



Audio Engineering Society

Convention Paper 10653

Presented at the 154th Convention
2023 May 13–15, Espoo, Helsinki, Finland

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Adoption of AI Technology in the Music Mixing Workflow: An Investigation

Soumya Sai Vanka¹, Maryam Safi², Jean-Baptiste Rolland², and György Fazekas¹

¹Queen Mary University of London, London, UK

²Steinberg Media Technologies GmbH, Hamburg, Germany

Correspondence should be addressed to Soumya Sai Vanka (s.s.vanka@qmul.ac.uk)

ABSTRACT

The integration of artificial intelligence (AI) technology in the music industry is driving a significant change in the way music is being composed, produced and mixed. This study investigates the current state of AI in the mixing workflows and its adoption by different user groups. Through semi-structured interviews, a questionnaire-based study, and analyzing web forums, the study confirms three user groups comprising amateurs, pro-ams, and professionals. Our findings show that while AI mixing tools can simplify the process and provide decent results for amateurs, pro-ams seek precise control and customization options, while professionals desire control and customization options in addition to assistive and collaborative technologies. The study provides strategies for designing effective AI mixing tools for different user groups and outlines future directions.

1 Introduction

The field of AI-assisted music production has seen a significant increase in recent years, with a growing number of tools and techniques aimed at automating various aspects of the process. However, despite the potential benefits of these tools, many music producers and engineers remain skeptical of their ability to match the performance of human experts. This skepticism is driven by several factors, including fear of being replaced [1, 2], doubt that machines can perform as well as humans on subjective tasks [3, 4, 5], and distrust of AI recommendations due to a lack of understanding of the underlying reasoning behind the models' outputs [6, 7]. These reasons pose a major bottleneck to the adoption of these tools in users' workflows.

In this investigation, we examine the attitudes and perceptions of various user groups who employ AI tech-

nologies in mixing workflows and describe some use case scenarios and expectations from AI tools for each user type. This is achieved by conducting thorough interviews with professional engineers (pros), followed by a questionnaire with pros and professional-amateurs (pro-ams). We also qualitatively assess the discussions on various internet forums to construct the bigger picture about the sentiments around these tools amongst a wider segment of users including beginners and amateurs. We are confident that the findings of this study will motivate the creation of suitable AI-based tools for mixing workflows that specifically address the needs and specifications of each user group in accordance with their expectations.

2 Background

2.1 Music Mixing

Mixing music is the process of bringing multiple audio tracks together to form a cohesive final product that evokes emotions and tells a story [8]. The job of the mixing engineer is to take the raw audio recordings and perform technical and creative transformations to align them with the artist's vision and create a polished, cohesive final mix. This includes tasks such as highlighting key elements, balancing instrumentation, and creating a unique sound. In addition to gain staging, techniques such as equalization, panning, compression, and modulation effects are used to shape the sound [9]. The mixed music is then passed on to the mastering engineer for further enhancements before it is ready for commercial release.

2.2 Democratisation of Music Production

Before the advent of digital technologies, recording and producing music required expensive equipment and specialised skills that were beyond the reach of most musicians. This meant that those with financial resources and infrastructural support had a better chance to access professional studios and distribute their music. However, with the introduction of personal computers, the internet and digital audio workstations (DAWs), the cost of recording, producing, and distributing music has significantly decreased, making it accessible to a much wider range of users including musicians [10]. The removal of the traditional barriers of cost and skill has allowed a larger number of musicians/artists/producers to enter the industry, creating a more diverse segment of user groups for these tools. The following section introduces the three categories of users identified from existing literature [11, 12, 13].

2.2.1 Amateurs

An amateur in the context of mix engineering is someone who is new to the field and may not have formal training or experience in mixing. They may be hobbyists or musicians who are just starting out and learning the basics of mixing music [14]. They are typically less skilled and do not get paid for mixing music. Users may continue to remain in this category for a prolonged period of time if they only use their mixing skills occasionally.

2.2.2 Pro-Ams

Robert Stebbins, a sociologist coined the term “modern amateur” as someone who is skilled, dedicated and passionate but different from professionals and hobbyists [15]. More recent discussions on amateurism have recognized the impact of societal changes and the internet on the concept of amateurism, and have introduced the term “Pro-Am” to describe individuals who are seriously engaged in a field and work to professional standards but without the same infrastructural support enjoyed by professionals [16]. The Pro-Am is seen as a spectrum between full-fledged professionals and traditional amateurs [17]. They may have a passion for making and mixing music and a desire to improve their skills, but they may not be making a living from mixing music directly or this may not be their main focus in the field.

2.2.3 Professionals

A professional mixing engineer is an individual who is highly skilled in mixing music and is paid for their services. They have possibly received formal education or have tacit knowledge gained from a formal apprenticeship or internship in a traditional studio environment [13]. They have good connections in the music industry and have the infrastructural support to afford equipment and facilities. They are also expected to continuously improve and update their knowledge and skills to stay current in their field. They are considered experts in mixing and are held to a higher standard of performance than amateurs or hobbyists.

2.3 Smart tools for mixing workflows

In 1975, Dan Dugan presented an automatic mixing system at an Audio Engineering Society convention that focused on microphone mixing to reduce feedback [18]. Ever since, there has been extensive research in building assistive mixing tools and automatic mixing systems using knowledge-based, psychoacoustics, and perceptually-motivated approaches [19, 20]. With the success of machine learning and artificial intelligence (AI) in computer vision and other fields, these techniques have been applied to the audio domain including building smart tools for music mixing. These tools have ranged from mixing systems that take in stems and provide a mix [21, 22, 23, 24, 25, 26, 27, 28, 29] to AI-based plugins that emulate their analogue/digital

counterparts and apply processing directly to provided audio by analysing them [30, 31, 32, 33, 34, 35, 36, 37, 38, 39]. Plugins that employ rule-based processing have also been proposed previously [40].

Smart or AI-based tools aim to provide solutions for various tasks in the mixing workflows using state-of-the-art statistical software technology from the field of artificial intelligence. The development of smart plugins and tools for music mixing is motivated by the goal of automating technical and non-creative tasks in the mixing process, thus reducing the time and skills needed to mix a song [41]. Additionally, such tools aim to provide convenience and efficiency in the mixing process, allowing users to spend more time on the creative aspects of music production [19, 42].

Several AI-based commercial tools for music-mixing workflows have been made available in the market in recent years. For the convenience of discussion, we divide smart tools available for mixing workflow into two categories.

- **Automatic mixing:** Automatic mixing can be defined as any system that can create a mix given the stems or raw audio. In other words, Automatic mixing is a broader term that encompasses the use of AI to automate the entire mixing process. RoEx² and iZotope Neutron¹ offer automatic mixing services and tools, respectively.
- **Assistive mixing tools:** Assistive mixing tools use AI technology to assist in mixing workflows. These tools include AI-based plugins and processors such as smart equalizers, smart reverbs, smart vocal riders, etc. Such tools are offered by several plugin companies including iZotope¹, Sonible³, Focuserite⁴, and more.

However, the success of these tools in the marketplace depends on three factors as well explained in [41]:

1. **Precision and quality of the produced results:** Audio generated by AI-based models needs to be of high quality in order to be used by professionals in their workflows. However, current AI-based models still often struggle to generate high-quality output which causes resistance to their adoption in professional workflows.

²<https://www.roexaudio.com/>

³<https://www.sonible.com/>

⁴<https://focusrite.com/en>

2. **Seamless integration into existing workflows:** Professionals tend to develop a workflow in their practice over years of iteration that is best suited to maximise their efficiency and productivity. For the better acceptance of these tools, seamless integration into existing individual workflow is key [43].
3. **Interaction models that facilitate trust:** The biggest inhibition towards the use of AI-based tools is rooted in them being perceived as “black boxes” that lack interpretability [44, 45]. The lack of control for alteration to achieve desirable results is often a caveat for many professional engineers [46, 47].

In this work, we aim to understand the adoption of AI-based tools in music mixing workflows by identifying user groups and analyzing their use cases, expectations, and sentiments. Utilising semi-structured interviews, a questionnaire-based study, and internet forum analysis, we gather data and perform a qualitative analysis. The findings are presented and analysed, followed by a discussion on potential future directions for the design of AI-based mixing tools. The research concludes with a summary of the findings and relevant future directions.

3 Methodology

To counter for the diverse views and insights from individuals with varying skill levels in mixing, we deployed three mechanisms for data collection, namely structured, semi-structured, and unstructured, not in this particular order [48]. The benefits of each of these methods have been described in the literature [49] in detail. Each of the methods are somewhat complementary in terms of information derived from them.

3.1 Semi-Structured Interviews with Pros

Expert knowledge can be used for modelling the world [50]. While this viewpoint may be simplistic from an interpretive-constructive perspective, it is useful for conducting interviews to collect information. To start, we conducted semi-structured interviews with mixing engineers who are professionals in their field. To avoid the problem of selectivity which is very important for the validity of the investigation, multiple interviews (five) were conducted [51]. Since we targeted experts at this stage of the study, who represent a significantly smaller number of users than non-experts,

and who are difficult to reach, five expert participants were personally identified amongst our contacts based on the criteria described in Section 2.2 and invited to participate in individual interviews [52]. All expert engineers have been practising mixing for more than five years, and are considered advanced in their skills. They mix music for themselves and other artists. These mixing engineers have received formal training, get paid for their work, and have had their work published as traditional discographies and on the music streaming platforms. The objective for conducting interviews was mainly to facilitate some preliminary theory-building. The interviews focused on understanding the use, expectations, and sentiments around AI-based technology in mixing workflows amongst the participant engineers. The interviews were conducted via online video calls. These calls were recorded and then transcribed. We performed thematic analysis on the transcripts to develop themes as elaborated in Section 3.3

3.2 Questionnaire-based Structured Study with Pros and Pro-ams

A structured and standardised way of asking questions leads to comparable answers across participants [49]. We aimed to derive quantifiable data based on the themes developed from the interviews and check for agreement amongst a larger pool of mixing engineers. Hence, we conducted a questionnaire-based structured study with invited pro and pro-am mixing engineers based on the criteria described in Section 2.2. This study involved twenty-two participant engineers who considered their skill level in mixing to be advanced to expert and had more than five years of experience. The questionnaire comprised multiple choice questions and free form long answer questions.

3.2.1 Data from Internet Forums on AI-based Mixing and Mastering

Further, to expand the horizons of our results, we scraped textual data from discussions on various internet forums such as Reddit, Quora, and Twitter that were specifically focused on AI-based tools for mixing and mastering workflows [53, 48]. We collected data from nearly six different discussions, involving about thirty participants across all the threads. Since we were looking for recent discussions in this niche field, we could only find a limited number of threads. Also, the participant profiles were not clear in many of

these discussions. We assumed participants to be either amateurs or pro-ams.

3.3 Data Analysis

The collected data from different studies was both qualitative and quantitative in nature. Hence, the data were evaluated based on a mixed method, using both qualitative and quantitative research approaches. The semi-structured interviews were recorded and transcribed. We performed qualitative thematic analysis on the transcripts using grounded theory analyses [54, 55] that involved coding of the data, customisation of the code system, and construction of theories iteratively and inductively. We also performed thematic analysis on all the qualitative data collected from the questionnaire and forums using similar inductive, iterative, and grounded approaches. Qualitative analyses were completed in the MAXQDA⁵ software. MAXQDA is designed for computer-assisted qualitative and mixed methods data, text and multimedia analysis in academic, scientific, and business institutions. We first generated a series of “codes” that identified different subgroups of users of AI mixing tools. Over several iterations of the process, we deduced their use-case and sentiments around these tools. The multiple-choice questions from the questionnaire were visualised using numerical and statistical methods which in many cases complemented and strengthened the qualitative findings. The outcomes will be discussed in the following sections.

4 Results

Our investigation confirms the relevance of AI-based mixing tools in all three distinct groups of users: amateurs, pro-ams, and professionals as shown in Figure 1. We found that each user group has unique requirements for the intelligent tools and employs them for various purposes in their workflows. We discuss the results of all three studies together. We quote excerpts from semi-structured interviews and questionnaires indicated by ‘I’ and those from internet forums by ‘F’.

4.1 Amateurs

We observed that many forum participants who identified as amateurs were primarily musicians who create and compose music. They indicated limited knowledge

⁵<https://www.maxqda.com/>

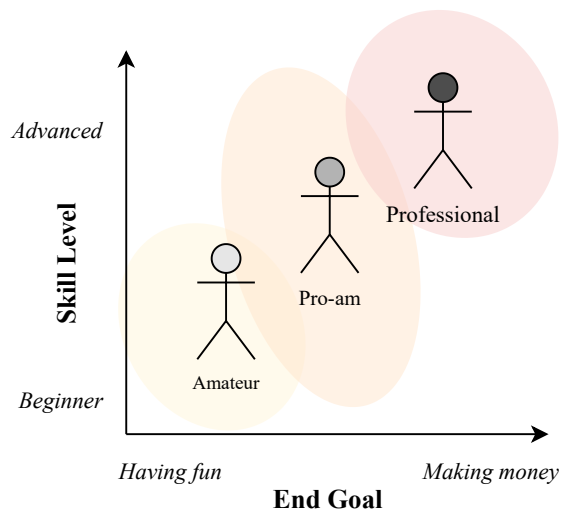


Fig. 1: Categories of users for smart tools

of music mixing and their main objective is to share their music with the world.

“The biggest barrier to me getting a recording out into the world is always the mixing process. It’s tedious, ... Give me a button I can push to automate most of the work.”

Several amateurs noted that mixing is the biggest hurdle in putting their music out. They also acknowledged that they are happy if the mix is decent enough.

“...give a beginner a simply decent mix. I’m not expecting to sound 100% perfect, but if they help, it would be enough for me.”

Fully Automatic mixing tools are more useful to this group since they typically don’t have the time or resources to invest in expensive equipment or hire a professional engineer. These tools can be a cost-effective solution for hobbyist musicians whose primary goal is creating and sharing music.

“Getting a mix or master 90% listenable is good enough for me to release music and lets me concentrate on what I enjoy.” [F]

Amateurs in the music production industry see AI-powered tools as a way to save time and trouble, and to produce a decent mix with minimal effort without a deep understanding of the technical aspects of mixing.

Due to their limited knowledge of the subject, they expect these systems to be highly autonomous almost like a “one-push magic button”. They also have lower expectations of the quality of the mix produced by these AI systems.

“I love recording my own songs, but I really can’t understand many of the concepts regarding mixing, I get easily lost with EQs, compression, etc. I’m basically an amateur, I respect the art of mixing, but that’s the phase where I lose the fun of making music.” [F]

Overall, we observed that amateurs are positively embracing this emerging technology.

4.2 Pro-ams

Our observations confirmed that pro-ams or semi-professional users have a higher level of technical skill than amateurs, but may still not have the same level of experience and expertise as professionals. Also, most often they don’t make money out of the craft, rather they produce and mix music for joy. They use smart mixing tools in a similar way to amateurs, but they also use them as a way to improve their skills and work towards becoming professional mix engineers. Pro-ams also use smart mixing tools as a way to quickly achieve a certain sound or style in their mixes.

“It could be used as a tool to learn the basics, then using the reference method, as well as trial and error, it’ll be another tool in the tool belt of learning. AI being used for basics, other methods to make more exciting mixes.” [F]

Given their higher level of expertise, pro-ams are well aware of the limitations of this technology but are willing to maximise the available smart tools to the best of their potential. They are cautiously optimistic about the future of these tools.

“None of this AI mixing is going to mix your music for you. It’s going to make basic suggestions of moves you can make to better process the audio. You, yourself, still have to take the brunt of the decision-making and the work.” [F]

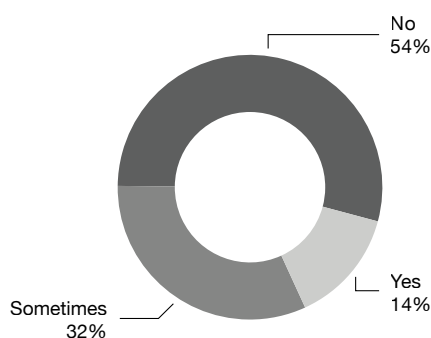


Fig. 2: Responses to the use of AI-powered tools in mixing workflow as reported by pro-ams and pros.

4.3 Professionals

When it comes to the experts in the field, professional engineers have a mixed view on the use of AI in mixing workflows. Some professionals use smart mixing tools as a way to improve their workflow and save time on repetitive tasks.

“How can we best use our time when we’re mixing a record or when we’re producing a record and having intelligent tools that take away some of these mundane tasks?” [I]

They may use these tools to assist them in the mixing process but ultimately rely on their own expertise and experience to create a polished final product. They believe these tools can provide a starting point and direction for the mix. Professionals may also use these tools as a way to experiment with new sounds and styles, or as a way to quickly achieve a certain sound or style in their mixes.

“You maybe can think of five amazing creative ideas in the mix if you’re lucky, whereas the system can maybe hit you with 100 great ideas and you can choose” [I]

“I’ve used Neutron’s masking feature to get some extra clarity but your talking one move of the potentially hundreds of thousands of moves involved in a production.” [I]

They expect these tools to be highly accurate, precise and customisable to their needs. They are certainly in favour of assistive and co-creative technologies that enable collaboration where the machine makes suggestions and the engineer can then adjust and fine-tune settings to their liking.

“...you know comparing your mix to what is going on on the reference song and recommending there is too much low end, there’s too much higher end, there’s too much of this and that.” [I]

“Machine could process as well and then give me the controls to just adjust everything. It would be much more helpful than starting from scratch.” [I]

Our investigation revealed that pros are currently using AI-based smart tools for tasks such as filtering, peak detection, pitch detection, mastering, equalisation, and sound enhancement.

However, a large number of professional engineers from our study said that they do not use smart AI tools in their mixing workflows as shown in Figure 2. These experts argue that AI-powered tools can’t fully replace the human touch and creativity required in the mixing process. They argue that these tools generate outputs using computational methods and do not understand the ‘feelings’ and ‘emotions’ required for mixing music intentionally. They argue that AI-controlled mixing could lead to a loss of control and precision in the final product.

“For a hobby, they can do that at home. On a professional level in a production setting, it won’t fly because client and corporate revisions are demanding and very precise in what they want on a granular level. It’s NOT about doing it faster but with precision.” [F]

Another expert pointed out that these tools are often met with scepticism in the court of public opinion, as they offer a “push-button magic solution” to something that otherwise requires a high-investment skill, which can invalidate the time and effort that many professionals have spent learning their craft. The idea that traditional methods of mixing are still superior and that learning by trial and error is the best way to master the art of mixing is still quite prevalent among the expert community.

In conclusion, some professionals see the potential of AI in improving their workflows, but ultimately it is up to individual preferences and workflows. Figure 3 presents the preference for usage of different categories of AI tools in mixing workflows as reported by pro-ams and pros. We see that there is more acceptance for automatic mixing over tools that apply audio effects directly to the given recordings. This could be explained as most often these automatic audio effect tools do not understand the context of the given stem in

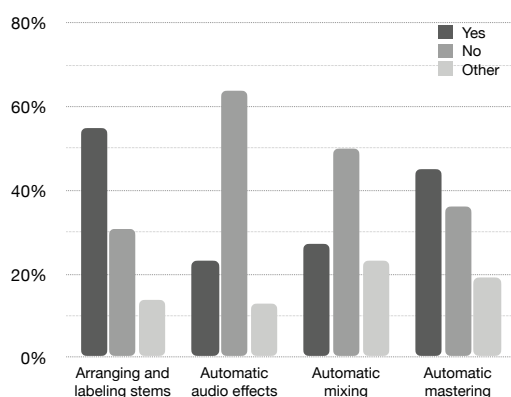


Fig. 3: Preference for the use of different AI tools in mixing workflow as reported by pro-ams and pros

the mix and are very generic in nature. We also observe that there is greater acceptance of labelling tools and automatic mastering tools. Many pros confirmed that they generally use auto-mastering tools to adjust the overall dynamics and loudness of a draft of the mix before sharing it with the client. They also explained that loading up multitrack in the DAW, labelling, and colour-coding stems before mixing is repetitive and takes up much time.

5 Discussion

“Mixing is without doubt an art, but there’s a lot that AI can help us with, and it’s an inevitable next step.” [F]

Most amateurs want to share their music with the world. However, the bottleneck to successful commercial release most often is mixing. Honing the skills for mixing or hiring a professional engineer is often an expensive affair. This category of users might employ automatic mixing services to create mixes that they believe are good enough for them to target a commercial release. Amateurs might also find use in assistive mixing tools that are corrective and enable smart editing [56]. This category of users are less interested in fine-grained control and customisation.

Pro-ams, on the other hand, are more intentional and particular about the quality of the final result and may be producing and mixing music for themselves or others. Pro-ams encompass the entire spectrum of users between amateurs and professionals. As mixing is

not their primary source of income, they are more exploratory and learning-focused. This category of users deploy AI-based tools in their workflows, however, given their higher levels of critical listening skills, they are looking for tools that are precise and offer more control. They use AI mixing tools as a starting point or for inspiration, but need more control over the final mix.

Finally, for professionals, mixing is an art form in itself. They are willing to spend more money on tools that they find useful and can help them speed up technical and repetitive tasks. However, they need a lot of control and customisation options. Table 2 lists smart tools under various categories and tasks that professionals ascertained will ease their mixing workflows. It should also be noted that there is less enthusiasm for fully automated mixing tools amongst experts as compared to those that provide some control. This could be because automated mixing is often seen as a “black box” with limited control. Through our interviews, it became evident that most professionals are open to using assistive and collaborative technologies in their workflows.

“...but I think where the computer system is interacting with the human, that’s where it’s really interesting.” [I]

We aggregate the use cases and expectations of various user groups from smart tools for mixing in Table 1.

5.1 Future of smart technology for mixing

The key takeaway of this investigation is that each user group has different goals and requirements, hence they might require tools that cater to their specific goals and needs. The future and utilisation of AI in mixing workflows are dependent on the design objectives of the tools. For the wider acceptance of these tools, it is important to build them such that they satisfy the three factors mentioned in Section 2.3. Here, we would like to propose a brief overview of possible directions.

5.1.1 Balance of Control and Automation

The right amount of control and automation has always been a matter of debate in human-computer interaction (HCI) [57, 58]. AI systems are often seen as black boxes that give out results that offer less control and interpretability. Our results show that each user group desires varying levels of automation and control. To counter this we propose designing AI tools that offer

Smart tools\User	Amateurs	Pro-ams	Professionals
Use-case	Create decent mix	Learning and exploratory tool	Automate repetitive and time-consuming tasks
	As a learning tool	To find a starting point	Co-creation and assistance
		Automate technical tasks	To find a starting point/direction for mix
		Creativity and inspiration	Creativity and inspiration
Expectations	Autonomous with less control	Advanced and more control	Highly advanced and wide range of control option
	Cost-effective	Accurate and precise	Accurate and precise
	Easy to use	Assistive	Assistive
		Cost effective	Easy integration in current workflow
		Easy integration into current workflow	Context-aware
Sentiment	Positive	Cautiously positive	Mixed

Table 1: Comparison of use-case, expectations, and sentiment amongst different categories of users of AI technology in mixing workflows

different levels of control and automation based on the user's level of expertise. This could be achieved by allowing the same plugin to have different modes, ranging from full automation, conditional automation, partial automation, assistive to no automation [41]. This would allow the user to salvage the same plugin/tool for multiple purposes based on the time available and the precision required. There have been works in this direction to design AI-based mixing systems that are interpretable and offer more control to the end user [27].

5.1.2 Seamless Integration

Professional mixing engineers often have an established workflow and familiar tools that helps them be efficient in their practice. Hence, we need to build tools and technology that have similar formats and configurations to what these users are familiar with. This implies turning research into products that are familiar, easy to access, navigate, and thus integrate into their existing workflows. For example, a model trained to apply automatic equalisation could be incorporated as a plugin that can be easily loaded into a DAW. This calls

for the creation of developer tools that enable quicker and easier translation of research into technology. This can be achieved by using frameworks and software development kits (SDK) like Neutone⁶, JUCE⁷, CMajor⁸, iPlug2⁹, and RTNeural¹⁰. This would allow researchers to get quicker feedback from the creatives and thus iterate over to create better tools with a clearer vision of the user needs. Another potential approach could be where the AI learns to control a DAW. This implies that given the stems in a project session, the system is able to predict the gain levels, panorama values, and audio effect chains with their parameters for stems to create a mix. This could be helpful for hobbyists and low-budget musicians to improve their productions, while professionals would use it to save time and create a boilerplate mix that can be refined later. This also opens the door for remodelling and redesigning existing DAWs

⁶<https://neutone.space/>

⁷<https://juce.com/>

⁸<https://cmajor.dev/>

⁹<https://iplug2.github.io/>

¹⁰<https://github.com/jatinchowdhury18/RTNeural>

Phase	Category	Tasks
Before Mixing	Mix preparation*****	Labelling and colour-coding tracks***
		Grouping and arranging tracks
		Importing session data
		Setting up sends, AUX, and buses**
	Editing	Trimming Silences
		Identifying and fixing clicks and pops
		Identifying and fixing phase issues
		Checking mono compatibility
During Mixing	Levels Balancing	Gain staging**
		Setting a rough fader mix***
	Spectral Corrections	Identifying and fixing masking issues**
		Auto-high pass filter for content with low bass content
		Auto EQ and EQ matching***
	Communicating	Getting artist feedback on time-stamps
	Co-creativity**	Suggesting ideas for mix**
		Recommendations for audio effect processing

Table 2: List of smart functionalities that mix engineers desire to ease their workflow; Frequency of the requests for the tool is represented by the number of asterisks (*)

to be able to use the increased power of consumer laptops and embedded hardware available in professional mixing consoles [59].

5.1.3 Context-Aware Systems

We observed that several pro-ams and pros admitted that AI-based tools are very generic and do not work well for outliers. They also expressed that these tools often only analyse the given track without understanding the context in the mix, which lowers trust and confidence in these tools.

"AI can't meaningfully interpret what your song calls for; its vibe. It can only do generic." [F]

This opens up the possibility to develop systems that are context-aware [60] and are trained on more diverse music and audio data to be able to work well for a variety of styles, genres and niche characteristics that are often a signature of innovation in the field.

These AI systems could also be made more contextually aware of the vision of the mix by allowing them to be queried by text [61], audio [29], and semantic descriptors [62, 63, 64]. Recent works have presented a state-of-the-art performance in text-based, melody-based, and semantic token-based music generation. We believe these systems could be further explored and fine-tuned for audio engineering applications.

Context-aware tools have the potential to bring in new talent, new genres and diversity into the music industry because of the ease of production that they offer. Amateurs might find it easier to produce and share music, thus potentially getting recognised for their craftsmanship. These tools might also help pro-ams improve skills and practice mixing professionally. Overall, these tools have the potential to improve mixing workflows and increase productivity and creativity.

6 Summary

The use of AI in mixing workflows is a complex and nuanced topic. While some see it as a potential game-changer, others are more skeptical about its ability to replace the human touch and creativity required in the mixing process. In this investigation, we identified three categories of users for the AI-based tool for mixing workflows. Based on the knowledge gained from them, we analysed and reported the sentiments and expectations of each of them. We argue that this information can be helpful and create a basis for further development and application of AI tools that can simplify the work of users in each category.

In conclusion, we believe that the user-specific insight that our study provides will aid researchers in building AI mixing technologies that are better accepted and integrated by the audio engineering community in the time to come. Additionally, AI-powered tools could become further precise, versatile and adaptable to different genres and styles of music, and able to work with a wider range of input audio. Another possibility is the development of AI that can understand and respond to the creative intent of the user, allowing for more intuitive and efficient collaboration between humans and machines in the music production process. The field of AI and Music is constantly evolving, and the future possibilities are only limited by the current technology and the imagination of the researchers.

7 Acknowledgements

We express our sincere gratitude to the AES Europe convention reviewers for providing valuable feedback on our work. We are truly appreciative of the contributions made by all the mixing engineers and participants who participated in the interviews and studies and shared their valuable knowledge with us. We extend our thanks to the Steinberg research and development team for their unwavering support and honest feedback throughout this project.

Additionally, we acknowledge the invaluable input of our colleagues, Christian Steinmetz, Gary Bromham, and Angeliki Mourgela, who engaged in insightful discussions on this topic and provided critical feedback that helped shape our work.

We are also grateful for the financial support provided by UK Research and Innovation [grant number

EP/S022694/1]. This funding has been instrumental in facilitating our research and enabling us to contribute to the field.

References

- [1] Campolo, A., Sanfilippo, M., Whittaker, M., and Crawford, K., *AI Now 2017 Report*, AI Now Institute at New York University, 2017.
- [2] Sturm, B. L., Iglesias, M., Ben-Tal, O., Miron, M., and Gómez, E., “Artificial intelligence and music: open questions of copyright law and engineering praxis,” in *Arts*, volume 8, p. 115, MDPI, 2019.
- [3] Jordanous, A., “Evaluating evaluation: Assessing progress and practices in computational creativity research,” *Computational creativity: The philosophy and engineering of autonomously creative systems*, pp. 211–236, 2019.
- [4] Ritchie, G., “The evaluation of creative systems,” *Computational creativity: The philosophy and engineering of autonomously creative systems*, pp. 159–194, 2019.
- [5] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J., “Building machines that learn and think like people,” *Behavioral and brain sciences*, 40, p. e253, 2017.
- [6] Ribera, M. and Lapedriza, A., “Can we do better explanations? A proposal of user-centered explainable AI,” in *IUI workshops*, volume 2327, p. 38, 2019.
- [7] Roy, Q., Zhang, F., and Vogel, D., “Automation accuracy is good, but high controllability may be better,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, 2019.
- [8] Miller, M., *Mixing Music*, Penguin, 2016.
- [9] Izhaki, R., *Mixing audio: concepts, practices, and tools*, Routledge, 2017.
- [10] Hracs, B. J., “A creative industry in transition: The rise of digitally driven independent music production,” *Growth and Change*, 43(3), pp. 442–461, 2012.

- [11] Sandler, M., De Roure, D., Benford, S., and Page, K., “Semantic web technology for new experiences throughout the music production-consumption chain,” in *2019 International Workshop on Multilayer Music Representation and Processing (MMRP)*, pp. 49–55, IEEE, 2019.
- [12] McGrath, S., Chamberlain, A., and Benford, S., “Making music together: an exploration of amateur and pro-am Grime music production,” in *Proceedings of the Audio Mostly 2016*, pp. 186–193, 2016.
- [13] Bromham, G., “How can academic practice inform mix-craft?” in *Mixing Music*, pp. 265–276, Routledge, 2016.
- [14] Hoare, M., Benford, S., Jones, R., and Milic-Frayling, N., “Coming in from the margins: amateur musicians in the online age,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1295–1304, 2014.
- [15] Stebbins, R. A., “The amateur: Two sociological definitions,” *Pacific Sociological Review*, 20(4), pp. 582–606, 1977.
- [16] Prior, N., “New amateurs revisited: Popular music, digital technology, and the fate of cultural production,” in *Routledge handbook of cultural sociology*, pp. 340–348, Routledge, 2018.
- [17] Leadbeater, C. and Miller, P., *The pro-am revolution: How enthusiasts are changing our society and economy*, Demos London, 2004.
- [18] Dugan, D., “Automatic microphone mixing,” in *Audio Engineering Society Convention 51*, Audio Engineering Society, 1975.
- [19] De Man, B., Reiss, J. D., and Stables, R., “Ten Years of Automatic Mixing,” in *3rd Workshop on Intelligent Music Production*, 2017.
- [20] Steinmetz, C. J., Vanka, S. S., Martínez Ramírez, M. A., and Bromham, G., *Deep Learning for Automatic Mixing*, ISMIR, 2022.
- [21] Tom, A., Reiss, J. D., and Depalle, P., “An automatic mixing system for multitrack spatialization for stereo based on unmasking and best panning practices,” in *Audio Engineering Society Convention 146*, Audio Engineering Society, 2019.
- [22] Moffat, D. and Sandler, M., “Machine learning multitrack gain mixing of drums,” in *147th Audio Engineering Society Convention*, 2019.
- [23] Martínez-Ramírez, M. A., Stoller, D., and Moffat, D., “A deep learning approach to intelligent drum mixing with the Wave-U-Net,” *Journal of the Audio Engineering Society*, 2021.
- [24] Steinmetz, C. J., Pons, J., Pascual, S., and Serrà, J., “Automatic multitrack mixing with a differentiable mixing console of neural audio effects,” in *ICASSP*, IEEE, 2021.
- [25] Colonel, J. T. and Reiss, J., “Reverse engineering of a recording mix with differentiable digital signal processing,” *The Journal of the Acoustical Society of America*, 150(1), pp. 608–619, 2021.
- [26] De Man, B. and Reiss, J. D., “A knowledge-engineered autonomous mixing system,” in *135th Audio Engineering Society Convention*, Audio Engineering Society, 2013.
- [27] Steinmetz, C. J., *Learning to mix with neural audio effects in the waveform domain*, Master’s thesis, Universitat Pompeu Fabra, 2020, doi:10.5281/zenodo.4091203.
- [28] Martínez-Ramírez, M. A., Liao, W.-H., Fabbro, G., Uhlich, S., Nagashima, C., and Mitsufuji, Y., “Automatic music mixing with deep learning and out-of-domain data,” in *ISMIR*, 2022.
- [29] Koo, J., Martínez-Ramírez, M. A., Liao, W.-H., Uhlich, S., Lee, K., and Mitsufuji, Y., “Music Mixing Style Transfer: A Contrastive Learning Approach to Disentangle Audio Effects,” *arXiv preprint arXiv:2211.02247*, 2022.
- [30] Martínez-Ramírez, M. A. and Reiss, J. D., “End-to-end equalization with convolutional neural networks,” in *21st International Conference on Digital Audio Effects (DAFx-18)*, 2018.
- [31] Sheng, D. and Fazekas, G., “A feature learning siamese model for intelligent control of the dynamic range compressor,” in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2019.
- [32] Singh, S., Bromham, G., Sheng, D., and Fazekas, G., “Intelligent Control Method for the Dynamic

- Range Compressor: A User Study,” *Journal of the Audio Engineering Society*, 69(7/8), pp. 576–585, 2021, doi:10.17743/jaes.2021.0028.
- [33] Ma, Z., De Man, B., Pestana, P. D., Black, D. A., and Reiss, J. D., “Intelligent multitrack dynamic range compression,” *Journal of the Audio Engineering Society*, 63(6), pp. 412–426, 2015.
- [34] Steinmetz, C. J. and Reiss, J. D., “Efficient neural networks for real-time modeling of analog dynamic range compression,” in *151st Convention of the Audio Engineering Society*, Audio Engineering Society, 2022.
- [35] Kuznetsov, B., Parker, J. D., and Esqueda, F., “Differentiable IIR filters for machine learning applications,” in *Proc. Int. Conf. Digital Audio Effects (eDAFx-20)*, pp. 297–303, 2020.
- [36] Colonel, J. T., Steinmetz, C. J., Michelen, M., and Reiss, J. D., “Direct design of biquad filter cascades with deep learning by sampling random polynomials,” in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3104–3108, IEEE, 2022.
- [37] Nercessian, S., “Neural parametric equalizer matching using differentiable biquads,” in *Proc. Int. Conf. Digital Audio Effects (eDAFx-20)*, pp. 265–272, 2020.
- [38] Colonel, J. T., Comunità, M., and Reiss, J., “Reverse Engineering Memoryless Distortion Effects with Differentiable Waveshapers,” in *153rd Convention of the Audio Engineering Society*, Audio Engineering Society, 2022.
- [39] Lee, S., Choi, H.-S., and Lee, K., “Differentiable artificial reverberation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30, pp. 2541–2556, 2022.
- [40] Wilmering, T., Fazekas, G., and Sandler, M. B., “High Level Semantic Metadata for the Control of Multitrack Adaptive Audio Effects,” *presented at the 133rd Convention of the AES, San Francisco, USA*, 2012.
- [41] Tsiros, A. and Palladini, A., “Towards a Human-Centric Design Framework for AI Assisted Music Production,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 399–404, 2020.
- [42] De Man, B., Stables, R., and Reiss, J. D., *Intelligent Music Production*, Routledge, 2019.
- [43] Long, D., Padiyath, A., Teachey, A., and Magerko, B., “The role of collaboration, creativity, and embodiment in AI learning experiences,” in *Creativity and cognition*, pp. 1–10, 2021.
- [44] Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S., “Explainable ai: A review of machine learning interpretability methods,” *Entropy*, 23(1), p. 18, 2020.
- [45] Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L., “Explaining explanations: An overview of interpretability of machine learning,” in *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pp. 80–89, IEEE, 2018.
- [46] Sterne, J. and Razlogova, E., “Machine learning in context, or learning from landr: artificial intelligence and the platformization of music mastering,” *Social Media+ Society*, 5(2), p. 2056305119847525, 2019.
- [47] Scurto, H. and Bevilacqua, F., “Appropriating music computing practices through human-AI collaboration,” in *Journées d’Informatique Musicale (JIM 2018)*, 2018.
- [48] Seale, C., Charteris-Black, J., MacFarlane, A., and McPherson, A., “Interviews and internet forums: a comparison of two sources of qualitative data,” *Qualitative health research*, 20(5), pp. 595–606, 2010.
- [49] Brinkmann, S., “Unstructured and semi-structured interviewing,” *The Oxford handbook of qualitative research*, 2, pp. 277–299, 2014.
- [50] Shaw, M. L. and Woodward, J. B., “Modeling expert knowledge,” *Knowledge Acquisition*, 2(3), pp. 179–206, 1990.
- [51] Bogner, B., M. W., A. Littig, *Interviews mit Experten*, Springer Fachmedien Wiesbaden, 2006.
- [52] Baker, S. and Edwards, R., “How many qualitative interviews is enough? Expert voices and early

- career reflections on sampling and cases in qualitative research,” *National Centre for Research Methods Review Paper*, 2017.
- [53] Bickart, B. and Schindler, R. M., “Internet forums as influential sources of consumer information,” *Journal of interactive marketing*, 15(3), pp. 31–40, 2001.
- [54] Braun, V. and Clarke, V., *Thematic analysis.*, pp. 57–71, 2012, ISBN 978-1-4338-1003-9.
- [55] Braun, V. and Clarke, V., “Using thematic analysis in psychology,” *Qualitative research in psychology*, 3(2), pp. 77–101, 2006.
- [56] Fazekas, G. and Sandler, M., “Intelligent Editing of Studio Recordings with the Help of Automatic Music Structure Extraction,” *presented at the AES 122nd Convention, Vienna, Austria, 2007.*
- [57] Xu, W., Dainoff, M. J., Ge, L., and Gao, Z., “Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI,” *International Journal of Human–Computer Interaction*, 39(3), pp. 494–518, 2023.
- [58] McFarlane, D. C. and Latorella, K. A., “The scope and importance of human interruption in human-computer interaction design,” *Human-Computer Interaction*, 17(1), pp. 1–61, 2002.
- [59] Turchet, L., Lagrange, M., Rottondi, C., Fazekas, G., Peters, N., Østergaard, J., Font, F., Bäckström, T., and Fischione, C., “The Internet of Sounds: Convergent Trends, Insights and Future Directions,” *IEEE Internet of Things Journal*, pp. 1–1, 2023, doi:10.1109/JIOT.2023.3253602.
- [60] Lefford, M. N., Bromham, G., Fazekas, G., and Moffat, D., “Context-Aware Intelligent Mixing Systems,” *Journal of the Audio Engineering Society*, 69, pp. 128–141, 2021, 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., and Frank, C., “MusicLM: Generating Music From Text,” 2023.
- [62] Stables, R., Enderby, S., De Man, B., Fazekas, G., and Reiss, J. D., “SAFE: A system for extraction and retrieval of semantic audio descriptors,” 2014.
- [63] Stables, R., De Man, B., Enderby, S., Reiss, J. D., Fazekas, G., and Wilmering, T., “Semantic description of timbral transformations in music production,” in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 337–341, 2016.
- [64] Turchet, L., Pauwels, J., Fischione, C., and Fazekas, G., “Cloud-Smart Musical Instrument Interactions: Querying a Large Music Collection with a Smart Guitar,” *ACM Trans. Internet Things*, 1(3), 2020, ISSN 2691-1914, doi:10.1145/3377881.