Audio Engineering Society

Convention Paper 10275

Presented at the AES 160th Convention
2026 May 28–30, Copenhagen, Denmark

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

NAVIQUAL: Creating Spatial Audio Quality Maps for Virtual Live Music Environments

Carl Timothy Tolentino[1] and Andrew Hines[1]

[1] *University College Dublin*

Correspondence should be addressed to Carl Timothy Tolentino (carl.tolentino@ucd.ie)

ABSTRACT

Live music environments can be simulated and evaluated through spatial audio and augmented reality (AR) technology. However, conducting perceptual studies on AR environments can be challenging, as multiple design considerations and uncontrolled variables come into play. Hence, we developed Naviqual, a tool to create a spatial audio quality map for a virtual live music environment. We generated objective quality contour and polar maps to predict the quality of experience (QoE) across listener locations and directions respectively. We found that these maps strongly aligned with perceptual evaluations by normal-hearing listeners through listening tests. We also found that localisation quality and signal-to-noise ratio both strongly predict QoE across listener translations, with the former outperforming the latter in predicting QoE across listener directions. Overall, Naviqual provides a QoE map for virtual live music environments robust across listener locations and directions, noise locations, music content, and room reverberation.

1 Introduction

Live music offers a unique experience compared to recorded music, enhancing the listeners' quality of experience (QoE) and expanding their social opportunities with peers and artists. However, live music can be impacted by external distractors such as audience noise and room reverberation. Audience participatory noise impacts the overall loudness levels by up to 10 dB, potentially violating regulatory loudness limits and thereby impacting the hearing health of the audience [1]. Consequently, hard-of-hearing (HoH) individuals become excluded from attending live music, as the increased loudness levels overload hearing aid and cochlear implant devices [2]. Unfortunately, sound engineers have no control over these external factors during performances.

To ensure a safe and enjoyable experience for the listeners, acoustic treatment of performance venues is conducted through multi-point measurements of loudness, clarity, and reverberation [3]. During performances, loudness levels are typically monitored using a noise dosimeter placed at the front-of-house (FOH) mixing desk [4]. However, measurements at this position usually underestimate the intensity especially at the most-exposed locations [4], signifying the need for measuring intensity distribution across the venue. Aside from safety, sound quality distribution is also important, which can be addressed by considering optimal loudspeaker configurations [5, 6]. However, it requires repeated calibrations and measurements across different scenarios, performances, and venues.

With the advancement of spatial audio and augmented reality (AR), simulating live music environments has

become plausible. Transitioning from stereo to spatial audio music has been shown to elevate the listeners' overall QoE [7]. Wearable AR glasses that render real-time six degrees of freedom (6DoF) spatial audio with room acoustics have been developed [8]. These egocentric devices can potentially shape the audio material based on a listener-centric approach, e.g. through source separation, beam-forming, 6DoF position and eye tracking, among others [9].

Despite these opportunities, perceptual studies on spatial audio with AR devices remain underexplored. A potential reason could be the challenging implications of conducting perceptual studies with AR, since the listener can freely navigate the AR environment in an uncontrolled fashion. In addition, multiple AR design considerations come into play, e.g. number of sound sources and their placements, content, room size and acoustics, among others. Without first understanding the spatial distribution of quality for a given virtual live music environment, perceptual studies with AR devices can be cumbersome and counterproductive.

Hence, we propose Naviqual[1], a publicly available tool to create spatial audio quality maps for virtual live music environments validated by human perceptual evaluations. Through these maps, we can identify optimal QoE locations and directions within the virtual environment given a specific sound scene. In this work, we aim to answer the following research questions:

RQ1: How can we create spatial audio quality maps for virtual live music environments that are aligned with perceptual evaluations?

RQ2: How robust are the spatial audio quality maps across sound scene variations, i.e. listener locations and directions, noise locations, music content, and room reverberation?

RQ3: To what extent do the objective quality maps align with perceptual evaluations?

2 Related Work

Previous work has explored perceptual studies involving spatial audio and AR. For instance, HoloBand is an AR music simulation software for spatial audio training with HoH listeners [10]. Although subjective evaluations with speech pathologists have been conducted, end-user perceptual evaluations have not yet been conducted for the software [10]. In another study, incorporating higher-order ambisonics (HOA) for spatial audio has been shown to increase listener engagement on 360° videos through a virtual reality (VR) device after analysing listener head pose, pupil dilation, and heart rate [7].

There have also been efforts on evaluating the spatial quality distribution in music venues. The EgoMusic data set contains live music performances captured egocentrically through wearable glasses with audience and musician perspectives [11]. Perceptual results have shown that the perceived quality decreased according to listener remoteness from the music sources [11]. Furthermore, beam-forming on the wearable microphone arrays did not improve the music quality as it impacted the music's natural fidelity [11]. The perceptual evaluation was limited to four locations in the venue, excluding the evaluation across listener directions. Meanwhile, Seat Sublime produced a QoE map across concert hall seats based on an individualised weighted distribution of acoustical, visual, and physical characteristics [12]. However, the QoE maps were not perceptually validated.

The perceptual quality of spatial audio signals can be predicted with the use of objective metrics. For instance, Ambiqual is a full-reference quality metric suitable for ambisonics data format [13]. By comparing the phase angle spectrograms, i.e. phaseograms, of a test signal against a reference signal, the listening and localisation quality can be derived. Meanwhile, Binaqual applies the same principle for binaural audio signals [14]. Both metrics strongly correlated with human perceptual evaluations [13, 14]. Another study proposed a spatial audio quality impairment metric known as the signal to spatial distortion ratio (SSR) [15]. It was shown to align with the signal-to-distortion ratio (SDR) metric for music source separation [15], but it was not validated with perceptual listening tests.

To synthesise spatial audio, existing tools can be utilised. For instance, Spatial Scaper is a library to simulate soundscape events based on 6DoF source and receiver locations, improving sound event localisation and detection for data-driven algorithms [16]. However, the program remains limited to multiple monaural room impulse responses (RIR), bypassing head-related transfer functions (HRTFs) for listener-centric binaural reproduction. In that regard, The SADIE Database offers a collection of 20 HRTFs (18 from humans) with

[1] https://github.com/QxLabIreland/naviqual

an option to incorporate room acoustics through binaural room impulse responses (BRIR) [17], allowing a more immersive listener experience through headphones. Perceptual tests revealed that the KU100 mannequin HRTFs provide superior audio quality compared to individualised HRTFs from the participants [17]. Taking a step further, the Sonicom HRTF dataset boasts an even greater number of HRTFs from 120 subjects, enabling modern machine-learning approaches to be investigated [18]. Utilising existing HRTF data sets, Binamix is a Python program to simulate binaural signals from SADIE [19]. It incorporates angle interpolations to allow continuous azimuth and elevation reproduction on missing IR angles [19]. However, the program is currently limited to 3DoF synthesis, excluding listener translations. Finally, Paraspax is a MATLAB program that simulates 6DoF BRIRs with listener translations and directions based on a single monaural RIR using parametric techniques [20]. However, perceptual studies have shown that the synthesised BRIRs were distinguishable from the reference BRIRs [20].

By outlining these research gaps, we highlight the novelty and relevance of Naviqual as a first step in providing spatial audio quality distributions for virtual live music environments with perceptual validations.

3 Methodology

In this work, we synthesised binaural audio signals based on listener location and direction within a virtual environment. We created objective quality maps for this virtual environment based on localisation quality (LQ) and signal-to-noise ratio (SNR). Then, we validated these maps with human perceptual evaluations through listening tests on representative listening locations and directions. Finally, we investigated the correlation between the objective metrics and perceptual results.

3.1 Spatial Audio Synthesis

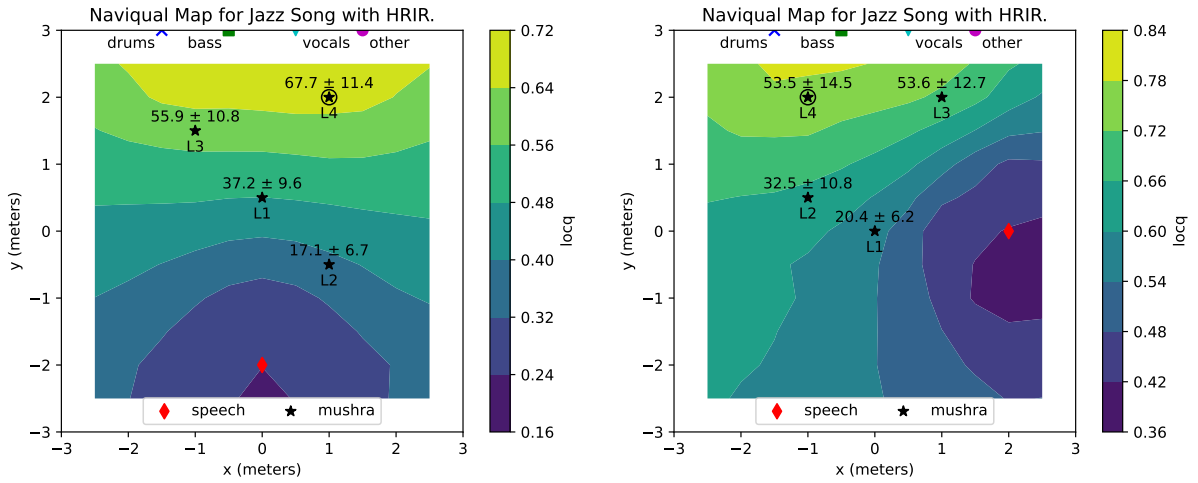
We used Binamix [19] to synthesise the binaural signals. This program uses head-related impulse responses (HRIRs) and BRIRs taken from the SADIE database [17]. Binamix generates a binaural signal by mixing a collection of tracks with the following parameters: azimuth, elevation, level, and reverberation.

In our live music simulation, we mixed down M music tracks $x_1(t), x_2(t), \ldots, x_M(t)$, with a distractor speech track $x_0(t)$, to generate the binaural signal $y(t)$ for the listener. We constrained our setup on the azimuth plane and did not apply source reverberations, thereby setting the elevation and reverberation parameters for all tracks to zero.

We calculated the azimuth and level parameters based on the listener location (x,y) and direction ϕ. Given source location (x_k, y_k), the source azimuth ϕ_k can be calculated based on the angle between (x,y) and (x_k, y_k), subtracted by ϕ. Meanwhile, the source level can be determined based on the inverse square law, i.e. the sound pressure level decreases by 6 dB for every doubling of the source distance to the listener, assuming lossless conditions. We assumed that the source level is maximum for a distance within 1 m based on the IASIG I3DL2 inverse distance clamped model [21]. Finally, we accounted for the sources' time delay of arrival based on their distances to the listener, assuming a sound velocity of $v_s = 343$ m/s.

Based on the source azimuth parameters, we obtained the corresponding HRIR or BRIR from the SADIE database $h_k(t)$, and convolved it with the input sound source $x_k(t)$, to produce the binaural sound source $y_k(t)$. We used the KU100 mannequin IR points, with 8802 HRIR points and 50 BRIR points, considering that it provided the superior perceptual quality among the HRTF set [17]. We applied *planar* angle interpolation for missing angles from the SADIE database, especially for BRIR synthesis which only contains 8 points with 45° azimuth resolution on the zero-elevation plane. Finally, the output binaural signal is obtained by summing the binaural sound sources after convolution.

In our virtual environment (see Fig. 1 and Fig. 2), the listener can navigate within a 5 m × 5 m space. The listener's experience was synthesised based on their location and direction. Four music sources corresponding to *drums*, *bass*, *vocals*, and *other* were placed within a stage area placed north beyond the listening space, at locations $(-1.5, 3)$, $(-0.5, 3)$, $(0.5, 3)$, and $(1.5, 3)$, respectively. Although available, we did not use the mastered mixes rendered on traditional loudspeaker setups, e.g. left and right speakers, to evaluate the robustness of Naviqual for a complex multi-point music soundscape. Meanwhile, the distractor speech source was positioned within the listening space, either at the south (S180) or east side (S270) at locations $(0, -2)$ and $(2, 0)$ respectively. We initially normalised the input sound sources $x_k(t)$ to -30 loudness units relative



(a) Jazz, HRIR, S180

(b) Jazz, HRIR, S270

Fig. 1: LQ contour maps and the MUSHRA scores of a jazz song with south (S180) and east (S270) distractor sources under anechoic conditions (HRIR).

to full scale (LUFS), and adjusted their loudness based on their distances to the listener.

We sourced the music tracks from the MUSDB18-HQ data set [22]. We selected three random songs with different genres, i.e. Jazz, Electronic (Elec), and Pop Rock (PopR). For listening test purposes, we segmented 7-second excerpts from these songs containing activity for all the music sources. Meanwhile, the speech tracks were sourced from the clean subset of the Microsoft Scalable Noisy Speech Dataset (MS-SNSD) [23]. We selected a random speech track for every sound scene generation.

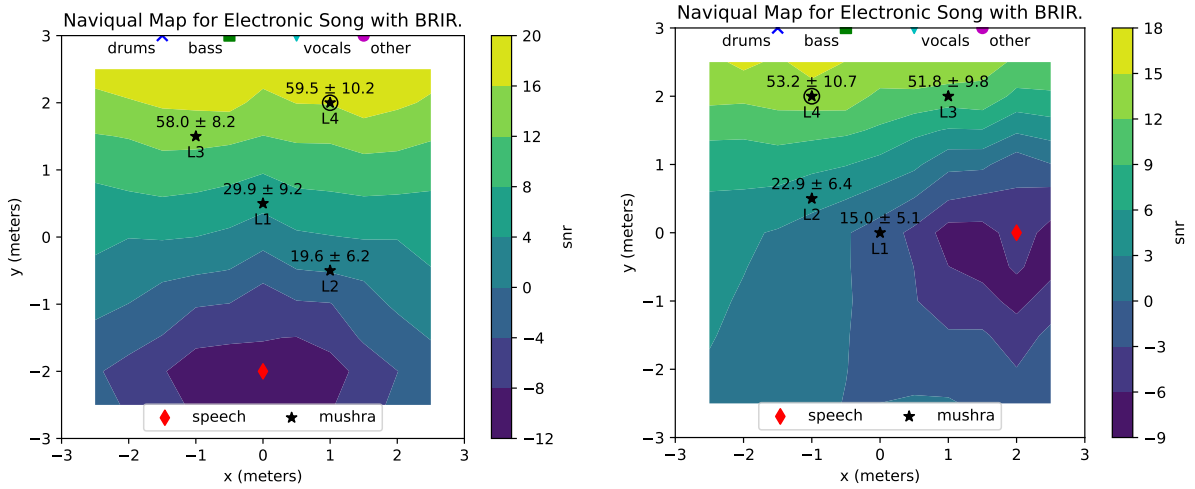
3.2 Objective Metrics

We used two objective metrics for the quality map generation, namely SNR (in LUFS) and localisation quality (LQ) derived from Binaqual [14]. The former measures the perceptual loudness difference between the uncorrupted binaural signal $y_{\text{music}}(t)$ and the distractor source $y_{\text{speech}}(t)$. Meanwhile, the latter measures the phaseogram degradations on the output binaural signal, i.e. $y(t) = y_{\text{music}}(t) + y_{\text{speech}}(t)$, compared to the uncorrupted reference signal $y_{\text{music}}(t)$. In general, the former mainly evaluates on intensity, while the latter mainly evaluates on phase.

We investigated how these metrics align with perceptual scores from listening tests. However, conducting a listening test across multiple listener locations, directions, noise locations, music content, and room acoustics can be cumbersome. Hence, we selected representative locations and directions in our virtual environment for the listening tests. Specifically, we selected two laterally varying (L2 and L4), two longitudinally varying (L3 and L4), and one central listening point (L1), all in the forward-facing direction $\phi = 0$. Then, we varied the listener direction on a selected fixed location (L4). We selected a side-facing direction (D1), an intermediate direction facing away from the distractor source (D2), a forward-facing direction (D4), and its 180-degree counterpart (D3). We used the objective metrics as a supplementary basis for selecting the listening points, ensuring that the perceptual evaluations span the listening space and the range of objective metric values.

3.3 Listening Tests

We conducted two listening test sessions inspired by the ITU-R BS.1534-3 MUltiple Stimuli with Hidden Reference and Anchor [24] through the web-based platform GoListen [25]. The two sessions focused separately on perceptual evaluations under anechoic (HRIR) and reverberant (BRIR) conditions. We recruited 15 and 12 normal-hearing adults for sessions one and two respectively representing various age brackets (18 – 25 years: 1, 1; 26 – 35 years: 6, 6; 36 – 45 years: 6, 3; 46 – 55 years: 2, 2). All participants have done MUSHRA-style listening tests on the GoListen platform prior to



(a) Elec, BRIR, S180

(b) Elec, BRIR, S270

Fig. 2: SNR contour maps and the MUSHRA scores of an electronic song with south (S180) and east (S270) distractor sources under reverberant conditions (BRIR).

the listening tests. They were requested to answer the tests at a quiet location with quality headphones.

Each trial contained four binaural signals under the following conditions: 2 IR types $\times$ 2 listener motion types $\times$ 2 distractor locations $\times$ 3 songs, a total of twenty four trials separated by two sessions. We employed a randomised block design, wherein the first block of six trials contained signals across listener locations, while the second block contained signals across listener directions. In the first block for both sessions, we included a hidden reference and anchor signal to assess listener reliability. We used the input music mixture $x_{\text{music}}(t)$ as the reference signal, and its corrupted version $x_{\text{music}}(t)+x_0(t)$ with -10 LUFS SNR as the anchor signal. For the post-screening, we excluded the participants whose ratings on at least two reference signals were below 80, or whose ratings on at least two anchor signals were above 40. We normalised the binaural signals to -30 LUFS using *pyloudnorm* [26] to eliminate loudness biases in the evaluation. Based on this post-screening criteria, we acquired 13 and 12 participant ratings for sessions one and two respectively.

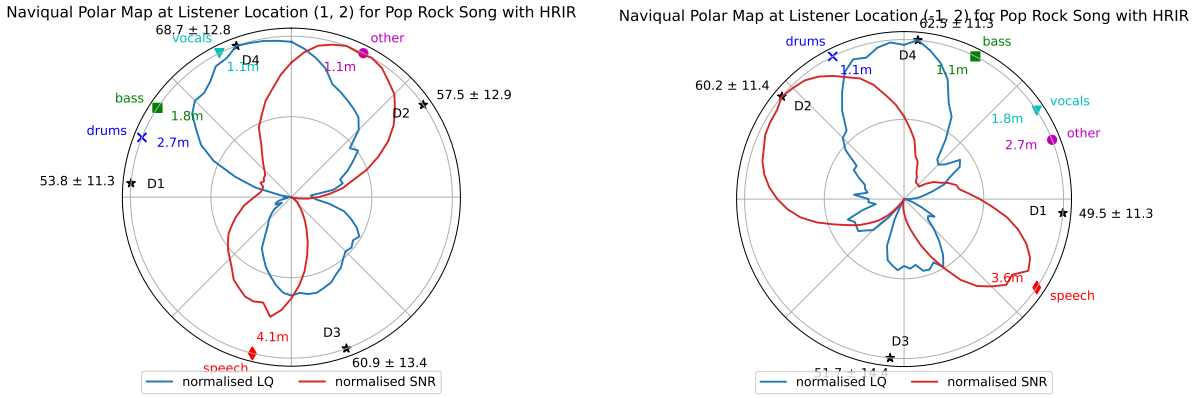
4 Results and Discussion

4.1 Quality Across Listener Locations

We created spatial audio quality maps for our virtual environment by calculating the objective metrics on a discrete set of listener locations and directions (RQ1). Fig. 1 shows the quality contour maps for a jazz song across listener locations facing north, i.e. $\phi = 0$, under anechoic conditions. These were generated using the *contourf* function of the *matplotlib* python package based on the Binaqual LQ values of signals synthesised at 0.5 m lateral and longitudinal resolution within the listening space. We also annotated the listening test points on each map with their average MUSHRA scores and 95% confidence intervals.

From these maps, we can observe how LQ changes according to the listener location and how it aligns with the MUSHRA scores. We can see that the optimal QoE lies within regions near the music sources, i.e. northwards. Correspondingly, these high QoE spots lie within regions away from the speech source, i.e. northwards for S180 and westwards for S270. Interestingly, we can also observe quality dependencies according to the music source positions, particularly from Fig. 1b wherein the listeners rated L3 more closely with L4 despite lying on different LQ levels. We hypothesise that some participants rated L3 higher than L4 due to the enhanced vocals despite having louder speech intrusion.

On the other hand, Fig. 2 shows the quality contour maps for an electronic song under reverberant conditions, generated based on SNR values. We can observe



(a) PopR, HRIR, S180

(b) PopR, HRIR, S270

Fig. 3: LQ and SNR polar maps and the MUSHRA scores of a pop rock song under anechoic conditions.

uneven contour lines which we attribute to the sparse BRIR azimuth resolution of the SADIE database. Nevertheless, the optimal QoE locations followed a similar trend with the anechoic conditions.

We observed that both LQ and SNR contour maps follow similar patterns across song genres and room conditions. We hypothesise that the fixed listener direction ($\phi = 0$) across listener locations primarily affects the intensity of the sound sources with minor effects on the phase, resulting in similar contour maps.

Based on these graphs and also evidenced in Table 1, we can confirm that the objective quality maps aligned with perceptual scores and are robust against noise locations, music content, and room acoustics (RQ2).

4.2 Quality Across Listener Directions

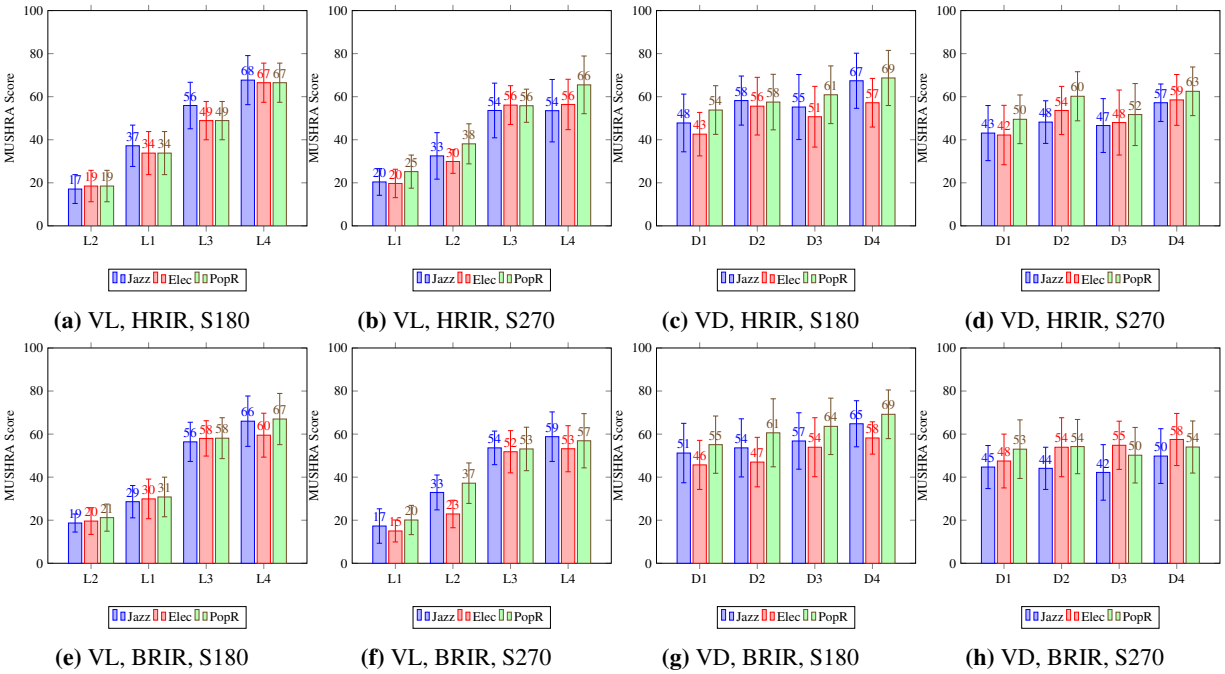
Subsequently, Fig. 3 shows quality polar maps across listener directions on a fixed listener location (L4) of a pop rock song under anechoic conditions. These were generated based on the LQ and SNR values of binaural signals synthesised across listener directions with 5° azimuth resolution. The objective metric values were normalised between zero and one to better visualise the optimal listening directions. The representative directions for the listening tests were annotated with their average MUSHRA scores and 95% confidence intervals. The sources were also placed along the map based on their azimuths, with their distances to the listener annotated.

From these polar maps, we can observe that the optimal perceptual quality lie on the northward direction, while the sub-optimal perceptual quality lie on the eastward or westward directions, facing the distractor source. We can also observe reduced quality for 180° counterparts, signifying front-back quality differences. Lastly, the intermediate directions provide mid-range quality scores.

Furthermore, we can observe subtle differences between the LQ and SNR values on these polar maps. For instance, the SNR polar maps mainly depend on the direction of the distractor source, with the minimum obtained from the directions perpendicular to it, and the maximum obtained from the directions facing away from it. This signifies that the intensity of the distractor source is maximised when an ear is directed towards the distractor source. Meanwhile, the LQ value is maximised depending on how well the music sources are distributed across the binaural signals while facing away from the distractor source. These observations signify that the optimal QoE direction can be based on multiple factors which include the distractor intensity and the balance of the music sources, similar to our findings from the contour maps. In this case, the Binaqual objective metric outperforms SNR in predicting the perceptual quality across listener directions, which can also be observed in Table 1.

4.3 Objective Metrics and Subjective Scores

Fig. 4 shows the participants' MUSHRA scores at the representative locations and directions in our virtual



(a) VL, HRIR, S180 **(b)** VL, HRIR, S270 **(c)** VD, HRIR, S180 **(d)** VD, HRIR, S270

(e) VL, BRIR, S180 **(f)** VL, BRIR, S270 **(g)** VD, BRIR, S180 **(h)** VD, BRIR, S270

Fig. 4: MUSHRA scores at various listener locations (VL) and directions (VD) across song genres, IR types, and noise locations.

environments. We can observe that perceptual QoE is primarily influenced by the listener location, followed by the listener direction which has a more subtle effect. In our study, we were interested on how the relative QoE scores were affected when varying the song genre, noise locations, and room reverberation (RQ2). Comparing the top and bottom rows, we can see that the score patterns remain similar for both anechoic and reverberant conditions. We can also see that the song genre does not largely affect the relative scores. This indicates that the perceptual quality variation within a pre-defined virtual environment is largely similar even when adding room reverberation or changing the music content.

We also validated whether the objective metrics that we used, i.e. Binaqual and SNR, can accurately predict these perceptual QoE scores (RQ3). Table 1 shows the extent of how the objective and subjective scores aligned based on their Pearson and Spearman correlation coefficients. We can observe that both metrics strongly correlate with the perceptual scores across listener locations, noise locations, song genres, and IR types, except for the listener directions. We can observe that LQ surpasses SNR in estimating the perceived quality across listener directions. We hypothesise that listeners rely more on inter-aural phase difference cues rather than intensity cues when evaluating the quality across directions, which can be measured more through Binaqual compared to SNR. A notable exception was found for a pop rock song with S270 speech source, wherein listeners preferred facing away from the speech source, which in this case the SNR metric surpasses Binaqual. Again, we note that this outlier may have also been affected by the sparse BRIR resolution of the data set. Nevertheless, we believe that an objective metric that combines both intensity and phase evaluations can be a better quality predictor compared to stand-alone metrics.

5 Conclusion

In this work, we evaluated the spatial audio quality of virtual live music environments across listener translations and directions with objective metrics and perceptual tests. We were able to generate objective quality contour and polar maps that strongly aligned with the perceptual quality scores from the listening tests. We highlighted their robustness across different sound

		VL, HRIR		VD, HRIR		VL, BRIR		VD, BRIR	
		LQ	SNR	LQ	SNR	LQ	SNR	LQ	SNR
Jazz	S180	1.00, 1.0	1.00, 1.0	0.99, 0.8	0.66, 0.8	1.00, 1.0	0.99, 1.0	0.99, 1.0	0.88, 0.8
	S270	0.96. 0.8	0.97, 0.8	0.96, 0.8	-0.26, 0.0	0.99, 1.0	1.00, 1.0	0.88, 0.8	-0.56, 0.0
Elec	S180	0.99, 1.0	0.98, 1.0	0.91, 1.0	0.57, 0.6	0.98, 1.0	0.99, 1.0	0.94, 1.0	0.54, 0.8
	S270	0.93, 1.0	0.97, 1.0	0.92, 0.8	0.43, 0.6	0.94, 1.0	0.97, 1.0	0.79, 1.0	-0.80, -0.8
PopR	S180	0.99, 1.0	1.00, 1.0	0.96, 1.0	0.27, 0.4	1.00, 1.0	0.99, 1.0	0.99, 1.0	0.25, 0.0
	S270	0.99, 1.0	1.00, 1.0	0.85, 1.0	0.44, 0.0	1.00, 1.0	0.99, 1.0	0.28, -0.2	0.83, 0.4

Table 1: Pearson and Spearman correlation coefficients between perceptual scores and objective metrics on different sound scenes across various listener locations (VL) and directions (VD).

scenes, i.e. listener locations and directions, noise locations, music content, and room reverberation. We found that both Binaqual and SNR strongly predict quality across listener translations, but Binaqual outperforms SNR in predicting quality across listener directions.

This work provides insights on room acoustic design with considerations on the spatial distribution of quality. For instance, Naviqual can be used as an initial step to identify loudspeaker placements in performance venues that optimises the acoustic quality across locations and orientations, addressing the need to conduct multi-point measurements and calibrations across venues, performances, and scenarios. It also provides insights on addressing quality degradation issues in live music environments contributed by external uncontrolled factors such as audience noise and room reverberation. Addressing these issues would enable both normal-hearing and hard-of-hearing individuals to access and enjoy live music. Finally, Naviqual can be used to streamline the design processes in AR environments for spatial audio perceptual studies, taking out the guesswork on spatial audio quality distribution.

Interesting future work on this project includes generating quality maps for more complex sound scenes, e.g. multiple noise sources. We plan to adapt the program to a wearable AR device wherein the listener can navigate the listening environment and evaluate the spatial audio quality on-the-fly, which can be validated by the objective quality maps. Finally, we plan to conduct AR perceptual studies with the hard-of-hearing community as well.

6 Acknowledgments and Ethical Compliance

This work was funded by the European Commission under the Marie Skłodowska-Curie Actions postdoctoral fellowship grant ID 101151676 (ARM4CI). Ethical approval for the subjective listening tests was obtained from the UCD Human Research Ethics Committee.

References

[1] Kok, M., Hill, A., Burton, J., and Mulder, J., "The influence of audience participatory noise on sound levels at live events." in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 265, pp. 1947–1954, Institute of Noise Control Engineering, 2023.

[2] Moore, B. C., "Listening to music through hearing aids: Potential lessons for cochlear implants," *Trends in Hearing*, 26, p. 23312165211072969, 2022.

[3] Tronchin, L., Bevilacqua, A., and Yan, R., "Acoustic Characterization and Quality Assessment of Cremona's Ponchielli Theater," *Applied Sciences*, 13(6), 2023, ISSN 2076-3417, doi: 10.3390/app13064057.

[4] Wiggins, I. M. and Liston, K. I., "SOUND DISTRIBUTION FOR SAFE LISTENING IN ENTERTAINMENT VENUES: A REVIEW OF EXISTING POLICIES AND REGULATIONS," in *ACOUSTICS 2020*, Institute of Acoustics, 2020, doi:10.25144/13328.

[5] Riionheimo, J. and Lokki, T., "Spatial Sound, Part 1: Exploring the Impact of Listening Room on Perceptual Immersion," *Journal of the Audio Engineering Society*, 72(12), pp. 913–929, 2024.

[6] Riionheimo, J. and Lokki, T., "Spatial Sound, Part 2: Exploring the Impact of Listening Room on Sound Character," *Journal of the Audio Engineering Society*, 72(12), pp. 930–944, 2024.

[7] Hirway, A., Qiao, Y., and Murray, N., "A Quality of Experience and Visual Attention Evaluation for 360 videos with non-spatial and spatial audio," *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(9), pp. 1–20, 2024.

[8] Jot, J.-M., Audfray, R., Hertensteiner, M., and Schmidt, B., "Rendering spatial sound for interoperable experiences in the audio metaverse," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–15, IEEE, 2021.

[9] Mehra, R., Brimijoin, O., Robinson, P., and Lunner, T., "Potential of augmented reality platforms to improve individual hearing aids and to support more ecologically valid research," *Ear and hearing*, 41, pp. 140S–146S, 2020.

[10] Ivanyi, B., Tsalidis, C., Naylor, S., Tjemsland, T. B., Adjorlu, A., Kepp, N. E., and Serafin, S., "HoloBand: An Augmented Reality Experience to Train Music Perception for the hard-of-hearing," in *2022 AES International Conference on Audio for Virtual and Augmented Reality, AVAR 2022*, pp. 1–10, Audio Engineering Society, 2022.

[11] Ragano, A., Tolentino, C. T., Szita, K., Barry, D., Panah, D. S., Murray, N., and Hines, A., "EgoMusic: An Egocentric Augmented Reality Glasses Dataset for Music," in *Proceedings of the 33rd ACM International Conference on Multimedia*, MM '25, p. 13104–13111, Association for Computing Machinery, New York, NY, USA, 2025, ISBN 9798400720352, doi: 10.1145/3746027.3758262.

[12] Salesin, J., Lokki, T., and Echenagucia, T. M., "Seat sublime: Focusing on the totality of the concertgoer experience–acoustical, visual, and physical," in *Proceedings of Meetings on Acoustics*, volume 58, p. 015001, Acoustical Society of America, 2025.

[13] Narbutt, M., Skoglund, J., Allen, A., Chinen, M., Barry, D., and Hines, A., "AMBIQUAL: Towards a Quality Metric for Headphone Rendered Compressed Ambisonic Spatial Audio," *Applied Sciences*, 10(9), 2020, ISSN 2076-3417, doi: 10.3390/app10093188.

[14] Panah, D. S., Barry, D., Ragano, A., Skoglund, J., and Hines, A., "BINAQUAL: A Full-Reference Objective Localization Similarity Metric for Binaural Audio," 2025, accepted for publication in the Journal of the Audio Engineering Society.

[15] Watcharasupat, K. N. and Lerch, A., "Quantifying spatial audio quality impairment," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 746–750, IEEE, 2024.

[16] Roman, I. R., Ick, C., Ding, S., Roman, A. S., McFee, B., and Bello, J. P., "Spatial scaper: a library to simulate and augment soundscapes for sound event localization and detection in realistic rooms," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1221–1225, IEEE, 2024.

[17] Armstrong, C., Thresh, L., Murphy, D., and Kearney, G., "A perceptual evaluation of individual and non-individual HRTFs: A case study of the SADIE II database," *Applied Sciences*, 8(11), p. 2029, 2018.

[18] Isaac, E., Rapolas, D., Thibault, V., T., H. A. O., Johan, P., J., T. A., and Lorenzo, P., "The SONICOM HRTF Dataset," *Journal of the Audio Engineering Society*, 71, pp. 241–253, 2023.

[19] Barry, D., Panah, D. S., Ragano, A., Skoglund, J., and Hines, A., "Binamix – A Python Library for Generating Binaural Audio Datasets," *Journal of the Audio Engineering Society*, (319), 2025.

[20] Arend, J. M., Garí, S. V. A., Schissler, C., Klein, F., and Robinson, P. W., "Six-degrees-of-freedom parametric spatial audio based on one monaural room impulse response," *Journal of the Audio Engineering Society*, 69(7/8), pp. 557–575, 2021.

[21] Software, L., "OpenAL 1.1 specification and reference," 2005.

[22] Rafii, Z., Liutkus, A., Stöter, F.-R., Mimilakis, S. I., and Bittner, R., "MUSDB18-HQ - an uncompressed version of MUSDB18," 2019, doi: 10.5281/zenodo.3338373.

[23] Reddy, C. K., Beyrami, E., Pool, J., Cutler, R., Srinivasan, S., and Gehrke, J., "A Scalable Noisy Speech Dataset and Online Subjective Test Framework," *Proc. Interspeech 2019*, pp. 1816–1820, 2019.

[24] "Method for the subjective assessment of intermediate quality level of audio systems," Standard ITU-R BS.1534-3, ITU-R, Geneva, 2015.

[25] Barry, D., Zhang, Q., Sun, P. W., and Hines, A., "Go listen: an end-to-end online listening test platform," *Journal of Open Research Software*, 9(1), 2021.

[26] Steinmetz, C. J. and Reiss, J. D., "pyloudnorm: A simple yet flexible loudness meter in Python," in *150th AES Convention*, 2021.