Practitioners' Perspectives on Spatial Audio: Insights into Dolby Atmos and Binaural Mixes in Popular Music

Christopher Dewey,[1], Austin Moore,[1] & Hyunkook Lee,[2]

Abstract

This paper presents the practitioners' perspective on mixing popular music in spatial audio, with particular emphasis on Dolby Atmos and binaural mixes generated by the renderers offered by both Dolby and Apple. It presents the results of a dual-stage study, which utilized focus groups with nine professional music producers and a questionnaire completed by 140 practitioners, all engaged in mixing music in spatial audio formats, particularly in popular music styles. Analysis of the focus groups and questionnaire revealed the continued influence of stereo approaches on mix engineers, partly due to its historical dominance as a production platform and consumers' continued use of headphones. It was also found that core elements of popular music productions, such as snare drums, tom-tom drums, kick drums, bass guitars, main guitars, and vocals, were less likely to have binaural processing applied (near, mid, and far in Atmos) compared with other sources. It was also shown there were perceived differences in the suitability of spatial audio mixing for specific genres, with Electronic Dance Music, Jazz, Pop, Classical, and World Music rated as the most suitable. Regarding the binaural renderers, there was less user satisfaction with the Apple device compared with Dolby's, and this dissatisfaction manifested mainly regarding the need for more user control. Finally, mix engineers were very aware of the importance of their mixes translating to smaller speaker systems and headphone playback, in particular. The results of this study serve as crucial information in understanding the professional music producer's perspective on spatial audio mixing of popular music and contribute to the ongoing development of spatial audio in popular music production.

1 INTRODUCTION

Despite being used in cinema and academic music composition for many years, spatial audio, and in particular Dolby Atmos and binaural versions of Atmos mixes rendered by the Dolby and Apple renderers, have only recently been embraced by the music industry as a viable playback format for popular music [1, 2]. The slow adoption of spatial audio by the music industry may be due to the challenges faced by Quadraphonic sound and 5.1 surround in the past, particularly regarding compatibility issues and speaker setups, which could have been more practical for many domestic listeners [3]. Furthermore, the rise of mp3 files during the early 2000s and the inclusion of music player apps in mobile phones in the mid 2000s shifted the focus further away from all forms of loudspeaker playback for the typical music consumer. Given this newfound interest in spatial audio the present study explores the impact of spatial audio music production on the work of professional mix engineers. Thus, a questionnaire was developed and circulated among professional mix engineers to understand their approaches to mixing, their workflow, and any perceived limitations with the current commercial systems.

To date, spatial audio research in music production has predominantly focused on acoustic recordings and psychoacoustics, and this is shown in the selected works cited here [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. It is worth noting that these studies encompass a broad spectrum of spatial audio formats and are not exclusively limited to Dolby Atmos mixes. Literature specifically on spatial audio in popular music production is minimal, with only a few studies identified during the review. Of note is the work by Lawrence [14] that provides an overview of formats, tools, and workflows for immersive audio production and highlights the benefits and opportunities of using height channels, multiple loudspeaker arrays and binaural

headphone applications. Malecki et al. [15] conducted a case study of mixing electronic dance music by creating an Amibsonics remix and binaural rendered version of a stereo production. Their findings highlighted how the spatial mixes differed not only in terms of the spatial attributes of the mix but also in the timbral presentation and dynamics. On a similar theme, Barboza et al. [16] highlighted the creative use of spatial audio in a Brazilian music production that was remixed using 3rd order Ambisonics, and their results showcased best practices in positioning sound sources in space. Sotgiu et al. [17] explored the listener's preference and sense of presence between stereophonic and binaural music and noted that a lack of standards in mixing practice was one of the challenges highlighted in their results (something the current study hopes to address). Finally Morell and Lee [18] discovered through a listening experiment that binaural processing on all musical sources in popular music mixes did not enhance immersion for the end listener. Of particular interest is they noted that the optimal use of binaural processing may depend on the music genre (again, something this current study hopes to address).

As can be seen, to the best of the authors' knowledge the current literature, particularly regarding popular music in spatial audio, could be much more extensive. To address this gap the present study will investigate spatial audio through the lens of experienced mix engineers to better understand their needs, preferences, and approaches to popular music production in spatial audio, specifically Dolby Atmos and binauralized mixes. This approach will likely identify potential enhancements in spatial audio technology and music production workflow to ensure that advancements in the field align with mix engineers' practical needs and creative aspirations and ultimately result in better quality spatial audio music productions for the consumer.

The paper adheres to the following structure. Firstly, it explores the initial steps in developing a questionnaire, which is a netnographic study, followed by an analysis of focus group discussions. Next, five hypotheses related to the literature review, netographic study and focus groups are presented to the reader. The paper then discusses the construction of the primary research tool, a questionnaire, to detail its design and sampling and distribution methods. In the final sections the questionnaire results are presented and discussed by exploring them as they pertain to the five hypotheses. As can be seen, the paper is structured so the reader can map the development of the study and easily make associations between each attribute of the work.

2 Netographic Study

In preparation for the focus groups, a short netographic study was conducted to gather information from online sources discussing spatial audio music production. This was necessary as the existing literature on spatial audio in popular music production and spatial audio music production was limited, and additional research was needed to develop the questionnaire. The findings from the netographic study informed the conceptual framework for the upcoming focus groups with professionals.

The study explored online material from the popular music production forum Gearspace [19] and video interviews with music producers engaging in spatial audio production [20, 21, 22] Netographic techniques have been used in previously in music production studies [23, 24, 25] and is a valuable method when the discourse of a particular area is mainly active within online communities. However, as is the case with the present study, netography is often used as part of a more extensive study [26] and, in this instance, was used as the 'front end' for the questionnaire's development, offering an unintrusive and immediate way to get a feel for the practitioners' perspective. The material was explored for patterns and themes using thematic analysis and elements of Grounded Theory. In this inductive qualitative method, a theory emerges from observations made during the coding process [27]. The use of grounded theory and related thematic analysis can be found in related literature [28, 29, 30] and the coding process used in the present study is described in detail in previous work [23]. The qualitative software NVivo was used for textual coding.

Emerging from the coding, the following five themes were identified:

- **Translation and Playback**- Specifically how Dolby Atmos decodes and stereo compatibility.
- **The Perceived Quality of Spatial Audio Mixes**-Including both original Atmos and its binaurally rendered versions.
- **Approaches to Production and Mixing**-Specifically the technical aspects of producing a mix as well as creative decisions, particularly the use of height channels and stems
- **Technical Requirements**- Pertaining to speaker configurations and audio levels for mastering Atmos mixes.
- **Music Business Related**-In particular, conversations around fees, probing whether spatial

audio in music production is on the cusp of becoming a significant trend or another element of marketing trickery.

The analysis from this study and the review of relevant spatial audio literature informed the development of the next stage, a series of focus groups with professional mix engineers.

3 Focus Group Discussions (FGDs)

Three focus group discussion (FGD) sessions were conducted that followed a semi-structured interview approach [31]. The discussions included the practitioners Andrew Scheps, Hans-Martin Buff, Adrian Hall, Nuno Fernandes, Mike Exeter, Gary Bromham, Phil Harding and Lachlan Goold. All participants had over 20 plus years of mixing experience, the majority with direct experience of mixing Atmos while the rest expressed a keen interest in this area. The authors of this paper also contributed to the discussions to keep the sessions focused and provide underpinning academic theory, particularly on psychoacoustics. The FGDs followed a semi-structured interview format, with topics from the literature review, netographic study and the authors' prior knowledge of the subject. Audio recordings of the sessions were made, transcribed to text and explored using the same method as described previously. In the third and final session, the results of the previous two sessions, which are presented below, were shown to the practitioners and shared via email to those who could not attend. The purpose was to get the participants' approval that the key areas of conversation had been captured in the thematic analysis and were indeed the salient areas in spatial audio music production. Developing questions through interaction with the end user is a similar method used by Zattra and Donin [32] and beneficial when academic literature is limited. All the participants approved that the analysis of the results accurately represented the FGDs.

3.1 Results of FGD 1

During the first FGD, the three core themes of discussion were **production and mixing** (75% of the discussion), **binaural playback** (21% of the discussion) and **business considerations** (4% of the discussion). Table 1 shows the core themes with their associated sub-themes.

The use of height and rear speakers was discussed, particularly regarding the perceptual effects and instrument choices. The general feeling was that not all instruments should be placed outside of the traditional stereo field and that there was perhaps an over expectation of what the height channels could afford the mixer. A typical approach, if appropriate for the mix, was to use the height channels for ambience and space and distribute layered recordings (pairs of double musical sources for example) around the height and rear channels.

Stem mixing emerged as an important topic, with mix engineers noting many spatial productions involved using pre-processed stems from the original stereo mix; this in turn led to the theme around how stereo mixing techniques are used in spatial mixes and if changes in mixing approach is necessary.

The role of mastering was introduced as at the time it was unclear to the participants how the spatial mixes were to be mastered; within this sub-theme there was debate around loudness unit (LU) levels and how they differed from stereo.

Additionally, the influence of music genres, especially regarding the creative possibilities in spatial audio mixing, was examined. It emerged that electronic music styles are more conducive to creative spatial mixing compared to traditional rock and pop band formats. Notably, these observations primarily relate to music productions originally intended for stereo reproduction. However, it was acknowledged that productions designed for spatial audio from inception could present a different paradigm.

The sound quality of the binaural renderers was an important theme, with participants not being satisfied with the overall timbral and spatial quality. Connected to this was the sub-theme headtracking and the role this may play on the perceived quality of localization in binaurally rendered mixes. A related sub-theme was the perceptual impact of binaraulizing instruments, and the general opinion was not to binarualize key elements in pop/rock mixes such as the vocals, snare drums, bass drums or bass instruments. It was felt that binaural rendering resulted in these sources losing impact and transient detail, which from a stereo mixing perspective is important. In fact, this idea is important, as there was a general feeling, over all focus group discussions that the spatial audio mixes should be an enhanced stereo mix; the terms "stereo plus" and " a larger version of stereo" were used by the participants.

The discussion briefly explored the innovative use of binaural plugins in audio mixing. One participant highlighted their use of these tools, emphasizing that they serve not only for spatial audio rendering but also as creative aids in the mix. This approach allows for the addition of unique spatial dimensions and

Table 1: Focus Group 1 Themes and Sub-Themes

Production and Mixing	Binaural Playback	Music Business
Use of Height Speakers	Renderer Quality	Mix Approval
Use of Rear Speakers	Binauralizing Music Sources	Back Catagloue Mixes
Stem Mixing	Binaural Plugins as Creative Processors	
The Impact of Genre on Mixing	Head Tracking	
The Role of Mastering		
Relevancy of Stereo Mixing Techniques		

artistic effects, enriching the overall sound in music production.

Regarding business considerations, there was talk around mix approval from labels, particularly artists; dislike of binaural mixes. There were also comments regarding mixes of back catalogue material and how they constituted a significant portion of one mixers work; therefore, they noted that the original stereo mixes were still held as the benchmarks of success when comparing the spatial mix.

3.2 Results of FGD 2

The themes emerging from the second FGD session included once again **binaural playback** (48% of the discussion), **production and mixing** (43% of the discussion) and **business considerations** (9% of the discussion). However, as can been seen in Table 2, the proportion of time spent on each topic shifted from the first FGD.

From the production and mixing theme, discussions surrounded the use of typical stereo mix bus processing techniques (stereo compressors, limiters, stereo EQ) and their suitability in spatial audio mixing. Mix translation involved a conversation around how mixes would translate onto other systems and mixing with headphones concerned the issue of whether spatial mixes could be completed almost entirely over headphones and if those mixes would translate appropriately over speaker systems. The spatial mix from stereo mix theme centred around the issue of trying to match the original stereo mix sound quality as a starting point for the spatial mix; the motivation for this being to provide a familiar audio experience for the listener, so their expectations were managed. This led to a conversation about creativity in spatial mixes and how sounds could be positioned and processed without affecting the musical intentions of the production, which is usually conceived in stereo. From the binaural playback theme, limitations of the current binaural renderers (Dolby and Apple) were again discussed at length, encompassing issues regarding the accuracy of localization and externalization.

3.3 Focus Group Consensus

In the third and final focus group meeting, the combined results of the previous two sessions were shown to the practitioners and shared via email to those who could not attend. The purpose was to get the participants' approval that the key areas of conversation had been captured in the thematic analysis and were indeed the salient areas in spatial audio music production. Developing questions through interaction with the end user has been used in previous studies [32] and is beneficial when academic literature is limited. All the participants were happy with the analysis of the results and approved that they accurately represented the focus group discussions.

4 Development of Constructs and Hypotheses

The results from the review of literature, netographic study and focus group discussions led to the creation of seven constructs, **speaker configurations, mix translation, mix creativity, renderer quality, renderer control, adherence to stereo conventions, and genre suitability** . These constructs were the basis for the development of the following hypotheses for the survey conducted.

- **Hypothesis 1**- When working in spatial audio, mixers tend to not binauralize elements that would typically be presented as a phantom center image.
- **Hypothesis 2**- Spatial audio suitability varies among music genres, influenced by their technical characteristics and consumer expectation.
- **Hypothesis 3**- Mix engineers have a negative opinion of the current binaural renderers. Furthermore, there is a difference in perceived sound

Table 2: Focus Group 2 Themes and Sub-Themes

Production and Mixing	Binaural Playback	Music Business
Mix Buss Processing	Virtual Reality and Binarual	Mix Approval
The Role of Mastering	Room Divergence Effect	
Mixing with Headphones	Headtracking	
Speaker Configurations	Renderer Quality	
Mix Translation	Renderer Control	
Creative Approaches to Atmos		
Spatial Mix versus Stereo Mix		

quality between the two most common renders in music production, the Dolby and Apple renderers.

- **Hypothesis 4**- Mix engineers adopt different approaches to audio processing when mixing in spatial audio.
- **Hypothesis 5**- Mix engineers are concerned with how their spatial mix will translate over smaller loudspeaker systems and headphones.

5 Survey

The following sections describe the questionnaire and sampling method used for the survey, and present and discuss the results of the survey thematically. For each hypothesis an overview of the results is presented and where appropriate, statistically analysed. Finally, the results are discussed, relative to the themes and hypotheses that were drawn from the two focus groups.

5.1 Questionnaire

Table 6 shows the questionnaire used for the survey. The majority of the items were developed around the seven constructs and five hypotheses detailed above. Additional related questions were also included for use in a future study.

5.2 Sampling method

To distribute the questionnaire, we used judgement sampling [33] and contacted mix engineers within our professional and academic networks who were currently working in spatial audio. Several engineers passed the questionnaire on to their contacts, utilizing snowball sampling. Additionally, professionals were contacted directed by email through the Dolby Atmos music studio list, a resource collated by Dolby that included 552 Dolby Atmos-equipped professional studios. All 552 studios were contacted from this list. Finally, links to the questionnaire were posted on social media and internet forums, but the uptake from this method was much lower than the direct email approach.

Table 3: Age group of respondents

Age Group	n
18-24	3
25-34	15
35-44	20
45-54	37
55-64	20
Prefer not to say	5

5.3 Demographics

259 practitioners engaged with the survey, with a completion rate of 54% (n = 140). The majority of respondents were male. Table 3 presents the age groups of the respondents and shows that over one third of the participants were aged between 45 and 54. Over 80% of the respondents had over 10 years experience of producing music. 67% of the respondents reported having produced over 20 Atmos mixes. These results support the view that the questionnaire was successful in recruiting professional music producers who were experienced in spatial audio based music production.

5.4 Results From Hypothesis 1

To explore the hypothesis that when working in spatial audio, mixers tend to not binauralize elements that would typically be presented as a phantom center image, the respondents were asked to assess the propensity to binauralise a range of musical sources typically found in popular music productions using a 5 point Likert scale (1: Never, 2: Rarely, 3: Sometimes 4: Often, 5: Always).

The musical sources that are typically presented as a phantom centre image include kick drum, bass instrument, snare drum and lead vocal. Arguably these four sources can also be considered as the principal elements in any given popular music production. Therefore, it was hypothesised that these four 'core' sources would deviate the least from conventional stereo mixing practice compared with other musical sources typically used in popular music production. The null hypothesis was all musical sources had an equal propensity for binauralisation.

Figure 1 presents the results for this question as a bar chart. As a 5 point Likert scale was used, the median ratings for each musical source are presented. Due to the ordinal nature of the data a test for normality was not performed. Subsequently a non-parametric Friedman test was conducted to assess whether there were significant differences in the distribution of ratings between the musical sources. There was a statistically significant difference between the musical sources ratings $\chi 2(19) = 397.81$, $p = > .05$, meaning the null hypothesis could be rejected. Pairwise comparisons of the musical sources assessed were performed with a Bonferroni adjustment. Table 4 in the Appendix presents an overview of the statistically significant differences. These results show that the snare drum, tom drum and kick drum, bass guitar, main guitar and vocal sources are less likely to be binauralised compared with the other sources assessed. With the addition of the tom drum and main guitars, this supports the view that the most important musical sources in a popular music production mix are less likely to deviate from standard stereo positioning compared with other, less important sources.

In terms of binauralisation, it is likely that the engineer would not want to add any additional ambience to the kick drum and bass instrument because both sources are low frequency dominant and provide a foundation to the mix. Arguably, this is also true for the snare drum as additional ambience may detract from the impact of careful designed reverb treatment by the mix engineer. Similarly, the vocals are typically the most important track in a popular music mix and are also likely to have genre specific time-based effects already applied. The result for main guitars and tom drums is interesting because these two sources were not specifically identified by the focus group participants as being musical sources that should be presented as a phantom centre image, due to the likelihood they would be panned off-centre. These results might be explained by a desire to preserve the transient detail, which is an important attribute in modern articulate guitar tones and impactful drums.

5.5 Results From Hypothesis 2

A series questions were posed to explore the hypothesis that certain genres are more suited to spatial formats than other genres. The null hypothesis was all popular music genres are equally suited to spatial formats.

Firstly, the respondents were asked to rate the suitability of a range of musical genres for Atmos and binaural playback respectively using a continuous scale in two separate questions. The continuous scale featured a label at either end (0: Very Unsuitable, 100: Very Suitable) and no additional markings. The raw genre ratings responses were normalised per participant to afford a relative comparison of genre suitability between the respondents. 26 respondents rated all genres 100% suitable for Atmos playback and 16 respondents rated all genres 100% suitable for binaural playback. To preserve this equally positive sentiment, in these instances the normalised value was 1. Figure 2 presents the mean normalised suitability ratings for all genres assessed as a bar chart with 95% confidence intervals.

For five of the genres assessed (EDM, Pop, Jazz, Classical and World Music) there were outliers present in the data below the lower quartile. This suggests conflicting opinions between the respondents. Given the subjective nature of the question, the outliers were not removed prior to statistical analysis using a repeated measures one-way ANOVA. A Shapiro-Wilk test showed that normality was violated for all thirteen genres ($p<.05$). As the direction of skew was inconsistent between genres, the data was not transformed prior to analysis. Mauchly's test of sphericity indicated sphericity had been violated, $\chi^2(77) = 248.887$, $p < .001$. Epsilon (ϵ) was .636, as calculated according to Greenhouse and Geisser (1959) and was used to correct the one-way repeated measures ANOVA. There was a statistically significant difference in normalised Atmos suitability ratings between the genres, $F(7.632, 717.454) = 27.308$, $p < .001$.

The post hoc tests show that the respondents deemed EDM, Pop, Jazz, Classical and World Music as the genres most suited to the Atmos format compared with the other genres assessed. Rock, Metal and Alt Rock were deemed the least suited to the Atmos format.

The normalised suitability genre ratings for the binaural renderer also shows outliers below the lower quartile for the genres EDM, Pop, Jazz, Classical and World Music genres. A Shapiro-Wilk test showed that normality was violated for all thirteen genres

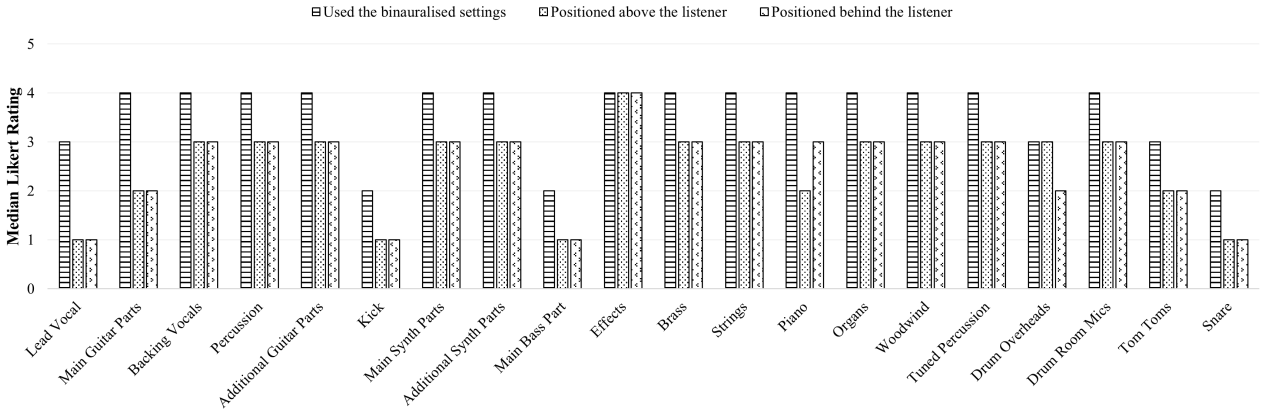

Figure 1: Bar chart showing the median ratings for the three questions that considered the frequency with which the respondents spatialised a range of musical sources typically found in popular music mixes. The median ratings for all three questions which consider the propensity for musical sources to be spatialised are presented together for ease of comparison.

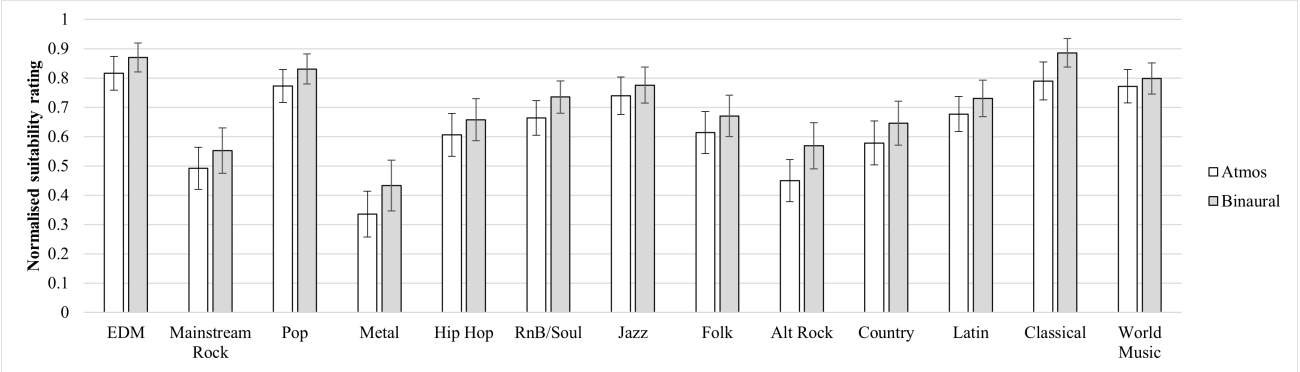

Figure 2: Bar chart of normalised genre suitability ratings for binaural and Atmos rendered mixes presented with 95% confidence intervals

(p<0.05). As the direction of skew was inconsistent between genres, the data was not transformed prior to analysis. Mauchly's test of sphericity indicated sphericity had been violated, $X^2(77) = 237.251$, $p <$ .001.

A repeated measures one-way ANOVA was conducted to statistically assess differences in the genre suitability ratings. Epsilon (ϵ) was .656, as calculated according to Greenhouse and Geisser and was used to correct the one-way repeated measures ANOVA. There was a statistically significant difference in normalised binaural suitability ratings between the genres, $F(7.874, 708.674) = 29.076$, $p < .001$. Table 5 in the Appendix presents an overview of the statistically significant differences between the genre ratings. Again these results show that the EDM, pop, classical and world music genres were most suitable and the rock derived genres were least suitable for binaural formats, relative to the genres assessed.

These results confirm the null hypothesis could be rejected for both playback formats. It is likely that the respondents were basing their suitability ratings for the two spatial formats on songs that typify the genre. Specific songs and sub-genres within each category may be more or less suited to this immersive format, and therefore the discussion of these results generalise based on typical characteristics of each genre. EDM and Pop genres are arguably more suitable for spatial formats because they are more synthetic in timbre and include extensive use of effects processing and features such as 'risers' and 'drops'. In contrast, mixes in the classical and world music genres recreate a live performance of acoustic instruments in a real-world acoustic space; a goal that lends itself to spatial formats. The negative result for the Rock and Metal genres likely relates to the perceived 'heaviness' and 'aggression' typified by this style of music requiring a mix that is 'upfront' and spatially dry.

To explore the genre suitability hypothesis further the respondents were also asked to rate their agreement to two statements for the same music genres using a five-point Likert scale (1: Strongly Disagree, 2: Disagree, 3: Neither Agree nor Disagree, 4: Agree, 5: Strongly Agree). The two statements were based on the opinions expressed in the two focus group discussions:

1. An Atmos mix should be clearly derived from the original stereo mix.

2. An Atmos mix should be completely new.

For both statements, most of the genres assessed show a lack of consensus between the respondents, with around 40% of respondents selecting the neutral rating, 30% selecting the two highest agreement ratings and 30% selecting the two lowest agreement ratings. A Friedman test showed that for both statements there was a statistically significant difference is the distribution of agreement ratings between the genres (Statement 1: $X^2(13) = 41.564$, $p < .001$, Statement 2: $X^2(13) = 94.070$ $p < .001$). Post hoc pairwise comparisons were performed with a Bonferoni adjustment for multiple comparisons. There was a significantly higher agreement rating that back catalogue mixes should be clearly derived from the original stereo mix compared with Jazz ($p = .008$), Classical ($p = .015$), World Music ($p = .016$), Folk ($p = .020$), EDM ($p = .024$) and Latin ($p = .028$).

It is likely that this opinion is because for all back catalogue mixes there exists a stereo reference mix (the original release). Therefore, this prior artefact is likely to provide a reference upon which the spatial audio mix should be based, which was a point raised during the FGDs.

Validating this assertion, when asked how frequently their Atmos mixes are based on pre-processed stems derived from a stereo mix 64% of respondents selected the highest two frequency ratings (4: Often, 5: Always). Comparatively, only 46% of respondents selected the highest two frequency ratings when asked how often they created their Atmos mixes from the raw multitrack. This supports the view that if a stereo mix already exists, the spatial mixes are likely to be based on this original artefact.

The post hoc tests for Statement 2 validate these results further. There was a significantly higher agreement rating that Atmos mixes for Latin (p =.037), Jazz (p =.010), EDM ($p = .009$), World Music (p =.005) and Classical ($p = .003$) genres be completely new compared with back catalogue mixes.

The post hoc results for Statement 2 also showed there was a significantly higher agreement rating that Classical mixes be completely new compared with Metal ($p = .038$) and Rock ($p = .037$).

Combined, these results support the view that some genres are more suited to spatial audio formats than others.

5.6 Results From Hypothesis 3

The respondents were asked to rate their satisfaction using the Dolby and Apple renders in terms of tonal and spatial quality and control using a five-point Likert scale (1: Extremely Dissatisfied, 2: Dissatisfied, 3: Neither Satisfied or Dissatisfied, 4: Satisfied, 5: Extremely Satisfied). A summary of the responses is shown in Figure 3.

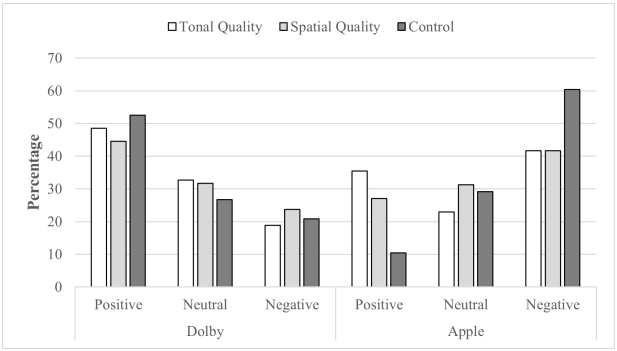

Figure 3: Bar chart of respondents satisfaction ratings for the Dolby and Apple renderers

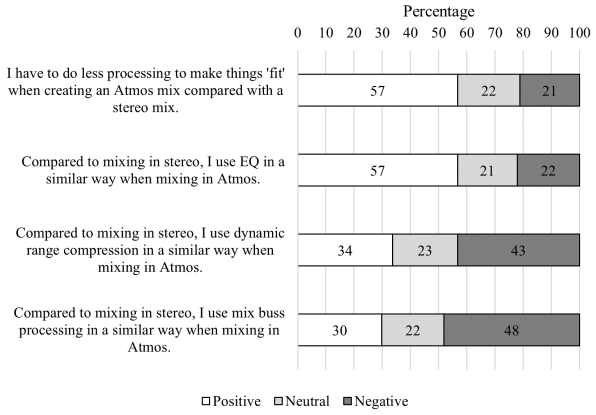

Figure 4: Respondents agreement ratings to the four mix processing statements considered

For all three metrics assessed, the respondents are $\leq$ 24% dissatisfied with the Dolby renderer and $\leq$ 61% dissatisfied with the Apple renderer, with most dissatisfaction being shown for the amount of user control offered by the Apple renderer. It is worth noting that this is the only attribute which received a majority dissatisfaction score. Thus, from the percentage of agreement and disagreement, it can not be inferred that participants are dissatisfied with the renderers, except regarding the control offered by the Apple render.

Participants selecting the two highest satisfaction ratings for each of the three attributes suggests that they are more satisfied with the Dolby renderer than the Apple renderer. To assess this, a Wilcoxon Signed Ranks Test was conducted for each of the three metrics. For tonal quality the two renders were rated comparably, with no significant difference in the median ratings, z = -1.735, p = .083. For spatial quality, the median rating for the Dolby renderer (4.0) was significantly higher than the Apple renderer (3.0), z = -1.970, p = .049. For control, the median rating for the Dolby renderer (4.0) was significantly higher than the Apple renderer (2.0), z = -4.653, p < .001. Therefore, we can reject the null hypothesis for control and spatial quality but not tonal quality. For clarity, it should be noted that the terms tonal and timbral quality were not explicitly defined in the questionnaire because the authors assumed that this was common vernacular within professional music producers. Future work should ensure these terms are explicitly defined to remove the potential for any ambiguity. Participants of the questionnaire expressed higher satisfaction with the Dolby renderer compared with Apple's and their dissatisfaction with most marked with the lack of control over the Apple renderer.

5.7 Results From Hypothesis 4

To explore the extent to which stereo mixing approaches inform the creation of an Atmos mix, the respondents were asked to assess the frequency with which they would first create a stereo mix and then spread out the musical sources to other locations when mixing in Atmos. When answering this question only 23% of respondents selected Never. In contrast, 41% of respondents selected Often or Always. This result suggests that stereo mixing approaches can be considered influential in the initial formation of the Atmos mix.

Four statements were posed to the respondents that considered different aspects of processing when mixing in Atmos, compared with mixing in stereo. The respondents were required to provide agreement ratings using a five-point Likert scale (1: Strongly Disagree, 2: Disagree, 3: Neither Agree nor Disagree, 4: Agree, 5: Strongly Agree). Figure 4 presents the results for all four statements as stacked barcharts.

The results show that overall the respondents believed that they had to do less processing to make the instruments fit when mixing in Atmos compared with stereo mixing. When comparing approaches to processing, over half of the respondents agreed that their approach to EQ processing was similar when mixing in Atmos. There was less consensus between the respondents when considering their approach to dynamics and bus processing with around half of the respondents disagreeing with the two statements and only a third agreeing with these two statements. Future work could explore the reasons for these differences in approach.

Additionally, participants were asked to rate how

frequently they would position a number of musical sources above or behind the listener when mixing in Atmos. Figure 1 shows the median rating for each musical source assessed. A Friedman test was conducted to statistically assess differences between the musical sources ratings ($p < .05$), meaning the null hypothesis could be rejected.

Pairwise comparisons of the musical sources assessed were performed with a Bonferroni adjustment (see Table 4). The results of the pairwise comparisons show that the snare drum, tom drum and kick drum, bass guitar, main guitar and vocal sources are less likely to be positioned above or behind the listener compared with the other sources assessed. Akin to the results presented in Section 4.4, this supports the view that the most important musical sources in a popular music production mix are less likely to deviate from standard stereo positioning compared with other, less important sources. The results also show that Effects, Backing Vocals and Additional Synth are the most likely candidates to be spatialised above or behind the listener.

5.8 Results From Hypothesis 5

To explore the hypothesis that mix engineers are concerned with how their spatial mix will translate over smaller loudspeaker systems and headphones, the respondents were asked a number of questions relating to mix translation and audio playback systems. The two most common Atmos mixing configurations were 7.1.4 and 9.1.4, with 69% respondents using 7.1.4 and 18% of respondents using 9.1.4.

The respondents were asked to rate the importance of checking their Atmos mixes over playback systems with a greater number of speakers using a five-point Likert scale (1: Not Important at all, 2: Slightly Important, 3: Moderately Important, 4: Important, 5: Extremely Important). Overall, the respondents did not deem the act of checking mixes on 'bigger' systems necessary with only 13% of the respondents selecting the two highest importance ratings and 58% selected Not Important at all. In contrast, 59% of respondents considered the act of checking their Atmos mixes on systems with fewer speakers to be Extremely Important.

The respondents were also asked to rate a range of commercial playback systems in terms of mix translation importance, using the five-point Likert agreement rating scale presented above. Figure 5 presents the median importance ratings for the playback systems assessed.

A Friedman test was conducted which revealed a statistically significant difference in the importance ratings between the playback devices, $X^2(5) = 191.18$, $p = > .05$. Pairwise comparisons with Bonferoni adjustments for multiple comparisons show that Headphones (Mdn = 5.0) and Atmos Loudspeaker systems (Mdn = 5.0) were deemed to be the most important playback systems in terms of mix translation. Headphones with headtracking, Atmos soundbars and stereo playback systems (Mdn = 4.0) are comparable and of secondary importance. The rating for Headphones was significantly higher than the ratings for Headphones with headtracking, Mono and Stereo playback systems ($p < .001$). The rating for Atmos Loudspeaker systems was significantly higher than Headphones with headtracking, Atmos soundbars and both Mono and Stereo playback systems (($p < .001$). Mono playback systems were deemed the least important overall, with ratings (Mdn = 2.0) significantly lower than all other playback systems assessed (($p < .001$).

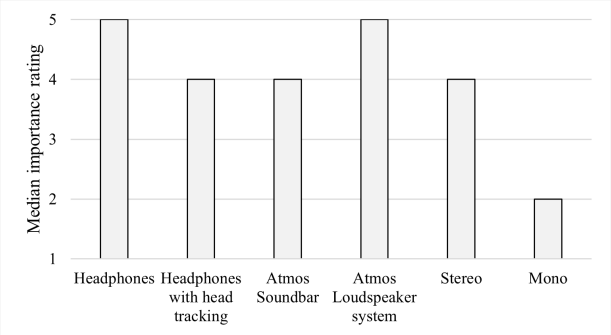

Figure 5: Median Likert importance rating for mix translation to a common range of audio formats

Overall, the respondents deemed headtracking to have neither a positive nor negative impact on their mix with 42% of respondents selecting this rating. There were mixed opinions on the importance of a personalised HRTF on music mixes with 42% of respondents selecting the highest two importance ratings and 31% selecting the two lowest importance ratings.

6 SUMMARY

This paper employed a three-step methodology: netographic study, focus group discussions (FGD), and questionnaire. The netographic investigation documented online discourse regarding spatial audio music production, highlighting five prominent themes: 'Translation and Playback,' 'The Perceived Quality of Spatial Audio Mixes,' 'Approaches to Production and Mixing,' 'Technical Requirements,' and 'Music

Business Related.' Subsequently, the FGD phase engaged industry professionals, who provided insights into the core themes of spatial audio production and mixing techniques, binaural playback considerations, and business-related aspects such as mix approval and back catalogue mixes.

The questionnaire results have highlighted that mix engineers are still working within some of the confines of the stereo paradigm. This is understandable as stereo has been the dominant playback system 60 year years, and many music listeners consume their music over headphones. It was shown that mixers tend not to binauralise core music elements, which would typically be presented as phantom stereo image. It also appears that mixers consider some genres of music to be more suitable for Atmos than others and that specific instrumental sources are more suited to placement outside of the typical stereo image. Regarding the two popular binaural renderers, mix engineers do not have a negative perception of their sound quality. They do, however, feel negative towards the lack of control offered by them. Finally, it was shown that mix engineers are mindful of mix translation to playback systems with limited loudspeakers (but not mono playback systems) and especially headphones. Further work intends to expand upon the findings from the survey by carrying out several semi-structured interviews with mix engineers and using participant observation to better understand their working practices in an ecologically appropriate environment, such as their own mixing rooms.

References

[1] "How To Stream Surround Sound Music: Dolby Atmos, TIDAL, Amazon Music," (2021), URL https://www.rollingstone.com/product-recommendations/electronics/how-to-stream-music-in-surround-sound-dolby-atmos-1136008/.

[2] C. Welch, "Apple Music's spatial audio is sometimes amazing but mostly inconsistent - The Verge," (2021), URL https://www.theverge.com/2021/6/9/22525028/apple-music-spatial-audio-dolby-atmos-hands-on.

[3] S. Guttenberg, "Why did SACD, DVD-A, and Blu-ray fail as music surround formats? - CNET," (2011), URL https://www.cnet.com/tech/home-entertainment/why-did-sacd-dvd-a-and-blu-ray-fail-as-music-surround-formats/.

[4] R. Wallis and H. Lee, "Vertical stereophonic localization in the presence of interchannel crosstalk: the analysis of frequency-dependent localization thresholds," *Journal of the Audio Engineering Society*, vol. 64, no. 10, pp. 762–770 (2016 October), doi:10.17743/jaes.2016.0039.

[5] R. Wallis and H. Lee, "The reduction of vertical interchannel crosstalk: the analysis of localisation thresholds for natural sound sources," *Applied Sciences*, vol. 7, no. 3, p. 278 (2017 March), doi:10.3390/app7030278.

[6] J. Berg and F. Rumsey, "Identification of quality attributes of spatial audio by repertory grid technique," *Journal of the Audio Engineering Society*, vol. 54, no. 5, pp. 365–379 (2006).

[7] S. Kim, D. Ko, A. Nagendra, and W. Woszczyk, "Subjective evaluation of multichannel sound with surround-height channels," presented at the *Audio Engineering Society Convention 135* (2013).

[8] H. Lee and C. Gribben, "Effect of vertical microphone layer spacing for a 3D microphone array," *Journal of the Audio Engineering Society*, vol. 62, no. 12, pp. 870–884 (2015 January), doi:10.17743/jaes.2014.0045.

[9] J. Francombe, T. Brookes, R. Mason, and J. Woodcock, "Evaluation of spatial audio reproduction methods (part 2): analysis of listener preference," *Journal of the Audio Engineering Society*, vol. 65, no. 3, pp. 212–225 (2017).

[10] J. Francombe, J. Woodcock, R. J. Hughes, R. Mason, A. Franck, C. Pike, *et al.*, "Qualitative Evaluation of Media Device Orchestration for Immersive Spatial Audio Reproduction," *J. Audio Eng. Soc*, vol. 66, no. 6, pp. 414–429 (2018 June), doi:10.17743/jaes.2018.0027.

[11] J. Kelly, W. Woszczyk, and R. King, "Are you there?: A literature review of presence for immersive music reproduction," presented at the *Audio Engineering Society Convention 149* (2020).

[12] H. Lee, "A conceptual model of immersive experience in extended reality," (2020).

[13] E. Patricio, A. Ruminski, A. Kuklasinski, L. Januszkiewicz, and T. Zernicki, "Toward six

degrees of freedom audio recording and playback using multiple ambisonics sound fields," presented at the *Audio Engineering Society Convention 146* (2019).

[14] R. Lawrence, "Producing Music for Immersive Audio Experiences," in *Producing Music*, pp. 134–155 (Routledge, 2019).

[15] P. Malecki, M. Piotrowska, K. Sochaczewska, and S. Piotrowski, "Electronic music production in ambisonics-case study," *Journal of the Audio Engineering Society*, vol. 68, no. 1/2, pp. 87–94 (2020 January), doi:10.17743/jaes.2019.0048.

[16] J. R. Barboza, E. Magalhães, and G. Bernardes, "Towards Best Practices in Spatial Audio Post Production: A Case Study of Brazilian Popular Music," presented at the *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–7 (2021 November), doi:10.1109/I3DA48870.2021.9610908.

[17] A. De Sotgiu, M. Coccoli, and G. Vercelli, "Comparing the perception of'sense of presence'between a stereo mix and a binaural mix in immersive music," presented at the *Audio Engineering Society Convention 148* (2020).

[18] P. Abehsera Morell and H. Lee, "Binaural Mixing of Popular Music," presented at the *Audio Engineering Society Convention 151* (2021 October).

[19] Chevron, "Dolby Atmos for Music POLL - Will It Take off? - Gearspace.com," (2021 Sep.), URL `https://gearspace.com/board/so-much-gear-so-little-time/1361167-dolby-atmos-music-poll-will-take-off.html`.

[20] P. L. A. Pro, "The Future of Audio with Dolby Atmos with F. Reid Shippen - YouTube," (2021 Oct.), URL `https://www.youtube.com/watch?v=TxFtNjXmv28`.

[21] P. L. A. Pro, "Mixing with Dolby Atmos: Brad Wood (Liz Phair, Smashing Pumpkins, The Bangles) - YouTube," (2021 Sep.), URL `https://www.youtube.com/watch?v=KLBYxtRLwJw`.

[22] Produce Like A Pro, "Mixing In Dolby Atmos with Dave Way," (2021 Mar.), URL `https://www.youtube.com/watch?v=Kky2u5XzZBU`.

[23] A. Moore, *An Investigation into Non-Linear Sonic Signatures with a Focus on Dynamic Range Compression and the 1176 Fet Compressor*, Ph.D. thesis, University of Huddersfield (2017).

[24] J.-P. Herbst and M. Mynett, "What is "Heavy" in Metal? A Netnographic Analysis of Online Forums for Metal Musicians and Producers," *Popular Music and Society*, vol. 45, no. 5, pp. 633–653 (2022 August), doi:10.1080/03007766.2022.2114155.

[25] T. Rice and A. Moore, "Towards a Lexicon for Distortion Pedals," in *Distortion in Music Production*, pp. 128–144 (Focal Press, 2023).

[26] L. Costello, M.-L. McDermott, and R. Wallace, "Netnography: Range of practices, misperceptions, and missed opportunities," *International Journal of Qualitative Methods*, vol. 16, no. 1, p. 1609406917700647 (2017), doi:10.1177/1609406917700647.

[27] W. Goddard and S. Melville, *Research methodology: An introduction* (Juta and Company Ltd, 2004).

[28] M. Ronan, N. Ward, and R. Sazdov, "Investigating the sound quality lexicon of analogue compression using category analysis," presented at the *Audio Engineering Society Convention 138* (2015 May).

[29] S. C. Brown and D. Knox, "Why go to pop concerts? The motivations behind live music attendance," *Musicae Scientiae*, vol. 21, no. 3, pp. 233–249 (2017), doi:10.1177/1029864916650719.

[30] A. Moore, "Towards a definition of compression glue in mastering," in *Mastering in Music*, pp. 44–59 (Focal Press, 2020).

[31] S. Brinkmann, "Unstructured and semi-structured interviewing," *The Oxford handbook of qualitative research*, vol. 2, pp. 277–299 (2014).

[32] L. Zattra and N. Donin, "A questionnaire-based investigation of the skills and roles of Computer Music Designers," *Musicae Scientiae*, vol. 20, no. 3, pp. 436–456 (2016), doi:10.1177/1029864915624.

[33] M. N. Marshall, "Sampling for qualitative research," *Family practice*, vol. 13, no. 6, pp. 522–526 (1996).

Christopher DeweyCDprofile.pngChristopher Dewey is a Senior Lecturer at the University of Huddersfield, UK. Christopher's research interests primarily revolve around the design and evaluation of user interfaces for creativity, with a particular focus on interfaces for popular music production. This focus is motivated by Christopher's passion for music performance/production and professional background in user interface design and development. Christopher completed an MSc by Research in 2014. This project focused on the development of a visually intuitive EQ user interface. Building on this work, in 2023 Christopher completed a PhD investigating the suitability of the audio mixing user interface for popular music production.

Austin MooreAMProfile.pngAustin Moore is a Senior Lecturer and Course Leader of Sound Engineering and Music Production at the University of Huddersfield, UK. After completing his first degree in Popular Music and Recording at the University of Salford, UK, he completed a PhD, which investigated the use of non-linearity in music production, focusing on dynamic range compression and the 1176 compressor. His research interests include music production sonic signatures, Atmos in music production, the user experience and semantic audio. He has a background in the music industry and spent many years producing and remixing various forms of electronic dance music and engineering in recording studios.

Hyunkook LeeHLProfile.png Hyunkook Lee is Professor of Audio and Psychoacoustic Engineering and the Director of the Applied Psychoacoustics Laboratory (APL) and the Centre for Audio and Psychoacoustic Engineering (CAPE) at the University of Huddersfield, UK. His research areas include the perception, recording, rendering and reproduction of 3D multichannel audio, virtual acoustics and binaural audio synthesis, and immersive audio experience. Before joining Huddersfield in 2010, he was a Senior Research Engineer with LG Electronics in South Korea, where he contributed to MPEG audio codec standaridzation activities and developed spatial audio algorithms for mobile devices. He received a a PhD in spatial audio psychoacoustics from the Institute of Sound Recording at the Univeristy of Surrey in 2006. Hyunkook is a Fellow of the AES, a Technical Associate Editor of the AES Journal and the Vice Chair of the AES High-Resolution Audio Technical Committee. He was the co-chair of the 2023 AES International Conference on Spatial and Immersive Audio.

Table 4: Results of pairwise comparisons with Bonferroni adjustments for the questions that considered the frequency with which the respondents $\star$ = used the binauralised settings, † = positioned above the listener, $\diamond$ = positioned behind the listener. Statistically significant differences ($p < .05$) are indicated by the numerals associated with each cell in the Table

	Median values ($\star$,†,$\diamond$)	Kick	Bass	Snare	Toms	Lead Vox	Tuned Perc	Drum OH	Main Guitar	Percussion	Piano	Drum Room	Brass	Add Guitar	Woodwind	Organ	Main Synth	Strings	Backing Vox	Add Synth	Effects
Kick	2,1,1						$\star$†$\diamond$	$\star$†	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$
Bass	2,1,1						†$\diamond$	†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$
Snare	2,1,1						†$\diamond$	†$\diamond$	$\star\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$
Toms	3,2,2						†$\diamond$	†	†	†$\diamond$	$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$
Lead Vox	3,1,1						†$\diamond$	†		†$\diamond$	†$\diamond$	†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$	$\star$†$\diamond$
Tuned Perc	4,3,3							$\diamond$											$\diamond$	†$\diamond$	†$\diamond$
Drum OH	4,3,2								†	$\diamond$	†		$\diamond$	$\diamond$	$\diamond$	$\diamond$	$\diamond$	$\diamond$	$\diamond$	$\diamond$	†$\diamond$
Main Guitar	4,2,2									†$\diamond$		†	$\diamond$	†$\diamond$	†$\diamond$	$\diamond$	†$\diamond$	†$\diamond$	†$\diamond$	†$\diamond$	†$\diamond$
Percussion	4,3,3										†										†$\diamond$
Piano	4,2,3											†		†$\diamond$				†	†$\diamond$	†$\diamond$	†$\diamond$
Drum Room	4,3,3																				†$\diamond$
Brass	4,3,3																			†	†$\diamond$
Add Guits	4,3,3																				†
Woodwind	4,3,3																		†		†$\diamond$
Organ	4,3,3																			†	†$\diamond$
Main Synth	4,3,3																				†$\diamond$
Strings	4,3,3																				†
Backing Vox	4,3,3																				
Add Synth	4,3,3																				
Effects	4,4,4																				

Table 5: Overview of the significantly different genre ratings for Atmos (⊙) and binaural (⋆) formats from post hoc analysis

	EDM	Rock	Pop	Metal	Hiphop	RnB/Soul	Jazz	Folk	Alt Rock	Country	Latin	Classical	World Music
EDM		⊙		⊙	⊙⋆	⊙ ⋆		⊙ ⋆	⊙ ⋆	⊙ ⋆	⊙ ⋆		
Rock			⊙ ⋆	⊙ ⋆		⊙ ⋆	⊙ ⋆				⊙ ⋆	⊙ ⋆	⊙ ⋆
Pop				⊙ ⋆	⊙ ⋆			⊙ ⋆	⊙ ⋆	⊙ ⋆			
Metal					⊙ ⋆	⊙ ⋆	⊙ ⋆	⊙ ⋆	⊙ ⋆	⊙ ⋆	⊙ ⋆	⊙ ⋆	⊙ ⋆
Hiphop												⊙ ⋆	⊙ ⋆
RnB/Soul									⊙ ⋆			⊙ ⋆	
Jazz									⊙ ⋆	⊙ ⋆		⊙ ⋆	
Folk												⊙ ⋆	⊙ ⋆
Alt Rock											⊙ ⋆	⊙ ⋆	⊙ ⋆
Country												⊙ ⋆	⊙ ⋆
Latin													
Classical													
World Music													

Table 6: List of Questions

Question	Response Method	Hypothesis
How frequently do you use the binaural settings (near, mid and far) for the following instruments in a mix?	5 point uniploar scale	1
Move the sliders to rate the suitability of each genre for a Binaural rendered mix	100 point continuous scale	2
Move the sliders to rate the suitability of each genre for an Atmos mix rendered mix	100 point continuous scale	2
An Atmos mix should be clearly derived from and an extension of the original stereo mix	5 point bipolar scale	2
An Atmos mix should be completely new and only use aspect of the stereo mix unless absolutely necessary	5 point bipolar scale	2
Please provide a comment to justify your rating.	Text box	3
How satisfied are you with the spatial quality of the Dolby Binaural Renderer?	5 point bipolar	3
Please provide a comment to justify your rating.	Text box	3
Briefly list the strengths and weaknesses of the Dolby Binaural Renderer.	Text box	3
How satisfied are you with the tonal quality of the Apple Binaural Renderer?	5 point bipolar scale	3
Please provide a comment to justify your rating.	Text box	3
How satisfied are you with the spatial quality of the Apple Binaural Renderer?	5 point bipolar	3
Please provide a comment to justify your rating.	Text box	3
How satisfied are you with the tonal quality of the Dolby Binaural Renderer?	5 point bipolar style	3
How frequently do you create an Atmos mix from pre-processed stems derived from a stereo mix?	5 point unipolar scale	4
How frequently do you create an Atmos mix from a raw multi-track i.e. not from pre-processed stems?	5 point unipolar scale	4
I have to do less processing to make things 'fit' (EQ, compression, limiting, gating etc) when creating an Atmos mix compared with a stereo mix.	5 point bipolar scale	4
Compared to mixing in stereo, I use mix bus processing in a similar way when mixing in Atmos.	5 point bipolar scale	4
Compared to mixing in stereo, I use EQ in a similar way when mixing in Atmos.	5 point bipolar scale	4
How often do you position the following audio sources above the listener?	5 point uniploar scale	4
How often do you position the following audio sources behind the listener?	5 point uniploar scale	4
When creating an Atmos mix, I first create a stereo mix and then spread out audio elements to other locations	5 point unipolar scale	4
What is your current Atmos mixing configuration?	Single Choice	5
How important is it for you to check your Atmos mix over a playback system involving a greater number of speakers than your main mix room?	5 point unipolar scale	5
How important is it for you to check your Atmos mix over a playback system involving a lesser number of speakers than your main mix room?	5 point unipolar scale	5
How important is it that the consumer has a personalized Head Related Transfer Function (HRTF)?	5 point unipolar scale	5
Will head tracking have a negative or positive impact on your mix when listened to by the consumer?	5 point uniploar scale	5
How important it is that your mix translates on the following playback systems?	5 point unipolar scale	5