Audio Engineering Society

Convention Express Paper

Presented at the 156th Convention

2024 June 15-17, Madrid, Spain

Subjective Evaluation of Binaural Renderers in Music Composition and Mixing

Haoping Zhang[1], Amandine Pras[1]

[1] *School of Arts and Creative Technologies, University of York, Heslington, York YO10 5DD, United Kingdom*

Correspondence should be addressed to Haoping Zhang (haoping.rec@gmail.com)

ABSTRACT

This study explores listeners' subjective evaluation of four binaural renderers that were used to mix experimental and popular electronic music, namely Technology 1 Binaural panner (T1B), Technology 2 Atmos panner (T2A), Technology 2 Binaural panner (T2B), and Technology 3 Atmos panner (T3A). We collaborated with seven performers and composers to produce six tracks. Subsequently, 32 participants completed an online survey that compared two binauralized versions of each track. We assessed their immersive experience and preferences, and asked them to describe the differences that they could perceive between the two versions. Findings indicate that participants can perceive differences between two versions of a binaural mix rendered through two different tools. However, significant differences in immersion ratings and preferences remain stimuli dependent.

1 Introduction

1.1 Background

Binaural refers to a two-channel audio format that differs from stereo because signals were recorded through a dummy head and/or filtered by a combination of time, intensity, and spectral parameters intended to mimic localization cues when listening to sound with two ears. [1] The effectiveness of a binaural renderer relies on the precision of the filtering that is applied to non-binauralized sound sources, to emulate the cumulative effects of their reflections on a chest, head, pinna, and inside ear canals. [2-3] By applying head-related transfer function (HRTF) filters that combine the frequency-dependent impact of auricles and other body measurements on the sound, along with interaural differentiation cues, listeners can externalize sound sources and perceive accurate localization cues through headphones monitoring.

Due to individual body measurement differences, HRTFs must be individualized for a binaural renderer to perform perfectly. [4] Nevertheless, previous research has showed that we can compensate for the use of non-individualized HRTFs on headphones by "adding reverberation-related cues and/or dynamic binaural rendering that matches listeners' self-initiated head movements." [5] In this study, we only used static binaural rendering with non-individualized HRTFs. Therefore, we experimented with adding natural and digital reverberation in our mixes according to the theory that "reverberant sounds are more likely than anechoic sounds to be perceived as externalized." [6]

Although the first binaural invention named 'Teatrophone' by Clément Ader goes back to the late 19th century, and HRTFs have been measured from the late 1970s, [7] the general audience's interest in the binaural format only began in increase in the 2010s, alongside the growth of augmented and virtual

reality applications, including video games. [8] Since then, the amount of research assessing the effectiveness of binaural renderers in creating immersive experiences for music listening and/or sound for digital media has rapidly expanded. [9-13] Nevertheless, binaural for music production has yet to find its commercial path.

Realizing that "75% of [their] listeners stream their music through headphones", [14] Dolby incorporated binaural CODECs in the Atmos renderer, a surround sound technology that is primarily dedicated to representing object-based spatial audio mixes on speakers. Because Atmos panners are now available in most popular digital audio workstations, mixing engineers can easily experiment with the format. The future will tell us whether Dolby's approach will convince the general audience to listen to music in binaural. Our purpose for this study is to examine whether listeners' experience of immersion differ when a binaural mix is rendered through different tools that are commonly used in studio production and live engineering, including Atmos panners.

1.2 Spatial audio in composition

The experience of listening to spatial audio tends to be associated with a sense of realism, naturalness, presence, immersion, and being surrounded. [11, 15-18] These characteristics can contribute significant elements to the toolkit of composers to enhance their artistic vision. [19] More than half of the composers who completed Peters et al.'s survey reported using spatialization effects for live electronics and/or prepared electronics also known as fixed media. [20] Moreover, multiple responses suggested that the listener's experience can be augmented through immersing them in sound. We could thus expect composers to frequently use spatialization techniques and experiment with available tools. However, the integration of spatial concepts into compositional processes remains limited.

Peters et al. noted barriers for composers to experiment with source spatialization, such as the complexity of setting up multi-loudspeaker systems that work well for all audience seats in a live performance venue; a steep learning curve for new technologies; usability issues; and a lack of reliability across non-standardized playback environments. [19-21] For instance, according to composer Natasha Barrett, "as soon as you leave that private listening space, your composed space may collapse because you are no longer in control of your environment." [19] Whereas Atmos offers a standardized solution to overcome some of these barriers in studio production, its use requires sound engineering knowledge, both in terms of "attempting realism" for classical and acoustic productions [12, 22-24] and "creating virtual worlds" for popular and electronic productions [10, 25-27]. This is why we proposed that professional composers and performers collaborate with us, and encouraged student composers to create their music with spatialization audio tools.

1.3 Mixing music in binaural

According to Snow, binaural audio "transports the listener to the original scene." [28] Similarly, Fontana et al. stated that binaural could facilitate the perception of a space compared to stereo [29], therefore adding depth and more realistic spatial features to compositions. Also, binaural headphone monitoring can stimulate musicians' creativity [30] and enhance their immersion in performance [5] by emphasizing the "illusion of reality" and/or the "reality of illusion". [31]

As for any reproduction format, recording and mixing techniques for binaural need to be adapted to the type of sound sources and musical genre expectations. For example, a binaural production of acoustic recordings requires the setup of several microphone systems to capture the auditory scene in a particular venue, [22-23] and a binaural production of popular music necessitates specific mixing techniques, and special attention to which sources to externalize or not, [26-27] e.g. a hip hop beat may be more effective if its main components are not externalized. [33] We drew upon this knowledge to explore binaural production approaches with experimental and contemporary composers and performers for this study.

1.4 Research questions

RQ1: How do different binaural renderers affect listeners' preferences and experience of immersion when the mixing parameters for each version are set to be identical?

RQ2: Do participants' gender, personal monitoring preferences between headphones and speakers, and number of years of audio education and/or experience impact their immersion ratings?

RQ3: To what extent do binaural renders affect artistic aspects of the mix?

2 Production of six stimuli

Table 1 features the musical and binaural technology information of the six stimuli that were produced by

the authors and MA students involved in this study.[1] For four of the tracks, the sound sources were spatialized and balanced to best convey the intentions of the composers and performers with whom we collaborated. Also, sonic textures were designed according to musical genre expectations and artist(s)' recording references. Two MA students mixed their own electronic compositions with the help of author HZ. We detail the production process of each track in the next subsections.

After a first binaural mix of a track was completed and validated by the artists (if applicable), the mixers replicated the audio spatialization cues using a different binaural render to create a version of the mix to be compared with the first one. To achieve this, when mapping the spatial cue parameters of the first mix in the 3D audio panner of the tool used for the replication, the mixers ensured that each individual object or instrument of the mix would be perceived coming from the same exact location. In the cases where adaptations to technology differences and limitations were needed, e.g. when passing from a spherical to a Cartesian coordinate system or vice versa, they used their technical ear to match the perceived localization of the sound sources across the two versions. Any other signal processing such as equalization and dynamic range compression that was used to create the first mix remained the same in the replicated one, making the binaural rendering process the only factor that could impact the perception of immersion and sonic textures.

Table 1. Information about the six stimuli in the order they were presented in the survey

Stimulus # Track Title	**Artist(s)**	**Duration**	**Genre**	**Instrumentation & DAW**	**First mix**	**Replicated mix**
S1. *Fusing É-Toilium*	Ulrica Dúo	10'24''	Electroacoustic	Recordings and electronics (Pure Data)	T1B	T2B
S2. *Improvisation*	Marjolaine Charbin	14'50''	Free improvisation	Prepared piano and voice	T2B	T1B
S3. *Suibian* [*Casual*]	Yifei Wu	2'57''	Electronic dance music	MIDI instruments (Ableton Live)	T3A	T1B
S4. *Places*	Eva Blanche	3'59''	Trip hop	Taishogotos , piano, voice, cello, synth bass, electronic percussion and keys	T3A	T2B
S5. Excerpt of *ReCantata*	Lore Lixenberg, Mattias Petersson, George Kentros	5'33''	Experimental electronic baroque	Voice, violin and electronics (Max/MSP)	T2A	T3A
S6. *Jijiji*	Yiqi Cai	3'40''	Electronic composition	MIDI instruments (Logic Pro X)	T2A	T3A

2.1 *Fusing É-Toilium* (S1)

The first stimulus (S1) was composed by the Ulrica Dúo, which features Colombian composers Juan Hernández Vega (JHV), PhD student at the University of Leeds, and Ángela Hoyos Gómez (AHG), PhD student at the University of Huddersfield. S1 was edited by Linyuan Wang and mixed by Tianyi Liang under the supervision of HZ. It superposes two previous composition projects that were presented in quadraphonic installations at the Market Gallery in April 2023, namely *Helium Burning*, for which JHV worked with Portuguese London-based artist Inês Rebelo, to stage the exploration of his microsound transformation techniques when sonifying nucleosynthesis data with sound synthesis; and *Es-Tela Est-Toile Confab*, for which AHG collaborated with Amy Chen, Lecturer in the Department of Fashion and Textiles, to highlight the presence of women in astronomy using immersive textiles, and the vocal processing of talks on stellar evolution by Julieta Fierro, Catherine Walsh, and Teresa Paneque.[2]

The first edit and mixing steps consisted of positioning the Gallery's quadraphonic speaker systems in a 2D 360˚ plan for each of the two

[1] These tracks were engineered as part of the Music as Audio, Engaging with Research, and Production Portfolio modules of the Master of Art in Music Production and Audio Cultures led and taught by author AP.

[2] A. H. Gómez, A. Chen, J. Hernández, and I. Rebelo, "Starts | Cultures of Sound," 8th May 2024, https://starts2023.github.io/

compositions, and placing *Es-Tela Est-Toile Confab* above *Helium Burning* in the 3D audio space. The engineers refined the first draft based on the composers' feedback to create a final version that conveyed their vision of "a sonic mix for the merging interlacing converging, diverging, diffusive, and wrapping meetings of the works."[3]

2.2 *Improvisation* (S2)

S2 was performed on a Fazioli F278 grand piano by Marjolaine Charbin (MC), a French London-based experimental musician and free improviser. It was recorded in a workshop taught by the authors to MA students in the Rymer Auditorium. A range of microphone systems was set up to enable options for a spatial audio mix. These included a Blumlein inside the piano; four small diaphragm condensers to emphasize specific piano preparation techniques; and an AB pair with omnidirectional microphones, an ORTF pair, and the Neumann KU 100 dummy head placed in front of the piano to capture the sound image from the audience's perspective. In addition, we used two omnidirectional microphones to capture the surround sound at the back of the auditorium, and two omnidirectional microphones to capture height.

For this improvisation, MC layered rhythmic patterns on the piano keyboard, drone-like textures created with small objects vibrating on the strings, and fragments of melodies and voice "to immerse herself and the audience in a trance-like imaginary space from which an abstract story can unfold." Interestingly, she used objects that can sustain the resonance in the piano and leave her hands free for playing other parts. She also played with a contact microphone whose signal was routed to a speaker close to the piano to add textures.

The mixing process was carried out in an online meeting on Zoom, with MC and AP guiding Shuli Mo through her exploration of binaural panning options with the different microphone systems. The end result involved a minimum of signal processing, and mostly relied on muting some systems and balancing others, to retain the rawness and rich textures of the performance.

2.3 *Suibian* [*Casual*] (S3)

S3 was composed and mixed by Yifei Wu, who was inspired by approaches to "decolonise music" introduced by Dr Philip Burnett in a lecture. [32] Her approach consisted of eliminating Western orchestration and focussing on arranging drums and percussion instruments. She also used a phaser to modulate impulsive and transient sound sources to obtain a flowing sensation in the 3D audio space. In the Atmos panner, all objects were set to 'close' for better clarity.

2.4 *Places* (S4)

S4 was composed and primarily performed by Eva Blanche (EB), a French-American singer-songwriter who came to York for a week in January 2023, to collaborate with the authors and MA students. She recorded overdubs of backing vocals, taishogotos and grand piano on the MIDI arrangements, and lead vocal recordings that she came with. During her residency, we introduced her to Gaia Blandina who improvised cello parts on the track. Back home, EB asked Simon Hedges to record an electric bass line that would fit with the new vibe of the track.

HZ edited and mixed the track according to EB's requests and feedback, with special attention to the integration of the taishogotos within the rest of the instrumentation. She commented that the binaural mix accentuated a spectrum gap between the drums and bassy instruments on the one hand, and the high, floating voices on the other hand, as if there were a silent instrument that was covered by the synths upon their entry, halfway through the track. In other words, not having sources in the medium frequency range made the overall piece sound enigmatic: It could be "the missing place" or "the next place" that the voice is calling for. In the mixing process in Atmos, some objects were placed 'close' and some 'far' to enhance the sensation of depth in the mix.

2.5 *ReCantata* (S5)

S5 consists of an experimental and electronic interpretation of the Aria "*O Ewigkeit, du Donnervort*" from Jean-Sebastian Bach's 60th Cantata by London-based singer Lore Lixenberg, Swedish violinist George Kentros, and Swedish composer Mattias Petersson who performed on live electronics. S5 was recorded by the authors and MA students in the Sir Jack Lyons Concert Hall during a workshop on 30th January, 2023. The voice and violin were close miked to allow for clean real-time processing in MaxMSP and for creative spatialization in the 3D audio space.

[3] Excerpt of the track description written by the composers for the artists' intention questions in the Qualtrics survey.

The electronics were recorded through DI boxes and played live through two monitors that faced the vocalist and the violinist in the Hall, with AB with omnidirectional microphones in between. We also positioned two omnidirectional microphones beyond the performance space, to have options for the mix.

S5 was edited on Pyramix and mixed in Atmos by Yufan Pan (YP), following the trio's aim "to go beyond reproducing the sound that was imagined existing at the time of the composer's writing." Their purpose was "to perform re-imaginings of ancient music [while] dialoguing with the composer as if they were still alive and opening up the score for further reinterpretation whilst still keeping the integrity of the music." They commented that "Yufan's binaural imagining of the performance fit very well with the ethos of the trio who also felt that in view of when Cantata 60 was written, the subject and reason behind fear of death was perfect for a post-pandemic world." This comment refers to YP's use of the violin spot mic as an object moving in circle in the height plan.

2.6 *Jijiji* (S6)

S6 was composed and mixed by Yiqi Cai (YC). Also inspired by Burnett's lecture on "decolonising music", he used MIDI samples and the 3D audio space to stage African aborigines living a peaceful life before being colonized, then being distressed by the invasion, and finally struggling to regain freedom.

3 Listening Survey Methods

3.1 Survey design

An online survey with a mixed-method approach was designed by HZ and Zhao Deng (ZD) on Qualtrics to collect comprehensive data regarding listeners' subjective evaluation of the four different binaural renderers used to mix the six stimuli, namely Technology 1 Binaural panner (T1B), Technology 2 Atmos panner (T2A), Technology 2 Binaural panner (T2B), and Technology 3 Atmos panner (T3A). The survey included three sections: an informed consent; demographic and monitoring preference questions; and the comparative listening test using the six stimuli that each featured the binaural mix of a track rendered through two different tools.

After listening to each stimulus that alternated the two binaural versions of the mix every 30 seconds, with corresponding visual cues as A or B, participants were asked whether they could perceive audible differences, and to rate how immersive they felt when listening to each version, from 'not immersive at all' to 'strongly immersive', on a 5-point Likert scale. The third question aimed to collect their preference between the two versions by selecting 'A', 'B', or 'not sure', followed by a prompt to describe all the differences that they could perceive. Finally, they were invited to indicate which version they thought best conveyed the artist(s)'s intentions, based on a description written by the composer(s) or the main performer of the track.

3.2 Participants

We recruited listeners from our respective networks via email and social media to participate in the online survey on a voluntary basis. Although we received 80 responses, only 32 were valid after manual screening. We considered a response valid if the participant had completed all the questions and had entered different responses across stimuli. For instance, if all their ratings of immersion were the same across stimuli, we did not consider the response valid.

These 32 respondents included 15 females, 14 males, and three who preferred not to disclose their gender. They were based in seven countries and territories, namely the UK (20), Mainland China (5), Canada (2), the Netherlands (2), Australia (1), France (1), and Hong Kong (1). More than a third (14 out of 32) were Master's students at the University of York. Sixteen of them were between 18- and 25-years-old; 11 between 26 and 30; and 6 older than 30. Also, six respondents disclosed having hearing damage and tinnitus, with four in the 18–25 and two in the 26–30 age group.

Among the 32 participants, 11 reported that they did not have any training or experience in audio, nine reported less than three years of audio education or experience, four reported three to five years, and seven over five years. One participant did not specify. Sixteen participants preferred monitoring music on speakers and 12 on headphones. Four did not state a monitoring preference.

3.3 Analysis approach

Both quantitative and qualitative analyses were conducted to thoroughly examine the research questions. The Likert-scale measurements of immersion ratings are treated as continuous, with 'not immersive at all' scored as 1 and 'strongly immersive' as 5. In the next section, we present descriptive statistical analyses through tables and a figure; and we report the t-test results about the significance of differences in listeners' immersion ratings between

two versions of each stimulus; the results of the chi-square tests that were conducted to assess the impact of participants' gender and monitoring preferences on immersion ratings; and the result of an ANOVA that was computed to assess whether the number of years of audio education and/or experience had an impact on immersion ratings. Also, we list the sound criteria that were coded from the listeners' descriptions of the two versions. We also summarize the predominant comments that we identified for each renderer.

4 Listening Survey Findings

4.1 Perception of differences between versions

Table 2 shows that all 32 listeners could perceive differences between the TB1 and TB2 versions of S1; 91% of them could perceive differences between the T3A and T2B versions of S4; 84% between the T2B and T1B versions of S2; and 78% between the T3A and T1B versions of S3, and between the T2A and T3A versions of S5 and S6.

Table 2. Percentage of listeners who can perceive differences between the two versions of the mix

	Yes	No
S1	100%	0
S2	84%	16%
S3	78%	22%
S4	91%	9%
S5	78%	22%
S6	78%	22%

4.2 Immersion ratings

Figure 1 presents the mean and standard deviation of immersion ratings for both binaural versions of the six stimuli, and Table 3 indicates the probability value of one-tailed t-test results. We observe significant differences in listeners' immersion ratings between the two versions of S6 and S3, with the T2A version of S6 being rated as more immersive than the T3A version, and the T1B version of S3 being rated as more immersive than the T3A version. The other comparisons did not lead to significant results.

Statistical test results show that:

- Gender had an impact on participants' immersion ratings, χ^2 (4, N=29) = 13.41, p = 0.009 (<0.05), with female participants rating immersion significantly higher (M = 3.6) than male participants (M = 3.2).
- The relationship between participants' monitoring preference for speakers vs. headphones and their immersion ratings was not significant, χ^2 (4, N=28) = 8.11, p = 0.088 (>0.05).
- The relationship between the participants' number of years of audio education and/or experience, and their immersion ratings was not significant, F (3, N=32) = 1.19, p = 0.313 (>0.05).

Table 3. p-value of one-tailed t-test results, with grey cells indicating significant results

	***p*-value**
S1	0.166
S2	0.142
S3	0.004
S4	0.065
S5	0.080
S6	0.00004

4.3 Preferences between versions

Table 4 shows that 69% of the listeners preferred the T2A version of S6 (first mix), and 16% the T3A version. In contrast, 53% preferred the T3A version of S4 (first mix), and 28% the T2B version. Also, 50% preferred the T2B version of S1 (replicated mix), and 31% the T1B version. Considering all stimuli, the first mix was selected 45% of the times, and the replicated one 34%.

Table 4. Percentages of listeners' preference for a specific binaural version, with *indicating which version corresponds to the first mix

	T1B	T2A	T2B	T3A	Not sure
S1	31%*		50%		19%
S2	37%		44%*		19%
S3	44%			31%*	25%
S4			28%	53%*	19%
S5		44%*		28%	28%
S6		69%*		16%	16%

4.4 Descriptions of perceived differences

We extracted 421 phrasings from listeners' free-format descriptions of the differences they perceived between the two binaural versions of each stimulus. These phrasings were coded into 31 sound criteria that were grouped into six broad categories. Here we report on the criteria that included at least three phrasings:

- Sound quality criteria that are commonly used to describe a mix regardless of format (232 occurrences), namely *frequency response* (42),

panning choices (42), *dynamic range* (25), *clarity* (25), *proximity* (23), *definition* (18), *depth* (14), *balance and cohesion* (12), *loudness* (12), *richness* (12), *brightness* (4), and *tonality* (3).

- Spatial audio attributes (77) listed in Agrawal et al.'s definition of immersion [16] and Berg's definition of envelopment [15], namely *sense of space* (33), *immersion* (13), *realism* (9), *envelopment* (7), *presence* (6), *naturalness* (5), and *surround sound image* (4).
- Other spatial audio attributes (44), namely *plane level information* (17), *perception of movement* (11), *sense of a sound source* (10), and *height information* (6).
- Other descriptions of perceptual differences and preferences (27), namely *agreeableness* (6), *punchiness and bouncing* (4), *a certain instrument that is preferred or disliked* (4), *engagement and excitement* (4), *noisy* (3), *phasy* (3), *preference statement without a reason* (3).
- Comments about *reverberation effects* (9).

It is worth noting that, although we standardized the loudness of all stimuli, *loudness* was mentioned 12 times in the responses, indicating that the perception of dynamic range can be impacted by the renderer. For instance, a listener who has more than five years of audio engineering education and/or experience stated, "*The biggest difference I notice here, is that T2A is louder* [than T3A], *which is making it hard for me to trust the other things I'm hearing, as I will always prefer the louder of the two, and know I can convince myself I'm hearing other things*."

Our grouping of phrasings per binaural renderer shows that T2A was qualified as very immersive. Sources externalized through T2A were perceived as somewhat closer than when externalized through T3A, with a more pronounced frequency response in the medium and high frequencies. Also, clarity was mentioned more often for T2A than for T3A. The frequency response through T3A rendering was reported as the most balanced, with positive comments on medium and low frequencies. T3A was perceived as clear in general.

Sources externalized through T1B were often perceived as distant, with a noticeable bottom end and top end. T1B was described as creating a good sense of space, depth, and immersion. Some participants mentioned that its rendering enhanced clarity but other reported muddiness. T2B was generally preferred for its representation of low frequencies. Sources externalized through T2B were in general perceived closer than those externalized through other renderers. TB2 was also generally preferred for its amount of low frequencies. As for T1B, whereas some participants mentioned that T2B enhanced clarity, others reported muddiness.

4.5 Meeting the artists' intentions

Table 5 shows that 56% of the participants selected T3A (first mix) as the binaural version of S4 that best conveyed EB's intentions, and 31% selected T2B. Also, 53% selected T2A (first mix) as the binaural versions that best conveyed YC's intentions, and 13% selected T3A. Participants provided reasons for their choice only 24 times out of the 192 selections. Most of the answers indicated that the perceived differences between the two versions did not alter the narrative or conceptual aspects of the composition. Some also mentioned that they could not understand the artist(s)'s description of their intention. Considering all stimuli, the first mix was selected 40% of the times, and the replicated one 31%.

Table 5. Percentage of listeners who selected a specific binaural version that best conveys the artist(s)'s intention, with *indicating which version corresponds to the first mix

	T1B	**T2A**	**T2B**	**T3A**	**Not sure**
S1	34%*		47%		19%
S2	31%		31%*		38%
S3	37.5%			25%*	37.5%
S4			31%	56%*	13%
S5		41%*		25%	34%
S6		53%*		13%	34%

5 Discussion

5.1 Findings of the comparative study

Overall, our findings suggest that listeners could perceive audible differences introduced by different binaural renders, sometimes experienced immersion differently between two versions, and could state a clear preference for one version over another. They could also describe, with details, the differences that they perceived. However, these differences depended on the stimuli, which means that we cannot conclude that one binaural render was significantly more effective in enhancing listeners' immersion than another. We only noticed that T2A had the highest immersion ratings of all four renderers.

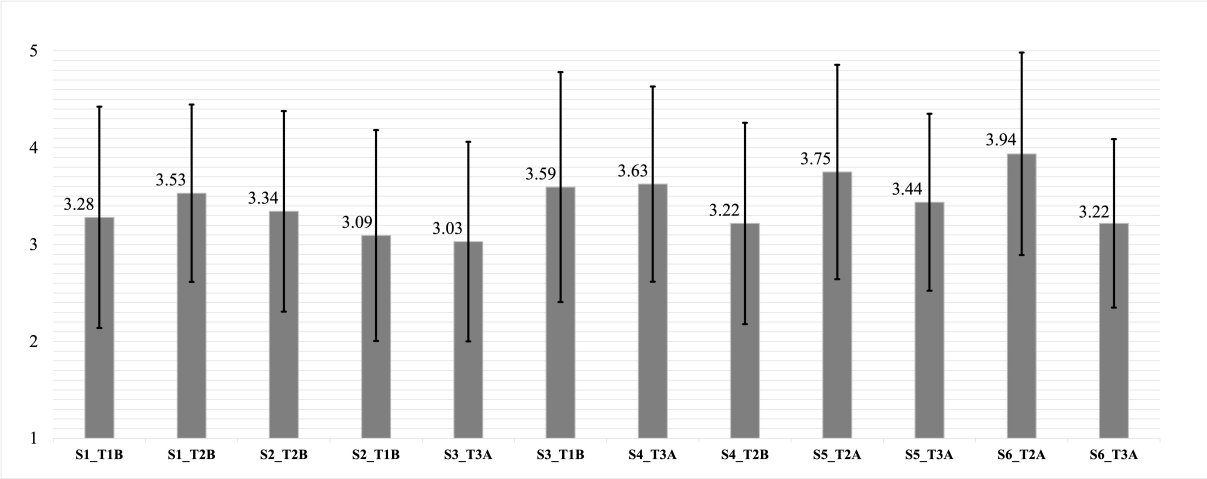

Figure 1. Immersive ratings' mean and standard deviation for the two versions of the six stimuli

Interestingly, out of the 421 phrasings extracted from the listeners' descriptions of the differences they perceived between versions, more than half (232) were coded into sound quality criteria that are commonly used to describe a mix regardless of format, with *frequency response* and *panning choices* being the most frequently mentioned codes (42 times each). We noticed that the reproduction of low frequencies (15), high frequencies (10), and the overall frequency response (10) played an important role in listeners' preferences, with the perception of low frequencies being associated with loudness. Also, comments about panning choices referred to wide or narrow (30), and spacious or crowded (12). These observations suggest that binaural renders impact the frequency response and perception of panning choices more than spatial audio attributes. Among spatial audio attributes, the sense of space was the most reported criterion. Immersion was only mentioned 13 times, which indicates that immersion is not much impacted by different binaural renderers.

Whereas the first mix was selected more often than the replicated one for listeners' preferences and the version that best conveyed the artist(s)'s intentions, we cannot conclude that it had a strong impact on listeners' choices. This means that the artistic aspects of a binaural mix are likely to remain when being rendered through different tools or CODECs.

We found a significant relationship between the gender of the participants and their immersion ratings, with female participants rating higher than male participants in general. Further research needs to be conducted to explain the findings. We did not observe any significant relationship between participants' monitoring preferences for speakers or headphones and their immersive ratings on the one hand, and between their number of years of audio education and/or experience and their immersive ratings on the other. This implies that listening to music in binaural on headphones does not need any training and can be enjoyed by a large range of listeners.

5.2 Contributions and limitations

Our research process confirms that composers and performers are interested in spatial audio and likely to creatively engage with binaural tools if they can collaborate with or be guided by sound engineers who have expertise in recording and mixing techniques. Our description of each stimulus' production process, with details about the artistic context and expectations, contributes practical knowledge in recording and mixing for spatial audio formats. Also, the list of codes that we identified in the qualitative analysis extends the research about the perception of spatial audio for music production and beyond.

The main limitation of this study is that two versions of a mix that was compared for each stimulus could have been rendered through four different tools This limited our ability to aggregate responses across stimuli to draw clear conclusions about the impact of a specific renderer on immersion and other sound criteria. Also, our sample included a limited number of participants to conduct statistical analysis. This is because the valid response rate of the survey was only 39.5%, with an average completion time of 48 minutes, which is long for an unpaid study.

6 Acknowledgement

We would like to warmly thank Ángela Hoyos Gómez, Juan Hernández, Marjolaine Charbin, Eva Blanche, Lore Lixenberg, George Kentros, and Mattias Petersson for collaborating with us, sharing their artistic vision, and writing a description of their intention for the survey. We also thank Yifei Wu and Yiqi Cai for their composition work; Zhao Deng, Tianyi Liang, Shuli Mo, Yufan Pan, and Linyuan Wang for their contributions to the production of stimuli and research process; Stef Conner for her review; and all the participants for their time to complete the survey. The listening test procedure was approved by the Research Ethics Committee of the School of Arts and Creative Technologies of the University of York (UK) in May 2023.

References

[1] A. Roginska, "Binaural Audio Through Headphones," in: "*Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*", A. Roginska, and P. Geluso, editors. New York: Routledge, pp. 88 (2018).

[2] E. M. Wenzel, D. R. Begault, and M. Godfroy-Cooper, "Perception of Spatial Sound," in "*Immersive Sound: the Art and Science of Binaural and Multi-Channel Audio*", Edited by Agnieszka Roginska and Paul Geluso. New York: Routledge (2018).

[3] W. A. Yost, and R. Dye, "Fundamentals of Directional Hearing," in *Seminars in Hearing*, 18 (4), pp. 321-344, New York: Thieme Medical Publishers (1997).

[4] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?" in *J. Audio Eng. Soc.*, Vol. 44, No. 6 (1996).

[5] V. Bauer, D. Soudoplatoff, L. Menon, and A. Pras, "Binaural Headphone Monitoring to Enhance Musicians' Immersion in Performance," in: "*Advances in Fundamental and Applied Research on Spatial Audio*," B. F. G. Katz, P. Majdak, editors. IntechOpen (2022). doi:10.5772/intechopen.91556.

[6] V. Best, R. Baumgartner, M. Lavandier, P. Majdak, and N. Kopčo, "Sound Externalization: A Review of Recent Research," in *Trends in Hearing*, Volume 24: 1-14 (2020).

[7] B. Boren, "History of 3D Sound*," in: "Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*", A. Roginska, and P. Geluso, editors. New York: Routledge (2018).

[8] K. Sunder, "Binaural audio engineering," in: "*3D Audio*", J. Paterson, and H. Lee, editors. Abingdon, Oxon; New York, NY: Routledge (2022).

[9] G. Reardon, A. Genovese, G. Zalles, P. Flanagan, and A. Roginska, "Evaluation of Binaural Renderers: Multidimensional Sound Quality Assessment," in AES Conference on Audio for Virtual and Augmented Reality, Redmond, USA (2018).

[10] B. Martin, R. King, B. Leonard, D. Benson, and W. Howie, "Immersive Content in Three Dimensional Recording Techniques for Single Instruments in Popular Music," in AES 138th Convention, Warsaw, Poland (2015).

[11] J. Kelly, R. King, and W. Woszczyk, "A Novel Spatial Impulse Response Capture Technique for Realistic Artificial Reverberation in the 22.2 Multichannel Audio Format," AES 147th Convention, New York, USA (2019).

[12] W. Howie, D. Martin, T. Kamekawa, J. Kelly, and R. King, "Comparing immersive sound capture techniques optimized for acoustic music recording through binaural reproduction," AES 150th Convention, online (2021).

[13] L. Brümmer, "Composition and Perception in Spatial Audio," in *Computer Music Journal*, 41:1, pp. 46-60 (2017).

[14] E. Pfanzagl-Cardone, "The Dolby Atmos System," in "*The Art and Science of 3D Audio Recording*," Springer International Publishing AG (2023).

[15] J. Berg, "The Contrasting and Conflicting Definitions of Envelopment," in AES 126th Convention, Munich, German (2009).

[16] S. Agrawal, A. Simon, S. Bech, K. Bæntsen, and S. Forchhammer, "Defining Immersion: Literature Review and Implications for Research on Audiovisual Experiences," *J. Audio Eng. Soc.*, vol. 68, no. 6, pp. 404-417 (2020).

[17] J. Kelly, W. Woszczyk, and R. King, "Are you there?: A Literature Review of Presence for Immersive Music Reproduction," in AES 149th Convention, online (2020).

[18] G. Reardon, J. S. Calle, A. Genovese, G. Zalles, M. Olko, C. Jerez, P. Flanagan, and A. Roginska, "Evaluation of Binaural Renderers: A Methodology," in AES 143rd Convention, New York, USA (2017).

[19] F. Otondo, "Creating Sonic Spaces: An Interview with Natasha Barrett," *Computer Music Journal*, 31:2, pp. 10-19 (2007).

[20] N. Peters, G. Marentakis, and S. McAdams, "Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative Analysis," in *Computer Music Journal*, 35:1 (2011).

[21] N. Peters, "Sweet [re]production: Developing sound spatialization tools for musical applications with emphasis on sweet spot and off-center perception," PhD Thesis, McGill University, Canada (2010).

[22] C. Eaton, and H. Lee, "Subjective evaluation of three-dimensional, surround and stereo loudspeaker reproductions using classical music recordings," in *Acoust. Sci. & Tech.* 43, 2 (2022).

[23] W. Howie, R. King, and D. Martin, "A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound," in the 141st AES Convention, Los Angeles, USA (2016).

[24] W. Howie, T. Kamekawa, and M. Morinaga, "Case Studies in Music Production for 3D Audio Reproduction with Bottom Channels," in the International Conference on Spatial & Immersive Audio, Huddersfield, UK (2023).

[25] C. Guastavino, B. F. G. Katz, J. Polack, D. J. Levitin, and D. Dubois, "Ecological Validity of Soundscape Reproduction," in *Acta Acustica united with Acustica* (2004).

[26] W. Howie, "Pop and Rock music audio production for 22.2 Multichannel Sound: A Case Study," in the 146th Convention of the Audio Engineering Society (2019).

[27] B. Martin and R. King, "Three-Dimensional Spatial Techniques in 22.2 Multi-channel Surround Sound for Popular Music Mixing," in the 139th Convention of the Audio Engineering Society (2015).

[28] W. Snow, "Basic Principles of Stereophonic Sound," in *Journal of the SMPTE* Vol. 61, pp. 568 (1953).

[29] S. Fontana, A. Farina, and Y. Grenier. "Binaural for popular music: a case of study," in the Proceedings of the 13th International Conference on Auditory Display, Montréal, Canada, 2007.

[30] V. Bauer, H. Déjardin, and A. Pras, "Musicians' binaural headphone monitoring for studio recording," in the AES 144th Convention, Milan, Italy (2018).

[31] V. Moorefield, *The Producer as Composer: Shaping the Sounds of Popular*, MIT Press, Cambridge, USA (2005).

[32] P. Burnett, E. Johnson-Williams, and Y. Liao, "Music, empire, colonialism: sounding the archives," in *Postcolonial Studies*, Vol. 26, No. 3, 345-359 (2023).

[33] K. Turner, and A. Pras, "Is Binaural Spatialization the Future of Hip-Hop?", in the 147th AES Convention, New York, USA (2019).

[34] A. Pras, C. Guastavino, and M. Lavoie, "The Impact of Technological Advances on Recording Studio Practices," *J. of the American Society for Information Science and Technology*, 64 (3): 612-626 (2013).