forum acusticum 2026
8-12 September · Graz, Austria

AUDITORY-MODEL PREDICTIONS OF QUALITY OF EXPERIENCE IN IMMERSIVE MUSIC CONCERTS

Nicolas Epain[1], Etienne Corteel[1]

[1] *L-Acoustics, 13 rue Levacher Cintrat, 91460 Marcoussis, France*

Abstract

Immersive live music productions promise three key benefits for audiences: audio-visual consistency (you hear sound where you see it), improved source separation and intelligibility through spatial unmasking, and new creative possibilities for mixing engineers. Yet it remains unclear how to assess whether a given venue is suitable for this type of production, based on its measurable acoustic characteristics. Using virtual acoustics and binaural auditory models, we analyze how the presence of diffuse reverberation, specified by a Reverberation Time (RT), direct-to-diffuse ratio (DRR) and Initial Time Delay Gap (ITDG) govern three perceptual predictors of Quality of Experience: source-localization accuracy, localization blur, and binaural advantage. In our simulation results, we observed that the DRR was the primary determinant of all three predictors, while the RT played a secondary role. ITDG primarily reduced localization blur, with minor effects on accuracy and spatial unmasking. Its benefits were most apparent at moderate DRRs. These results help define what makes a venue good or poor for immersive live events from a QoE standpoint and provide actionable criteria to secure the "what you hear is what you see" promise and enhance spatial unmasking in production.

Keywords: *reverberation, spatialization, localization, spatial unmasking, auditory models*

1. Introduction

Relative to conventional left–right production, immersive live sound offers improved audio–visual consistency, greater separation and intelligibility via spatial unmasking, and increased creative freedom for mixing engineers [1]. These advantages ultimately depend on the audience's capacity to perceive multiple sources simultaneously in three dimensions —a capacity gated not only by system design and rendering algorithms, but also by the venue's acoustic characteristics, which shape sound propagation temporally, spectrally, and spatially. This motivates the question: which room-acoustic characteristics determine a venue's suitability for immersive concerts, from a QoE standpoint?

Room acoustics shape spatial hearing via the balance and timing of direct and reflected energy. In this study we focus on three descriptors: the direct-to-reverberant energy ratio (DRR), the reverberation time (RT), and the initial time-delay gap (ITDG). Prior work indicates that DRR strongly conditions localizability and perceived clarity [2], while RT modulates spaciousness and can degrade cues when the late field dominates [3]. Early-to-late timing determines whether reflections stabilize or blur the auditory image via precedence mechanisms: for short delays the first wavefront anchors perceived direction and later energy is down-weighted [4]. Consequently, longer ITDGs primarily reduce localization blur, especially when DRR is modest.

Room acoustic conditions also impact separation and intelligibility. The ability to spatially separate a target from an interferer yields benefits through a better-ear SNR and binaural unmasking, but room reflections incoming from many directions can severely impede this effect. This can be estimated using perceptual models such as that proposed by Jelfs [5]. While these models are mostly focusing on speech, their mechanisms remain relevant as predictors of separation for complex program material such as music, provided results are interpreted comparatively across conditions. In parallel, binaural localization models such as that proposed by Dietz [6] enable repeatable estimation of direction and uncertainty from binaural signals, linking reverberation parameters to audio-visual consistency.

This paper proposes a model-based evaluation framework that ties venue acoustics to three perceptual predictors of QoE for immersive concerts: (1) localization accuracy, (2) localization blur (image stability), and

***Corresponding author**: nicolas.epain@l-acoustics.com.*

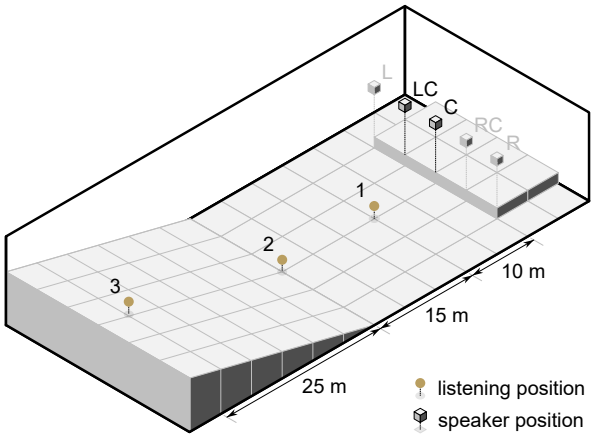

Figure 1: Visual representation of the simulation geometry.

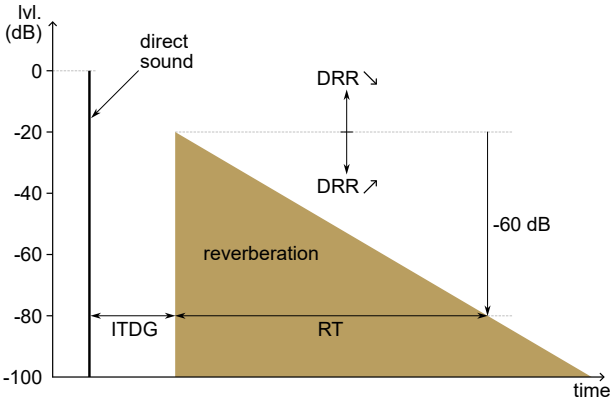

Figure 2: Model employed for the generation of the Binaural Room Impulse Responses.

(3) binaural advantage (a proxy for spatial unmasking). We used synthetic room impulse responses (RIR) to manipulate DRR, RT, and ITDG independently and applied established binaural models for direction estimation and unmasking. We then mapped how these room acoustic parameters govern spatial predictors, which we assume to closely align with QoE. Our results are intended to help practitioners measure what matters in a venue, predict QoE across the audience, and inform system design when constraints prevent ideal conditions.

The remainder of the paper is organized as follows. Section 2 outlines our simulation methodology. Section 3 reports simulation results. Section 4 translates our findings into room acoustics analysis guidance and discusses limitations.

2. Numerical simulations

In this section, we describe the geometry, synthetic BRIR construction, stimuli and binaural models we used to link room acoustic parameters to spatial predictors.

2.1. Simulation geometry

Figure 1 illustrates the simulation setup used in our study. We modeled the minimal L-ISA deployment in which 5 stacks of loudspeakers span the entire width of the scene. In our model, the scene is 20 m wide and the sources are regularly spaced, therefore the distance between two adjacent sources is 5 m. We only consider the source located at the center of the stage (C) and its neighbor on the left side (LC) in the simulations. These correspond to the positions where vocals and some of the main competing instruments would be panned in typical shows, and it constitutes a critical scenario in terms of spatial separation.

We consider three listening positions in the virtual venue, all three being aligned with the center of the stage. Position 1, 2 and 3 are located 10, 25 and 50 m from the stage, respectively. For simplicity we assumed that the sources were located at ear height for every position. Hence, Source LC is located at azimuth 26.6°, 11.3° and 5.7° relative to positions 1 to 3.

2.2. Room impulse response model

We model the room impulse responses (RIRs) as illustrated in Fig. 2. In our simulations the room reverberation only consists of a diffuse reverberation tail. This tail is characterized by: the DRR, which sets the overall energy of the reverberated sound field with respect to the direct sound; the RT, which defines the slope of the reverb energy decay, modeled as exponential; and the ITDG, which we define as the time delay separating the onset of the direct sound impulse and the beginning of the reverb tail.

2.3. Binaural RIR synthesis

To synthesize Binaural RIRs (BRIRs), we started by generating the part corresponding to the reverberated sound field. For each of the two sources, we first synthesized two channels of mutually uncorrelated Gaussian white noise, with duration 1.5 times the RT. We then converted these signals to the time-frequency domain using a short-time Fourier transform (STFT). In the time-frequency domain, we applied an exponential decay matching a constant RT value τ for frequencies ranging from 0 to 8 kHz, and matching an RT linearly decreasing from τ to $\tau/4$ for frequencies between 8 KHz and the Nyquist frequency. A low-pass gain factor with a slope of -30 dB per octave was then applied to time frequency bins above 8 kHz, and an inverse STFT was applied to obtain two time-domain reverberant tails. Lastly, we convolved the obtained signals with a matrix of FIR filters calculated such that the output signal inter-correlation matched the frequency-dependent pattern observed for binaural signals in the presence of a diffuse field. A fade-in consisting of a half Hanning window was applied to the beginning of the tail. The duration of the fade-in was a quarter of the ITDG.

For the direct sound part, we used the set of anechoic HRIRs measured for the Neumann KU100 manikin by Bernschütz [7]. For each listening position, we

simply selected the pairs of measured HRIRs that best matched the direction of the sources. For source LC, this resulted in using the HRIRs measured for directions $(\theta, \varphi) = (0, 26°)$, $(0, 12°)$ and $(0, 6°)$ for listening positions 1, 2 and 3 respectively. The reverberant part was then normalized in amplitude relative to the direct part, such that the ratio of the energies matched the desired DRR value. Lastly, the direct and reverberant part of the BRIRs were concatenated, setting the time gap between the two to match the relevant ITDG.

2.4. Source localization metrics

To estimate the accuracy of source localization in the presence of reverberation, we first generated a source signal consisting of a series of 30 ms long bursts of pink-noise, at a rate of four bursts per second. The total duration of the signal was:

$$d = \frac{\lceil 4RT \rceil}{4} + 1 \,, \tag{1}$$

in other words the source signal was longer than RT and comprised of an integer number of 250 ms burst-silence sequences. This signal was then convolved with the BRIRs corresponding to Source LC, to obtain the binaural signals corresponding to a given listening position. The binaural signals were time-cropped, keeping only the last second of signals. Hence the cropped signals always started at the onset of a noise burst, consisted of exactly four bursts and contained reverberation caused by preceding bursts, modeling the audience experience when listening to a percussive signal in steady state conditions.

We used the model proposed by Wierstorf [6], [8] (as implemented in the Auditory Model Toolbox (AMT) for Matlab [9]) to predict the perceived source direction as well as the localization uncertainty (predicted standard deviation). The azimuthal error was then computed as the absolute angular distance between the predicted perceived azimuth and the actual direction of the source:

$$\mathrm{err}_{\mathrm{az}} = |\hat{\varphi} - \varphi_i| \,, \tag{2}$$

where $\hat{\varphi}_i$ denotes the predicted azimuth and φ_i is Source LC's direction relative to listening position i. Note that, in the following we refer to the localization uncertainty as the localization *blur*.

2.5. Spatial unmasking metric

In order to evaluate the benefit in clarity brought by the frontal system in the presence of reverberation we estimated the spatial unmasking, *i.e.* the increase in speech intelligibility of a target source when the target and interferer are spatially separated, using the model presented by Jelfs *et al* [5]. We used the implementation provided in the AMT, which takes the BRIRs of the target and interferer as input. In our model, Source C was the target while Source LC was the interferer, therefore we predicted the spatial unmasking based on the corresponding simulated BRIRs.

2.6. Factorial design

We evaluated the spatial QoE metrics for a total of 1215 parameter combinations, spanning:

- 9 RTs: 0.25, 0.35, 0.5, 0.71, 1, 1.41, 2, 2.83 and 4 s,
- 9 DRRs: -12, -9, -6, -3, 0, 3, 6, 9, and 12 dB,
- 5 ITDGs: 6.25, 12.5, 25, 50 and 100 ms, and

the 3 listening positions. For each of these parameter sets, we generated 50 sets of BRIRs using the framework described above. In the following we report the median of the metric values obtained for these 50 trials.

3. Results

In this section we report the results of our numerical simulations.

3.1. Source localization error

Figure 3 reports median azimuthal localization as a function of RT and DRR, with separate panels for ITDG and listening position. Two trends dominate. First, error is primarily governed by DRR: performance is near its floor at high DRRs and degrades rapidly as DRR decreases (i.e., as reverberant energy becomes dominant). Second, RT has a strong effect but it is not uniform across DRR: longer RT values disproportionately increase error when DRR is already low, whereas at high DRR the effect of RT is comparatively limited. Position also contributes a systematic shift in performance: errors are lowest at the front position and increase towards the back, consistent with the reduced source azimuth (hence reduced binaural cue salience) at larger listening distances in the chosen geometry. By contrast, ITDG seem to produce only modest changes in localization error.

To quantify these observations, we ran a factorial analysis of variance (ANOVA) on cell means (*i.e.*, averaging the 50 trials for each RT, DRR, ITDG, position combination). All parameters and interactions were statistically reliable (all $p \leq 0.00244$). In line with the figure, the largest F-ratios were obtained for DRR ($F(8, 1021) = 1063.3$, $p \approx 0$), RT ($F(8, 1021) = 403.0$, $p \ll 10^{-6}$), and the RT×DRR interaction ($F(64, 1021) = 65.86$, $p \ll 10^{-6}$), confirming that both factors matter, but also that their interaction is central to explaining localization error. Because our aim is to understand what drives venue suitability, we report effect magnitudes using η^2, defined here as the proportion of total variance explained by each term. Under this definition, DRR explains the largest share of variance in cell-mean error ($\eta^2 \approx 46\%$), followed by RT×DRR ($\eta^2 \approx 23\%$) and RT ($\eta^2 \approx 18\%$). Together, these three terms account for about 86% of the variance, which formalizes the figure-driven conclusion that energy balance (DRR) is the primary lever, and that RT is most consequential in interaction with DRR (i.e., it amplifies error mainly when the direct field is not dominant). All remaining effects are secondary in variance terms. Position explains

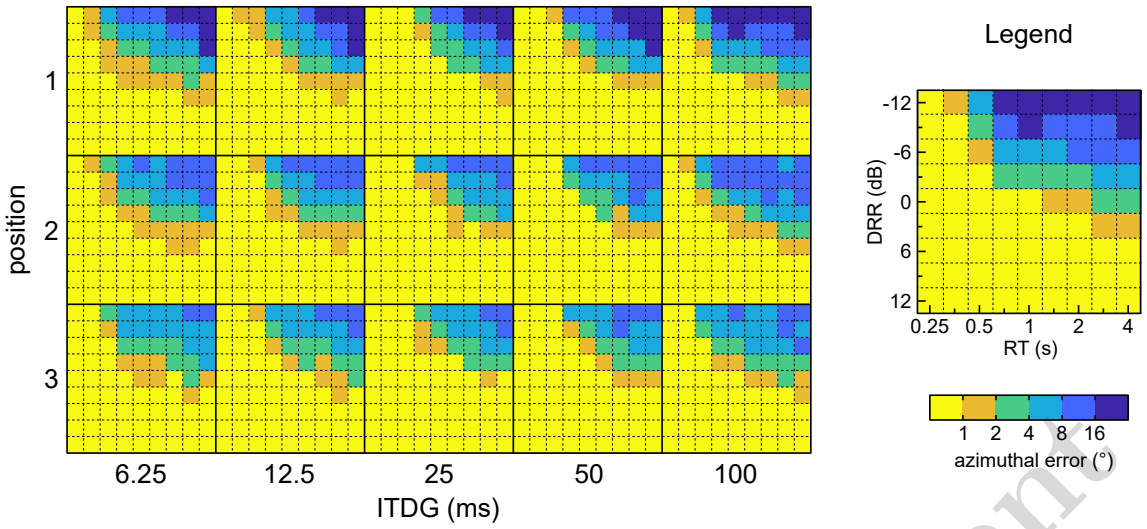

Figure 3: Source localization error, as a function of the RT, DRR, ITDG and position (brighter is better).

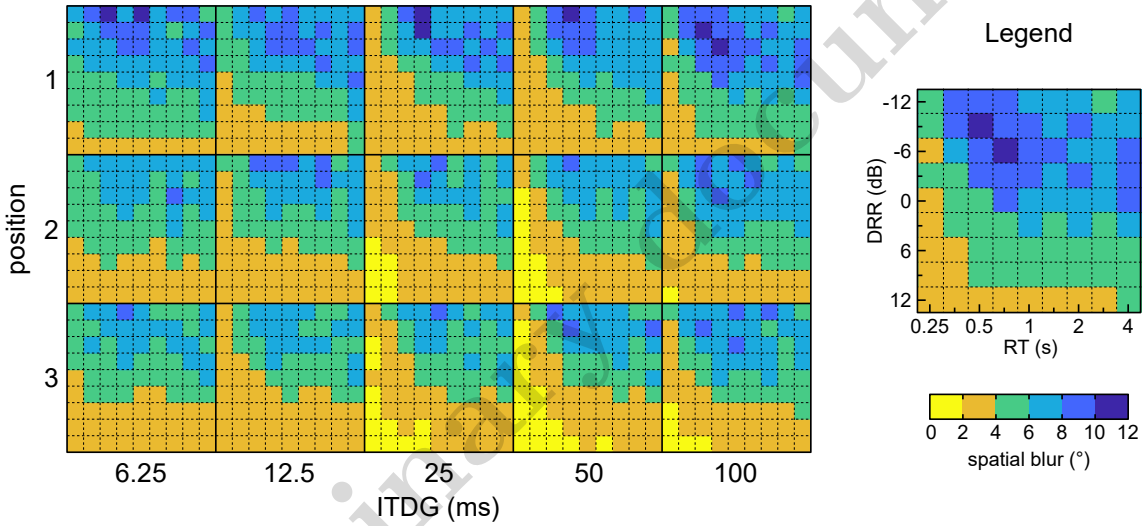

Figure 4: Source localization blur, as a function of the RT, DRR, ITDG and position (brighter is better).

1.4%, while DRR×position contributes 2.8%, consistent with a position-dependent sensitivity to DRR visible across panels. ITDG contributes 1.0% and its interactions are similarly small, indicating that ITDG produces statistically reliable but comparatively modest changes in accuracy within the explored parameter space.

3.2. Localization blur

Figure 4 reports localization blur as a function of RT and DRR, with panels for ITDG and listening position. Two trends are immediately apparent. First, blur is strongly governed by DRR: as DRR decreases, blur increases markedly, whereas at favorable DRR it collapses towards a low baseline across most RT values. Second, position induces a clear shift: blur increases from the front (Position 1) to the back (Position 3), consistent with progressively smaller source azimuths and thus weaker binaural cue salience.

Beyond these dominant effects, RT and ITDG shape blur in complementary ways. Increasing RT tends to increase blur, but this effect is most visible when DRR is not favorable (i.e., when the late field is sufficiently strong to perturb binaural cues). In contrast, ITDG produces a more selective, "stabilizing" effect: longer gaps generally reduce blur, with the most noticeable benefits occurring at intermediate DRR where blur is neither at floor nor fully dominated by reverberation. The ANOVA supports this hierarchy in terms of variance explained. The largest contribution is DRR ($\eta^2 \approx 57\%$; $F(8, 1021) = 1956.1$, $p \approx 0$), followed by RT ($\eta^2 \approx 20\%$; $F(8, 1021) = 686.9$, $p \approx 0$) and position ($\eta^2 \approx 6.5\%$; $F(2, 1021) = 893.0$, $p \ll 10^{-6}$). Interactions are present but smaller: RT×DRR accounts for 6.2% ($F(64, 1021) = 26.77$, $p \ll 10^{-6}$), indicating that the influence of RT depends on the

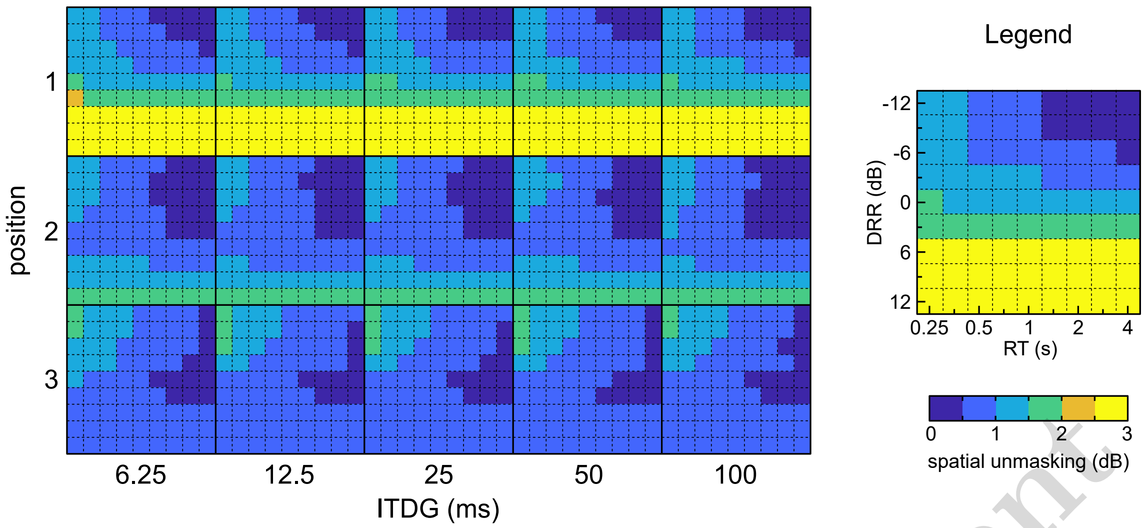

Figure 5: Spatial unmasking, as a function of the RT, DRR, ITDG and position (brighter is better).

direct-to-reverberant balance. Importantly, ITDG explains a non-negligible share of variance for blur ($\eta^2 \approx 3.1\%$; $F(4, 1021) = 215.7$, $p \ll 10^{-6}$), with an additional RT×ITDG contribution ($\eta^2 \approx 2.6\%$; $F(32, 1021) = 22.10$, $p \ll 10^{-6}$), consistent with the ITDG panels in Fig. 4. By contrast, ITDG×position is not supported ($F(8, 1021) = 0.26$, $p = 0.978$), suggesting that the effect of ITDG on blur is consistent across positions.

3.3. Spatial unmasking

Figure 5 reports spatial unmasking as a function of RT and DRR, with panels for ITDG and audience position. Interpreted relative to the anechoic baselines (Position 1/2/3: 5.88/2.35/0.91 dB), the maps primarily describe how much unmasking is retained or lost once reverberation is introduced: at each position, favorable DRR conditions tend to preserve higher values, whereas increasingly reverberant conditions (lower DRR and higher RT) reduce binaural advantage. The position dependence is therefore best understood as a combination of (i) a strong geometric ceiling already present in anechoic conditions (front≫mid≫back), and (ii) a room-dependent modulation that is most visible when DRR becomes unfavorable.

A notable exception appears for positions 2 and 3 at short RT, where decreasing DRR can locally yield higher predicted unmasking. A plausible interpretation is that, in these configurations (small target–interferer angular separation, hence low anechoic advantage), adding diffuse, incoherent energy may increase the effective binaural unmasking term by introducing additional decorrelation in the BRIRs used by the model. In other words, at small azimuth separations and short RT, a modest reverberant component can (counter-intuitively) act as a decorrelating "helper" in the model.

These observations are supported by the ANOVA. In F-ratio terms, the dominant effects are position ($F(2, 1024) = 1.56 \cdot 10^5$, $p \approx 0$), DRR ($F(8, 1024) = 3.92 \cdot 10^4$, $p \approx 0$), and the DRR×position interaction ($F(16, 1024) = 2.45 \cdot 10^4$, $p \approx 0$), indicating that the influence of DRR on unmasking is itself strongly position-dependent. Expressed as a proportion of total variance, this hierarchy is even clearer: DRR×position explains 36%, DRR 29%, and position 29%. In other words, more than 90% of the variance in binaural advantage is accounted for by DRR, position, and their interaction, which matches the pattern seen in Fig. 5.

By comparison, RT contributes a smaller but still reliable effect ($F(8, 1024) = 5463.9$, $p \approx 0$), corresponding to $\eta^2 \approx 4\%$. The RT×DRR interaction is also reliable ($F(64, 1024) = 401.6$, $p \approx 0$; $\eta^2 \approx 2.4\%$), consistent with RT primarily modulating unmasking when the direct-to-reverberant balance is not favorable. In contrast, ITDG has a negligible influence on spatial unmasking in both the maps and the variance partitioning: while the main effect reaches significance ($F(4, 1024) = 16.38$, $p \ll 10^{-6}$), its magnitude is minimal ($\eta^2 \approx 6 \cdot 10^{-5}$), and neither DRR×ITDG nor ITDG×position are supported ($p = 0.804$ and $p = 0.965$, respectively). Overall, unmasking in these simulations is dominated by energy balance and geometry (DRR and position), with RT exerting a secondary modulation and ITDG contributing little to the binaural advantage metric.

3.4. Summary of our results

We now summarize the simulation outcomes and translate them into practical QoE implications for immersive concerts, *i.e.* the conditions under which audio-visual consistency and source separation can be expected to hold across the audience. We focus on the DRR for two reasons: (i) it is the main driver of the perceptual metrics in our results, and (ii) unlike RT, which is characteristic of the venue, DRR can be shifted by system design choices such as source

directivity control and coverage management.
This yields the following rules of thumb:

1. when DRR≥0 dB, azimuthal localization error remains below roughly 4° (i.e., comparable to typical human accuracy for broadband bursts in frontal directions), essentially regardless of RT and ITDG within the explored range;

2. when DRR≥0 dB and ITDG is within a normal range (12.5–50 ms), spatial blur is constrained to a moderate range (6–8°), with marked broadening mainly emerging only for very long RT (4 s); and

3. whenever DRR≤0 dB, predicted unmasking is typically close to 1 dB or below. For greater DRRs, spatial unmasking is mostly driven by position, such that listeners located at the back (Position 3) rarely exceed about 1 dB of binaural advantage regardless of other parameters. The counter-trend observed for Positions 2 and 3 at low DRR / short RT) is irrelevant because: (i) those conditions would in any case be associated with degraded overall quality under the localization-based QoE perspective and (ii) low DRRs with short RTs are unlikely in a venue large enough to host immersive concerts such as those considered here.

Taken together, these points support the following conclusion: achieving DRR≥0 dB is a key prerequisite for robust immersive QoE across the audience. Notably, it is also the lever most directly amenable to improvement through system design, in that DRR depends on source directivity and aiming.

4. Discussion

The present simulations suggest that the suitability of a given venue for immersive concerts is governed first and foremost by how much direct sound survives relative to late energy, i.e., by DRR, with RT acting mainly as a secondary modulator once DRR is already unfavorable and ITDG primarily influencing image stability rather than directional accuracy. This aligns with the venue-facing question posed in the introduction: the benefits of immersive reinforcement require that the auditory system can reliably extract spatial cues across audience zones, which becomes difficult when the late field dominates.
These findings translate into a simple assessment strategy: for a given venue, prioritize DRR measurement, use RT as an indicator of how quickly performance will degrade when DRR is poor, and treat ITDG as a secondary descriptor relevant to robustness in marginal cases. The unmasking analysis further highlights that geometry sets a baseline: in the present configuration, predicted binaural advantage is strongly position-dependent even in anechoic conditions, and reverberation largely determines how much of that baseline is retained.

The scope of these conclusions is defined by the deliberate simplifications of the model. We used a diffuse late field without discrete early reflections, point sources, and a speech-derived binaural advantage metric; future work should incorporate structured early reflections, loudspeaker-like directivity (and distance-dependent DRR), and music-specific perceptual validation to sharpen assessment criteria. Despite these limitations, the study provides a compact answer to the practical question "what makes a venue good or bad for immersive live events": maintaining a sufficiently favorable DRR is a necessary precondition for robust audio-visual consistency and stable spatial impressions, with RT and ITDG playing secondary, conditional roles once DRR is controlled.

References

[1] L-Acoustics, "L-ISA Hyperreal sound: Why amplitude-based algorithms are best suited to frontal loudspeaker cconfigurations?" White paper, 2020. [Online]. Available: https://www.l-acoustics.com/wp-content/uploads/2021/07/L-ISA_Panning_Technology.pdf.

[2] A. Haapaniemi and T. Lokki, "Direct-to-reverberant ratio threshold for localization in concert halls," *Acta Acust. united Ac.*, vol. 104, no. 6, pp. 1019–1029, 2018.

[3] S. Devore, A. Ihlefeld, K. E. Hancock, B. Shinn-Cunningham, and B. Delgutte, "Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain," *Neuron*, vol. 62, no. 1, pp. 123–134, 2009.

[4] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, 1999.

[5] S. Jelfs, J. F. Culling, and M. Lavandier, "Revision and validation of a binaural model for speech intelligibility in noise.," *Hearing research*, vol. 275, no. 1-2, pp. 96–104, 2011.

[6] M. Dietz, S. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Communication*, vol. 53, no. 5, 2011.

[7] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proc. of AIA-DAGA 2013*, Mar. 2013.

[8] H. Wierstorf, A. Raake, and S. Spors, "Binaural assessment of multi-channel reproduction," in *The Technology of Binaural Listening*, J. Blauert, Ed., Springer Berlin, Heidelberg, 2013.

[9] Majdak, P., Hollomey, C., and Baumgartner, R., "Amt 1.x: A toolbox for reproducible research in auditory modeling," *Acta Acustica*, vol. 6, no. 19, 2022. DOI: 10.1051/aacus/2022011.