V., 和 Kot- siantis, S., “可解释人工智能：机器学习可解释性方法综述”，熵，23(1)，第 18 页，2020 年。[45] Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., 和 Kagal, L., “解释解释：机器学习可解释性概述”，载于*2018 IEEE* 第五届数据科学与高级分析国际会议（*DSAA*），第 80–89 页，IEEE，2018 年。[46] Sterne, J. 和 Razlogova, E., “情境中的机器学习，或从 LANDR 学习：人工智能与音乐母带处理的平台化”，社交媒体+与社会，5(2)，第 2056305119847525 页，2019 年。[47] Scurto, H. 和 Bevilacqua, F., “通过人类与人工智能协作借鉴音乐计算实践”，载于音乐信息学日（*JIM 2018*），2018 年。[48] Seale, C., Charteris-Black, J., MacFarlane, A., 和 McPherson, A., “访谈与互联网论坛：两种定性数据来源的比较”，定性健康研究，20(5)，第 595–606 页，2010 年。[49]Brinkmann, S., “非结构化与半结构化访谈”，《牛津定性研究手册》，2，第 277–299 页，2014 年。[50] Shaw, M. L. 和 Woodward, J. B., “专家知识建模”，知识获取，2(3)，第 179–206 页，1990 年。[51] Bogner, A., Littig, B., 和 Menz, W.,《专家访谈》，Springer Fachmedien Wiesbaden，2006 年。

[52] Baker, S. 和 Edwards, R., “定性访谈需要多少次才够？关于定性研究中抽样与案例的专家见解及早期职业生涯反思”，国家研究方法中心评论论文，2017 年。

[53] Bickart, B. 和 Schindler, R.M.，“互联网论坛作为具有影响力的消费者信息来源”，《互动营销期刊》，15(3)，第 31–40 页，2001 年。

[54] Braun, V. 和 Clarke, V.，主题分析。，第 57–71 页，美国心理学会，2012 年，ISBN 978-1-4338-1003-9。

[55] Braun, V. 和 Clarke, V.，“在心理学中使用主题分析”，《心理学定性研究》，3(2)，第 77–101 页，2006 年。[56] Fazekas, G. 和 Sandler, M.，“借助自动音乐结构提取对录音室录音进行智能编辑”，发表于奥地利维也纳举行的 *AES* 第 *122*届大会，2007 年。[57] Xu, W.、Dainoff, M. J.、Ge, L. 和 Gao, Z.，“向与人工智能系统进行人机交互转变：人机交互专业人士在实现以人为本的人工智能方面面临的新挑战与机遇”，《国际人机交互期刊》，39(3)，第 494–518 页，2023 年。[58] McFarlane, D. C. 和 Latorella, K. A.，“人机交互设计中人为中断的范围与重要性”，《人机交互》，17(1)，第 1–61 页，2002 年。[59] Turchet, L.、Lagrange, M.、Rottondi, C.、Fazekas, G.、Peters, N.、Østergaard, J.、Font, F.、Bäckström, T. 和 Fischione, C.，“声音互联网：趋同趋势、见解与未来方向”，《*IEEE* 物联网期刊》，第 1–1 页，2023 年，doi:10.1109/JIOT.2023.3253602。[60] Lefford, M. N.、Bromham, G.、Fazekas, G. 和 Moffat, D.，“情境感知智能混音系统”，《音频工程协会期刊》，69，第 128–141 页，2021 年，doi:10.17743/jaes. 2020.0043。[61] Agostinelli, A.、Denk, T. I.、Borsos, Z.、Engel, J.、Verzetti, M.、Caillon, A.、Huang, Q.、Jansen, A.、Roberts, A.、Tagliasacchi, M.、Sharifi, M.、Zeghidour, N. 和 Frank, C.，“MusicLM：从文本生成音乐”，2023 年。

[62] Stables, R.、Enderby, S.、DeMan, B.、Fazekas, G. 和 Reiss, J. D.，“SAFE：一种用于提取和检索语义音频描述符的系统”，载于*ISMIR 2014*，2014 年。

[63] Stables, R.、De Man, B.、Enderby, S.、Reiss, J. D.、Fazekas, G. 和 Wilmering, T.，“音乐制作中音色变换的语义描述”，载于 第 *24* 届 *ACM* 国际多媒体会议论文集，第 337–341 页，2016 年。

[64]Turchet, L.、Pauwels, J.、Fischione, C. 和 Fazekas, G.，“云端智能乐器交互：使用智能吉他检索大型音乐合集”，*ACM* 互联网物联网汇刊，1(3)，2020 年，ISSN 2691-1914，doi: 10.1145/3377881。