Audio Engineering Society

Convention Paper 10229

Presented at the AES 158th Convention
2025 May 22–24, Warsaw, Poland

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.

Immersive Music Production Workflows: An Ethnographic Study of Current Practices

Marcela Rada[1], Russell Mason[1], and Enzo De Sena[1]

[1]*Institute of Sound Recording, University of Surrey, Stag Hill, Guildford, GU2 7XH, UK*

Correspondence should be addressed to Marcela Rada (c.rada@surrey.ac.uk)

ABSTRACT

This study presents an ethnographic analysis of current immersive music production workflows, examining industry trends, tools, and methodologies. Through interviews and participant observations with professionals across various sectors, the research identifies common patterns, effective strategies, and persistent obstacles in immersive audio production. Key findings highlight the ongoing struggle for standardized workflows, the financial and technological barriers faced by independent artists, and the critical role of collaboration between engineers and creatives. Despite the growing adoption of immersive formats, workflows still follow stereo conventions, treating spatialization as an afterthought and complicating the translation of mixes across playback systems. Additionally, the study explores the evolving influence of object-based and bed-based mixing techniques, monitoring inconsistencies across playback systems, and the need for improved accessibility to immersive production education. By synthesizing qualitative insights, this paper contributes to the broader discourse on immersive music production, offering recommendations for future research and industry-wide best practices to ensure the sustainable integration of spatial audio technologies.

1 Introduction

The rise of immersive music production is reshaping the industry by introducing spatial audio technologies like Dolby Atmos and binaural audio. These formats shift the focus from traditional stereo to multi-dimensional soundscapes, transforming music composition, production, and consumption. As streaming platforms adopt immersive audio, demand for such content is rising.

This shift is both technological and creative, requiring producers and engineers to rethink workflows. Apple's Eddy Cue noted that mixing in Dolby Atmos is not a simple stereo conversion but demands careful adjustments by engineers and artists [25]. Financial incentives also drive adoption; Apple Music offers up to 10% higher royalties for Spatial Audio tracks [9], prompting major labels to reissue older music and release new tracks in both formats.

Industry adoption presents several challenges. For example, mixing engineers primarily manage spatial audio, while musicians and producers often lack the expertise to contribute meaningfully. Caroline Hilton of Universal Music Group (UMG) highlighted that while artists ultimately found their creative intent remained intact, they initially struggled to assess Atmos mixes due to unfamiliarity [13]. UMG has guidelines, but only for engineers [28], reflecting broader gaps in best practices. Existing resources focus on hardware and software setup rather than ensuring mix consistency across playback systems [8].

Despite immersive audio's growing presence, workflows still follow stereo conventions [4], treating spatialization as an afterthought [5]. This reliance on stereo-oriented workflows creates challenges: many engineers use binaural monitoring due to cost and space constraints, which aligns with consumer habits but complicates translation to speaker systems [22]. Furthermore, the lack of standardized approaches results in inconsistencies across studios and artists.

New workflows are emerging, prioritizing spatialization earlier in the production process [5, 7, 4]. Based on interviews and participant observations, this study examines the current state of immersive music production, technological advancements, and evolving best practices.

The paper is structured as follows: Section 2 reviews existing literature on immersive music production, Section 3 outlines the methodology used to collect and analyze data, Section 4 presents results, and Section 5 discusses them. Finally, Section 6 concludes with key insights and future directions.

2 Existing Literature on Immersive Music Production

This section addresses the research question: *What key trends, tools, and techniques shape commercial and immersive music production workflows?* It does so by reviewing key studies on immersive music production. The literature offers insights into the integration of spatial audio in popular music, examining both technical and creative challenges. It highlights the intersection of technological innovation and artistic creativity while identifying gaps for future research.

2.1 From Stereo to Immersive

Many engineers still rely on stereo mixes as the foundation for their Dolby Atmos productions, adapting them rather than creating spatial mixes from scratch [4]. Dewey et al. [7] found that engineers use similar EQ techniques in both stereo and Atmos, but opinions on compression and bus processing differ. Approximately 50% of respondents use compression differently in Atmos, indicating evolving dynamics in spatial mixing. These findings suggest that immersive formats may require less processing due to the increased space available for sound placement [7]. However, further research is needed, particularly through semi-structured interviews and participant observations, to refine workflow strategies.

2.2 Recommended Mixing Techniques

Immersive mixing incorporates techniques from stereo production while requiring adaptations to address spatial complexity. One key difference is the need for greater dynamic control, as excessive dynamic range becomes more noticeable in multi-speaker arrays [17, 29]. To accommodate this, engineers often avoid master bus compression in immersive formats, ensuring flexibility across different downmixes [17].

Another critical factor is spatial clarity, particularly in the low-frequency range. Localization blur increases at low frequencies, making precise placement more difficult [30, 26]. To counter this, engineers often reduce low-frequency content in height channels while boosting higher frequencies, which improves spatial precision and clarity [17].

The choice of object-based vs. bed-based workflows impacts immersion. While renowned engineers such as Bob Clearmountain and Chris Lord Alge prefer bed-based positioning, others, including Kybert, advocate for object-based mixing, arguing that individualized object placement enhances spatial engagement [4]. However, Greg Wells warns that excessive movement in the immersive field can be distracting rather than immersive, emphasizing the need for subtlety [4].

Finally, spatial depth - the perceived distance and three-dimensional placement of sounds within a mix, and mix balance remain crucial considerations. Engineers typically anchor core elements such as vocals, bass, and drums, in the left-right speaker pair, maintaining a sense of familiarity within the expanded soundstage. Meanwhile, reverb and backing vocals are often positioned in height and surround channels to enhance the immersive experience [4]. This suggests that while immersive formats provide new spatial possibilities, stereo mixing principles still play a foundational role in immersive production.

2.3 The Center Channel

One feature of multi-channel formats is the inclusion of a central speaker. Lawrence [17] suggests using it for key musical elements such as vocals, kick, snare, and bass, repurposing its traditional role from film dialogue. However, Bourbon et al. [4] notes that many engineers still prefer a phantom center, and one participant avoided the center entirely due to filtering introduced by the Dolby Atmos renderer during binaural playback.

Bourbon et al. [4] also highlight differing views on the use of the center and LFE (low-frequency effects) channels. Clearmountain favors placing vocals in the center, while Jessy Ray Ernster avoids it due to rendering artifacts. Both engineers approach the LFE channel cautiously, particularly in headphone-based listening.

As context for their study, King et al. [15] describe how phantom imaging results from interaural time and level differences, which can introduce crossfeed. This occurs when each ear receives input from both speakers at slightly different times and levels, potentially leading to frequency response inconsistencies, such as midrange dips at the listening sweet spot. A discrete center signal, by contrast, reaches both ears at the same time and typically provides a more even frequency response. The study itself compared phantom and discrete center sources using binaural renderers and dummy head simulations. Phantom centers showed tonal artifacts in binaural playback, including low-frequency boosts and a dip near 1 kHz. Discrete center sources maintained a flatter and more stable response. These differences were especially pronounced in off-center listening, where the phantom image deteriorated more quickly than the discrete center.

Dolby's spatialization settings also affect perception. While many engineers prefer to disable spatialization on LCR channels to reduce coloration, King et al. [15] recommend auditioning mixes with spatialization set to NEAR or MID, since most streaming platforms apply spatialization by default.

Headroom is another practical concern. Routing all central elements exclusively to the center channel can lead to overload, so engineers often distribute them across the LCR channels using divergence. This helps manage level while preserving spatial detail. Although phantom centers can enhance perceived width, they may require additional reflections or reverb in the side channels to compensate for tonal instability.

Mixers also face limitations when only stereo stems are available, reducing flexibility in treating center elements. Even in headphone-only workflows, crossfeed and inter-channel interference remain part of the binaural rendering process, which can affect imaging and tone.

These findings underscore the complexity of center channel use in immersive formats, where tonal stability, spatial clarity, and playback translation must all be carefully balanced.

2.4 Monitoring

Accurate monitoring is essential in immersive audio to ensure mix translation, meaning how a mix retains its intended spatial characteristics across different playback systems. Unlike stereo mixing, where translation is primarily tested on two speakers or headphones, immersive formats require compatibility across multiple speaker configurations and binaural playback. Engineers must balance optimizing for high-end multi-speaker systems and ensuring the mix translates well to consumer-grade setups.

To understand how professionals navigate this challenge, Dewey et al. [7] surveyed mixing engineers, finding that 69% used 7.1.4 setups and 18% used 9.1.4. Notably, 59% prioritized checking mixes on fewer speakers, while only 13% emphasized larger systems, highlighting a prevailing industry focus on downmix compatibility over high-end optimization.

To improve accessibility, binaural rendering tools allow engineers to monitor spatial mixes without the need for full speaker arrays, reducing production costs. These tools simulate multi-speaker environments over headphones, providing a practical alternative for independent producers. Several DAWs and third-party solutions offer integrated binaural renderers. However, some professionals recommend that final mixes still be reviewed in a 7.1.4 studio to ensure accurate translation across spatial audio systems [13].

King and Theriault [14] highlight additional monitoring challenges related to low-frequency content. The authors found that immersive mixes relying heavily on the LFE channel may produce inconsistent playback across platforms, particularly in binaural contexts. For example, Apple's Spatial Audio rendering showed a loss of around 3 dB at 58 Hz when LFE content was summed with other channels. However, they caution that rendering software is frequently updated, so these results are specific to the time of testing. Mixers are encouraged to stay informed about changes to playback engines and renderers.

King and Theriault [14] also note that engineers have limited control over how their masters are rendered across streaming services. Although a single ADM file is usually delivered, each platform may apply different spatialization and filtering settings, and might not preserve the mixer's intended configurations. These findings reinforce the importance of cross-platform

monitoring and careful LFE use to ensure consistent low-frequency translation in consumer playback environments.

While this hybrid workflow (binaural monitoring followed by speaker validation) is widely endorsed, empirical research on its effectiveness remains limited. More structured studies are needed to establish best practices for spatial mix translation and to refine workflows that balance both efficiency and accuracy.

3 Methodology

This section details the research methodology, outlining the approaches used for data collection and analysis.

3.1 Research Philosophy

Qualitative research has been successfully used to examine professional workflows across industries [2, 24, 3]. This approach is particularly suited for immersive music production because it captures the nuances of creative decision-making, which may not be easily quantified. This research prioritizes understanding individual experiences over generalizable laws, a paradigm that is referred to as “interpretivist paradigm" [23]. An overt research strategy ensured that participants understood the study’s purpose, beneficiaries, and outcomes, fostering engagement from skilled professionals [23].

3.2 Ethnography for Data Collection

Ethnographic research has been used to study digital workflows across creative fields. It has examined the digital practices of humanists—scholars and professionals engaged in the study of culture, history, philosophy, and the humanities [1]; music production workflows, including collaboration and metadata management [19]; and popular music studies, focusing on the lived experiences and social relationships of music-making [6]. It has also been applied to music-related fields, providing ecologically valid insights into musicians’ interactions with live performance technology [21] and analyzing how producers configure tools, manage metadata, and balance creativity with technical demands [20]. Additionally, studies of radio production teams, such as BBC Radio, have examined workflow challenges and contributed to tool development [2].

Ethnography was chosen as the data collection method for this study due to its ability to provide a holistic understanding of participants’ work environments and practices. Semi-structured interviews provided participants with the opportunity to share their experiences, challenges, and techniques. This format allowed for flexibility in exploring topics relevant to the study’s objectives. The interview structure covered six main topics:

1. **Educational and Training Background** – Participants’ formal and informal learning experiences.
2. **Work Setup** – Hardware and software configurations, recording and calibration techniques.
3. **Workflows for Immersive Music Production** – Differences between stereo and immersive workflows, creative approaches.
4. **Equipment** – Selection of tools, decision-making regarding hardware and software use.
5. **Challenges and Solutions** – Technical and logistical difficulties encountered and strategies used to address them.
6. **Limitations and Industry Changes** – Gaps in current technology, accessibility issues, and industry developments.

While all observation involves some degree of participation, this study adopted the complete observer approach, in which the researcher observes, records, and analyzes participants’ actions and decision-making without direct interaction, maintaining intellectual detachment and using directed questioning [23]. Observations were systematically documented, and the researcher regularly stepped back to conduct critical analysis, ensuring a comprehensive understanding of workflows, decision-making, and engagement with immersive music production technologies. This approach captured data that might not have been articulated in interviews, providing additional context and allowing for triangulation -the process of using multiple data sources to cross-verify findings and enhance research validity. For these observations, the researcher:

- Followed a free-flowing approach, allowing participants to work without structured tasks or external prompts.
- Documented data using detailed field notes rather than recordings.

- Informed the participants of the observations in advance, and the researcher remained passive and unobtrusive to minimize influence on workflow.

3.3 Sample Selection Methods

The participants interviewed and observed in this study were professional artists, composers, audio engineers, software and hardware developers, and music industry executives with significant experience in immersive music production at the time of data collection. This investigation used Mixed Purposeful sampling, which integrates two or more techniques to align data collection with research objectives, enable triangulation by examining the issue from different perspectives, and remain adaptable to various stakeholders, such as funders or policymakers [27]. To achieve this, the two complementary techniques employed were: Purposive Sampling, and Snowball Sampling.

3.3.1 Purposive Sampling

Purposive sampling involves selecting participants that meet specific criteria. Researchers using this method aim to cover all important aspects of the study while also creating a diverse group to examine how the chosen criterion varies across different backgrounds [23]. All participants in this study possessed a background in this field and meet at least one of the following criteria:

- **Professional Practice:** Actively working on immersive music productions at a professional level.
- **Technical Expertise:** Experience in designing and/or applying spatial audio workflows utilizing hardware and software.
- **Industry Integration:** Involvement with record labels and/or streaming platforms.
- **Technological Influence:** Having an input in the development of spatial audio technologies.
- **Resource Access:** Having access to spatial audio tools.
- **Industry Recognition:** Being recognized in the music industry for their work in immersive music production through awards such as Grammys or features in major industry publications.

Furthermore, Purposive Sampling aimed to ensure representation across two key dimensions:

- **Educational Background:** The sample includes participants with diverse academic backgrounds relevant to immersive music production.
- **Production Complexity:** Participants represent a range of complexities in their immersive music production setups.

While all participants shared a background in immersive music production, these additional criteria ensured diversity, incorporating a range of experiences and perspectives for a more nuanced understanding of immersive music production practices.

3.3.2 Snowball Sampling

Snowball sampling begins with a small group of initial contacts, expanding through referrals [23]. While this method can lead to homogeneity and exclusion of isolated individuals, these limitations can be mitigated by carefully selecting the initial sample, ensuring broad representation of major population characteristics, and making efforts to include hard-to-reach groups [23].

The researcher, leveraging industry experience, initially contacted colleagues involved in immersive music production, who then referred additional experts, ensuring a diverse participant pool. The advantage of Snowball sampling was that it allowed access to professionals, leveraged industry knowledge, and helped identify under-documented studies in the field.

3.3.3 Participants

While the study initially involved 30 participants, this paper presents a subset of those results which includes 11 participants. By selecting this smaller, heterogeneous group, the analysis captures a range of experiences and workflows while remaining focused. These preliminary findings informed the broader analysis of data from the remaining 19 participants in later stages of the study. This pool of participants is composed of 1 record label executive, 1 artist manager, 2 artists, 2 researchers, and 5 audio engineers. Table1 in Appendix A outlines how the 30 participants met the criteria mentioned above. The first 11 participants (highlighted in gray) are the participants chosen for this preliminary analysis. All 11 participants were interviewed. However, only participants 6, 7, and 8 were observed.

Classification by Level of Expertise

Participants were classified as beginner, intermediate, advanced, or expert in immersive music production based on a combination of objective indicators and self-reported competencies. These criteria included the number of immersive projects completed, the complexity of roles undertaken, familiarity with spatial audio tools, and recognition within professional networks such as music industry awards. Beginners had minimal exposure or had participated in only one project, while intermediate participants had contributed to multiple projects and demonstrated foundational proficiency with relevant software and mixing techniques. Advanced participants had extensive experience, often assuming lead roles and displaying a high level of expertise in spatial audio methodologies. Experts were those with long-term involvement in immersive music production, having directed major projects and earned recognition for their contributions to the field.

Classification by Production Complexity

The complexity of participants' typical productions was classified as standard, complex, or highly complex based on factors related to monitoring configurations and the routing of hardware and software components. Standard productions involved a single, well-established speaker array (for example, a 7.1.4 system) and a single renderer, typically the Dolby Atmos renderer, with minimal additional routing. Complex productions employed either multiple renderers or more advanced routing, which might include combining software- and hardware-based workflows or using more than one immersive audio format. Highly complex productions incorporated extended or custom speaker arrays beyond 7.1.4, used multiple renderers or proprietary technologies in tandem, and required detailed coordination among various hardware and software solutions. This classification took into account the choices made in monitoring setups, routing processes, and the level of technical integration required to support each production.

3.4 Data Analysis

Content analysis is a method of data analysis that examines recorded communications by identifying key elements, structuring them into a framework, and interpreting patterns within textual data [16]. This method categorizes content while considering participant perspectives and contextual influences, making it effective for analyzing interviews, surveys, media, and observations.

This study employed content analysis to systematically categorize data. Specifically, a *conventional* content analysis approach was used, allowing categories to emerge naturally rather than relying on predefined frameworks [12]. Through inductive category development, insights surfaced as data were grouped into broader patterns, moving from concrete observations to abstract interpretations [10].

3.4.1 Analysis Structure

Determining Classes and Categories

Codes are essential in content analysis, serving as tools to organize and interpret data. They are brief, descriptive labels that represent key ideas or attributes found in multiple parts of the text [16]. In conventional content analysis, the process of data coding involves defining common classes, special classes, and theoretical classes [11]. In this study, this process involved multiple phases.

Pre-Analysis:

During the initial engagement with the data, the researcher used the sampling criteria and participant demographics to establish common classes - everyday categories such as age, gender, and occupational roles, which can be useful for linking data to demographic characteristics [11].

- Job Roles: Record Label Executive, Artist Manager, Artist/Composer, Researcher, and Audio Engineer.
- Experience Levels: Beginner to expert experience in immersive music production.
- Work Environments: Commercial studio, personal studio, academic institution. These classes provided a descriptive framework to contextualize participants' responses and link them to their backgrounds. These common classes were determined before the detailed analysis began and served as a foundation for only organizing data into manageable units.

During Analysis:

Special classes involve specialized language used within specific groups, such as slang, acronyms, or professional jargon, allowing researchers to explore cultural or social distinctions within the data [11]. As the researcher immersed in the data (e.g., reading transcripts, reviewing observation notes), the following special classes were identified.

- Technical Terms: Spatial audio, binaural audio, Dolby Atmos, ambisonics, immersive formats, etc.
- Workflow Tools: Specific software or hardware mentioned (DAWs like Pro Tools, Nuendo, plugins, VR tools).
- Project Stages: Terms like pre-production, mixing, rendering, mastering for immersive outputs.

Post Initial Coding:

Theoretical classes are categories developed during data analysis and are often derived from theoretical models, allowing researchers to impose a theoretical framework on the empirical data and adjust it iteratively based on findings [11]. In this study, these were determined after the initial coding stage, as broader patterns and relationships between codes became evident. These classes were developed inductively and refined through multiple iterations of data organization, aligning with the principles of conventional content analysis. For example, categories such as:

- Effective Techniques: strategies or tools that work well.
- Challenges: technical limitations, software gaps, creative hurdles, collaboration issues.
- Limitations: time constraints, technology accessibility, budget issues, knowledge gaps.
- Workflow Patterns: recurring approaches across professionals that highlight consistency or innovation.

The iterative process of determining classes and categories ensured that they were rooted in the data itself, rather than being predefined or imposed externally. This approach allowed the analysis to remain responsive to participants' explicit descriptions and ensured that the findings reflected their lived experiences.

3.4.2 Approaches to Establishing Trustworthiness

To ensure consistency and accuracy of the data analysis, the following measures were implemented as recommended by Mason [18].

Triangulation: Data from interviews, observations, and field notes were cross-referenced to identify shared patterns and address inconsistencies. Triangulation helped reveal connections between participants' explicit practices and workflows.

Literal Readings: Literal reading focuses on the direct content, structure, and form of the data, rather than interpreting underlying meanings or researcher reflexivity. In the context of interview transcripts, this means analyzing the specific words, language, dialogue structure, and sequencing of interactions [18]. This approach was particularly important for documenting technical and procedural aspects of workflows.

Data Saturation: Data collection continued until no new patterns or categories emerged, ensuring that the findings comprehensively addressed the research questions.

4 Results

This section presents the preliminary findings of the study, reporting on the results derived from the analysis of the qualitative data collected through interviews and observations. The study examined the workflows employed in the creation of immersive music and identifies common patterns, including effective techniques, challenges, and limitations. The analysis addresses the research questions: *What workflows do industry professionals use in immersive music production?* and *What common patterns, effective techniques, challenges, and limitations exist across workflows?*

4.1 Category 1: Workflows

Fig.1 provides an organized view of immersive music production workflows, structured around four stages commonly identified by participants: data wrangling and template setup, stereo mix reconstruction, Atmos mix adjustments, and final checks and rendering. Arranging the findings according to these stages clarifies how participants move from initial preparation to delivering the final mix. Each stage includes the key tasks performed, the codes that emerged from the analysis,

and a summary of similarities and differences observed across multiple participants.

The findings highlight both consistent practices and variations across immersive music production workflows. The differences among participants appear to be influenced by factors such as workflow efficiency, creative intent, and technical constraints.

One area of consistency is in data wrangling and setup, where all participants employ structured file organization and some form of session template. However, the extent of customization varies. P6 and P11 prefer project-specific templates, while P7, P8, and P9 rely on a single standardized template that can be modified as needed. Metadata practices also vary, with some embedding detailed track information and others limiting metadata usage to basic file naming conventions.

In stereo mix reconstruction, all participants begin with a reference stereo mix to ensure continuity. The primary difference lies in how stereo effects are handled when transitioning to immersive formats. Some participants replicate stereo effects exactly, while others make adjustments to enhance spatialization. This suggests that while stereo reference alignment is a necessary step, individual approaches to adapting the mix for immersive playback differ.

The Atmos mixing stage shows the greatest variation among participants. The primary distinction is in bed configuration. Some participants place all tracks in beds, treating them similarly to a conventional mix. Others reserve beds only for reverb or certain effects, keeping dry elements as objects to retain flexibility. A smaller group avoids beds entirely, using only object-based placement. The choice of workflow depends on the balance between mix stability, creative control, and playback translation. Another area of variation is the use of multiple renderers, with some participants integrating additional immersive formats while others remain exclusively within Dolby Atmos.

Final checks and rendering also exhibit some divergence. While all participants test their mixes on secondary playback systems, the approach to binaural compatibility varies. Some participants create a dedicated headphone mix, while others rely on automated fold-down processing. These findings suggest that there is no single standard for how engineers ensure immersive music translates well across different playback environments.

4.1.1 Examples of Workflows

This preliminary analysis mainly focused on the workflows of mixing engineers. P7 begins with importing and aligning audio with the stereo master, then printing stems to create a mix. Side chains and plug-ins are set up, followed by EQ, compression, and limiting. Next, the loudness of the entire album is adjusted to -18 LUFS. Stems are measured, compared to the stereo master, and adjusted as needed before Atmos mixing begins. If drums are in stereo, they are duplicated and split into frequency bands (kick, snare/toms, hats) to create height. If reverb is baked into vocals, Supertone Clear is used for isolation.

P8 sets up a SoundGrid system, installs drivers on multiple computers, and configures IO for up to 256 streams. The Dolby Atmos Renderer connects to SoundGrid, with outputs assigned for 7.1.4, stereo, and binaural monitoring. If recording, an Ambisonics mic captures impulses, bypassing encoding for four discrete channels; otherwise, provided multitracks are processed. Spatialization occurs in a 7.1.4 setup with discrete outputs, ignoring the Renderer due to latency. Effects use multiple stereo reverbs routed via auxiliary returns, avoiding surround plugins. The mastering chain includes sidechain compression via auxiliary sends. The final mix is rendered as an ADM file.

P9 receives either mix-ready stems or raw multitracks. Raw multitracks require a rebuilt stereo mix, while stems are imported into a 48kHz, 24-bit Pro Tools session with Dante, Dolby Audio Bridge, binaural plugins, and preloaded reverbs. A mastered stereo reference is imported and aligned using phase reversal. Stems are grouped, arranged by instrument type, and processed with plugins. Mixing includes stereo reference listening, compression, limiting, and EQ. Objects and beds are assigned in the Dolby Renderer, with objects 1-10 for beds and others arranged front to back in pairs. Binaural settings are configured, the mix is finalized for consistency, and an ADM file is rendered.

P11 uses a standardized Pro Tools template, updated with new plugins and techniques. The template includes three reverbs, delays, crush tracks, and distortion, defining room configuration rather than specific placements. An object bed replaces the traditional bed for flexibility, with experiments in 9.1.6 and 11.1.6 for future speaker setups. Localization techniques guide

instrument placement: pads/synths use circular or back-to-front motion, guitars maintain stereo with leads off-front and reverb in the rear, and drums stay front with room/transient effects on sides or heights. Percussion is spread in height and room, while vocals use a phantom center with reverb across speakers. Backing vocals are positioned based on role, avoiding the center speaker except for Blu-ray and cinema. Mixes are tested on consumer devices for consistency before rendering the ADM file.

4.2 Categories 2 through 7

Categories 2–7 addressed challenges and innovations in immersive music production, including Dolby Atmos adoption, financial and technical barriers (P1, P2, P7–P11), and format competition (P7). Economic burdens impacted studios and artists (P2, P11), while technical issues like latency, metadata management, calibration inconsistencies, and missing stems (P7, P8, P10, P11) underscored the need for workflow standardization. The absence of formal education led participants (P4, P7, P9–P11) to rely on self-learning, mentorship, and external resources.

Collaboration helped integrate technical and artistic goals, with clear communication and accessible terminology (P1, P4, P5, P9) facilitating alignment. Networks like the Dolby Atmos Mixers Network (P9) fostered industry-wide collaboration. Creativity in immersive audio enhanced engagement (P4, P5), with participants exploring binaural audio, spatial installations, and acoustic reconstructions.

Mixing strategies varied, with custom templates, routing, and calibration used to optimize workflows (P1, P2, P7, P8, P10, P11). Some prioritized front-focused core elements (P7, P11), while others used spatial effects (P7, P8) for listener guidance. Object precision (P10), bed/discrete effect balance (P8), controlled reverb (P8, P11), and cross-device testing (P10, P11) ensured spatial clarity, while high-cost hardware (P9) remained a barrier.

Spatial audio advancements required better tools and research. Participants (P5, P6, P10) explored automation, ambisonic tools, and environmental acoustics, while collaborations with developers (P10) aimed to improve binaural rendering and Dolby Atmos workflows.

5 Discussion

This section offers a discussion of the results in relation to emerging practices.

5.1 Workflows in Immersive Music Production

The study found that participants used a structured but adaptable approach to immersive music production, generally consisting of four stages: data preparation and session setup, stereo mix adaptation (in cases where there was a previous stereo mix of the song), immersive mix refinement, and final review and rendering. While workflows differ, most professionals organize files, use session templates, and conduct iterative monitoring across different playback systems.

Variations in approach are identified in the complexity of the production, the level of technical expertise, and playback format considerations. Some professionals prefer an object-based mixing approach for control over spatial placement, while others use a hybrid or bed-based approach for efficiency. Monitoring practices also differ, with some using multiple renderers to evaluate mix translation across formats.

5.2 Common Patterns, Techniques, and Challenges

The analysis of participant workflows revealed recurring patterns in immersive music production.

- **Techniques and Practices:** Participants highlighted the importance of structured file management, consistent loudness normalization, and iterative listening across playback formats. Core elements such as vocals, bass, and drums are generally positioned at the front to maintain mix integrity, while spatial effects are applied selectively to enhance immersion. Binaural rendering is frequently used, though participants noted inconsistencies in its translation across different devices.
- **Challenges and Limitations:** Several technical and workflow-related challenges were identified. Participants reported issues with software stability, metadata inconsistencies affecting playback translation, and calibration difficulties in immersive monitoring environments. Some professionals noted that Dolby Atmos tools, particularly the

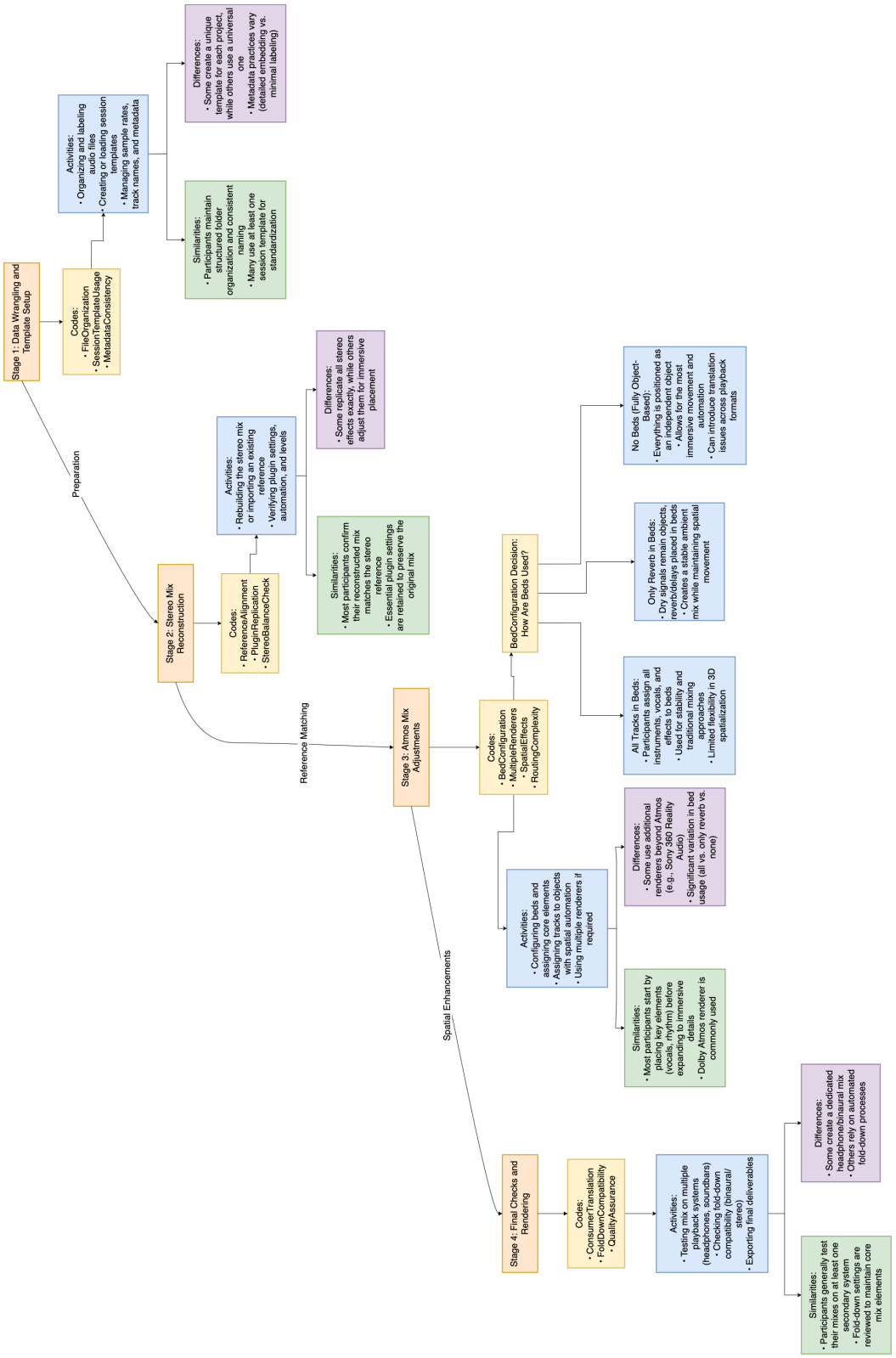


Fig. 1: Workflow Comparison

renderer, introduce latency and instability, requiring workarounds. Additionally, the financial cost of immersive audio setups, including hardware and software, was identified as a barrier for independent artists and smaller studios.

- **Collaboration and Communication:** The study highlighted the role of communication in bridging creative intent and technical execution. Engineers often translate technical concepts into accessible terms to help artists understand spatialization decisions. Some participants described using real-world analogies and demonstrations to improve artist engagement with immersive formats.
- **Access to Education and Training:** Many professionals relied on informal learning due to the absence of standardized training programs for immersive music production. Participants described self-directed learning through experimentation, mentorship from industry peers, and limited formal training from organizations like Dolby. The lack of structured educational resources has led to variation in skill levels and production consistency across the industry.
- **Future Considerations:** Participants discussed the need for improvements in immersive mixing tools, particularly in binaural rendering accuracy, automation for spatial workflows, and more reliable monitoring solutions. Some professionals suggested that advancements in AI-assisted spatial audio processing and metadata standardization could help streamline workflows. Industry-wide efforts to make immersive production more accessible, including cost-effective tools and improved educational resources, were identified as priorities for further development.

The findings indicate that immersive music production follows recognizable workflows, though individual practices vary due to technical constraints, financial limitations, and differences in training. Professionals have adopted adaptive strategies to navigate these challenges, but inconsistencies in tools, standards, and workflow methodologies remain obstacles to widespread industry adoption.

Addressing issues such as workflow efficiency, financial accessibility, and standardized training resources could support the broader adoption of immersive audio production. These insights contribute to the ongoing discussion on best practices, industry challenges, and the future development of immersive music workflows.

6 Conclusion

Immersive music production represents both a creative frontier and a technical challenge, requiring a balance between innovation and standardization. As this research highlights, industry professionals are actively developing new workflows, and the results of this preliminary analysis identify common patterns among mixing engineers. However, the field still lacks universally accepted best practices. Addressing issues such as financial accessibility, education gaps, and monitoring inconsistencies will be essential for ensuring a sustainable future for immersive music production.

This study is part of a broader research initiative aimed at developing comprehensive workflows that incorporate the roles of all professionals involved in the music production process. While the preliminary results focus on mixing engineers, future work intends to explore new workflows that involve engineers, producers, musicians, educators, and developers to provide a more holistic perspective on immersive audio practices. The findings will have implications not only for academia but also for the music industry, where they may inform the development of standardized workflows that accommodate diverse professional needs.

Furthermore, this research has the potential to benefit a range of industry stakeholders. Hardware and software developers may utilize these insights to refine their tools, commercial studios could leverage them to optimize production efficiency, and educators might apply them to bridge knowledge gaps in immersive audio training.

Acknowledgments

The research presented in this paper was funded by a Technē Doctoral Training Partnership (Workflows and Guidelines for Immersive Music Production) of the UK's Arts and Humanities Research Council (AHRC). It was also funded in part by the Engineering and Physical Sciences Research Council (EPSRC) under grant EP/X032914/1 (Challenges in Immersive Audio Technology). The authors wish to acknowledge the participants for their time, effort, and willingness to contribute to the advancement of this research.

References

[1] Smiljana Antonijevic and Ellysa Stern Cahoy. Researcher as bricoleur: Contextualizing humanists' digital workflows. *Digital Humanities Quarterly*, 12, 2018. ISSN 1938-4122.

[2] Christopher M. Baume. *Semantic Audio Tools for Radio Production*. PhD thesis, University of Surrey, 2018.

[3] Sarah Bolger. *Organisational learning from patient complaints: A case study exploring professional hierarchies and customary practices*. PhD thesis, University of Surrey, 2019.

[4] Andrew Bourbon, Matthew Watson, Dawid Stelmach, Michael Champion, Ansh Chomal, Steven Fenton, and Hyunkook Lee. An investigation into immersive contemporary popular music mixing practice through analysis of released music. In *AES International Conference on Spatial & Immersive Audio*, Huddersfield, UK, August 23–25 2023.

[5] Zack Bresler. Immersed in pop: 3d music, subject positioning, and compositional design in the weeknd's "blinding lights" in dolby atmos. *Journal of Popular Music Studies*, 33(3):84–103, 2021. ISSN 1533-1598.

[6] Sara Cohen. Ethnography and popular music studies. *Popular Music*, 12(2), 1993.

[7] C. Dewey, A. Moore, and H. Lee. Practitioners' perspectives on spatial audio: Insights into dolby atmos and binaural mixes in popular music. *Journal of the Audio Engineering Society*, 72(7/8): 504–516, Jul./Aug. 2024.

[8] Dolby Laboratories. Getting started with dolby atmos music, 2023. URL https://professionalsupport.dolby.com/s/article/Getting-Started-with-Dolby-Atmos-Music?language=en_US. Accessed: 2024-09-23.

[9] Chris Eggertsen. Apple music offers higher royalties for spatial audio streams, 2023. URL https://www.billboard.com/business/streaming/apple-music-higher-royalties-spatial-audio-streams-1235586800/. Accessed: 2024-09-01.

[10] Ulla H. Graneheim, Britt-Marie Lindgren, and Berit Lundman. Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Education Today*, 56:29–34, 2017.

[11] David E. Gray. *Doing research in the real world*. Sage, 2018.

[12] Hsiu-Fan Hsieh and Sarah E. Shannon. Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9):1277–1288, 2005.

[13] Sam Inglis. Mixing in atmos, June 2023. URL https://www.soundonsound.com/people/mixing-atmos-caroline-hilton. Accessed: 2024-09-01.

[14] Richard King and Marc Theriault. Lfe – friend or foe? the pros and cons of the low frequency effect channel. In *AES 154th Convention*, Espoo, Helsinki, Finland, May 13–15 2023. Audio Engineering Society. Convention Express Paper 104.

[15] Richard King, Marc Theriault, and George Massenburg. A practical approach to the use of center channel in immersive music production. In *AES 155th Convention*, New York, USA, October 25–27 2023. Audio Engineering Society. Convention Express Paper 168.

[16] A.J. Kleinheksel, Nicole Rockich-Winston, Huda Tawfik, and Tasha R. Wyatt. Demystifying content analysis. *American Journal of Pharmaceutical Education*, 84(1):7113, 2020. ISSN 0002-9459.

[17] Rob Lawrence. *Producing Music for Immersive Audio Experiences*, page 22. Routledge, 1st edition, 2019.

[18] Jennifer Mason. *Qualitative Researching*. SAGE Publications Ltd, London, third edition, 2018.

[19] Glenn McGarry, Peter Tolmie, Steve Benford, Chris Greenhalgh, and Alan Chamberlain. "they're all going out to something weird": Workflow, legacy and metadata in the music production process. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work*

and Social Computing (CSCW '17), pages 995–1008, New York, NY, USA, 2017.

[20] Glenn McGarry, Alan Chamberlain, Andy Crabtree, and Christopher Greenhalgh. The meaning in "the mix": Using ethnography to inform the design of intelligent tools in the context of music production. In *Proceedings of Audio Mostly 2021 (AM '21)*, pages 40–47, New York, NY, USA, 2021. ACM.

[21] Sean McGrath and Steve Love. The user experience of mobile music making: An ethnographic exploration of music production and performance in practice. *Computers in Human Behavior*, 72: 233–245, 2017.

[22] Sean Olive. The perception and measurement of headphone sound quality: What do listeners prefer? *Acoustics Today*, 18(1):58–67, March 2022.

[23] Karen OReilly. *Key Concepts in Ethnography*. SAGE Publications Ltd, London, 2012.

[24] Demesia Razo. *An Exploration of Professional Training and Professional Practice: Title IX Administrators and Meaning Making*. PhD thesis, University of North Texas, 2019.

[25] Micah Singleton. Apple music's eddy cue talks spatial audio, lossless and the future of music, August 2021. URL `https://www.billboard.com/pro/apple-music-eddy-cue-spatial-audio-lossless-future-of-music/`. Accessed: 2024-09-23.

[26] Rosanna C. G. Smith and Stephen R. Price. Modelling of human low frequency sound localization acuity demonstrates dominance of spatial variation of interaural time difference and suggests uniform just-noticeable differences in interaural time difference. *PLoS ONE*, 9(2):e89033, 2014.

[27] Harsh Suri. Purposeful sampling in qualitative research synthesis. *Qualitative research journal*, 11(2):63–75, 2011.

[28] Universal Music Group. Dolby atmos audio best practices, 2024. URL `https://contentguide.universalmusic.com/dolby-atmos-audio-best-practices/`. Accessed: 11-Feb-2025.

[29] Bosun Xie and Guangzheng Yu. Psychoacoustic principle, methods, and problems with perceived distance control in spatial audio. *Applied Sciences*, 11(23):11242, 2021.

[30] Xiangyang Zeng, Huiying Ma, and Haitao Wang. Localization performance under middle and low frequency sound source based on time reversal method in enclosed space. *IEEE Access*, 7:5649–5661, 2019.

Appendix A

Table 1: Participant Criteria (Rows in gray indicate participants who were included in the analysis presented in this paper)

Participant	Job Role	Professional Practice	Technical Expertise	Industry Integration	Technological Influence	Resource Access	Industry Recognition	Educational Background	Production Complexity	Observed
Participant 1	Record Label Executive	Yes	Advanced	Yes	No	Yes	Yes	Postgraduate	Standard	No
Participant 2	Artist Manager	Yes	Beginner	No	No	Yes	No	Undergraduate	Standard	No
Participant 3	Artist	Yes	Beginner	Yes	No	Yes	No	Undergraduate	Standard	No
Participant 4	Artist	Yes	Intermediate	No	No	Yes	No	No formal education	Complex	No
Participant 5	Researcher	Yes	Expert	Yes	Yes	Yes	No	PhD	Complex	No
Participant 6	Researcher	Yes	Expert	No	Yes	Yes	Yes	PhD	Complex	Yes
Participant 7	Audio Engineer	Yes	Expert	Yes	Yes	Yes	Yes	Undergraduate	Complex	Yes
Participant 8	Audio Engineer	Yes	Expert	Yes	Yes	Yes	Yes	Self-taught	Highly Complex	Yes
Participant 9	Audio Engineer	Yes	Expert	Yes	Yes	Yes	Yes	Self-taught	Standard	No
Participant 10	Audio Engineer	Yes	Expert	Yes	Yes	Yes	Yes	Self-taught	Complex	No
Participant 11	Audio Engineer	Yes	Expert	Yes	Yes	Yes	Yes	Postgraduate	Standard	No
Participant 12	Artist	No	Beginner	No	No	Yes	No	Self-taught	Standard	No
Participant 13	Artist	Yes	Advanced	No	No	Yes	No	Postgraduate	Complex	No
Participant 14	Artist/Record Label staff	Yes	Advanced	Yes	No	Yes	No	Postgraduate	Standard	No
Participant 15	Audio Engineer	Yes	Advanced	No	No	Yes	No	Undergraduate	Standard	No
Participant 16	Audio Engineer	Yes	Expert	No	Yes	Yes	No	Postgraduate	Complex	No
Participant 17	Audio Engineer	Yes	Advanced	No	No	Yes	No	Self-taught	Standard	No
Participant 18	Audio Engineer	Yes	Intermidiate	No	No	Yes	No	Self-taught	Standard	No
Participant 19	Audio Engineer	Yes	Advanced	No	No	Yes	No	Self-taught	Standard	No
Participant 20	Audio Engineer	Yes	Expert	No	No	Yes	No	Undergraduate	Standard	No
Participant 21	Audio Engineer	Yes	Advanced	No	No	Yes	No	Self-taught	Standard	No
Participant 22	Audio Engineer	Yes	Expert	Yes	Yes	Yes	Yes	Undergraduate	Complex	No
Participant 23	Audio Engineer	Yes	Expert	Yes	No	Yes	No	Self-taught	Complex	No
Participant 24	Audio Engineer	Yes	Advanced	No	No	Yes	No	Self-taught	Standard	No
Participant 25	Audio Engineer	Yes	Expert	No	No	Yes	No	Undergraduate	Complex	No
Participant 26	Audio Engineer	Yes	Advanced	No	No	Yes	No	Undergraduate	Standard	No
Participant 27	Audio Engineer	Yes	Advanced	No	No	Yes	No	Undergraduate	Standard	No
Participant 28	Composer	Yes	Advanced	No	Yes	Yes	No	Self-taught	Complex	No
Participant 29	Composer	No	Beginner	No	No	No	No	Self-taught	Standard	No
Participant 30	Researcher	Yes	Expert	No	No	Yes	No	Postgraduate	Standard	No