Audio Engineering Society

Convention Paper 10513

Presented at the 151st Convention
2021 October, Online

Perceptual Evaluation of Interior Panning Algorithms Using Static Auditory Events

Thomas Robotham[1], Andreas Silzle[1], Anamaria Nastasa[1], Alan Pawlak[1], and Jürgen Herre[1]

[1] *International Audio Laboratories Erlangen, Erlangen, Germany**

Correspondence should be addressed to Thomas Robotham (`thomas.robotham@iis-extern.fraunhofer.de`)

ABSTRACT

Interior panning algorithms enable content authors to position auditory events not only at the periphery of the loudspeaker configuration, but also within the internal space between the listeners and the loudspeakers. In this study such algorithms are rigorously evaluated, comparing rendered static auditory events at various locations against true physical loudspeaker references. Various algorithmic approaches are subjectively assessed in terms of; Overall, Timbral, and Spatial Quality for three different stimuli, at five different positions and three radii. Results show for static positions that standard Vector Base Amplitude Panning performs equal, or better, than all other interior panning algorithms tested here. Timbral Quality is maintained throughout all distances. Ratings for Spatial Quality vary, with some algorithms performing significantly worse at closer distances. Ratings for Overall Quality reduce moderately with respect to reduced reproduction radius and are predominantly influenced by Timbral Quality.

1 Introduction

Vertically extended loudspeaker configurations, in conjunction with panning techniques such as Vector Based Amplitude Panning (VBAP) [1], allow auditory events to be positioned in azimuth and elevation anywhere around the sweet spot, completely immersing the listener. The development of ISO/MPEG coding standards, such as the MPEG-H 3D audio codec [2], allows immersive content to be broadcast and played back as flexibly and faithfully as possible using the delivered positional metadata. Recently, there has been the desire to further enhance this listening experience by reproducing auditory events within the internal reproduction space. The inclusion of a distance parameter into rendering algorithms departs from auditory events being positioned only at the convex hull of the loudspeaker configuration, but also allows them to be authored closer to the listener within the reproduction space. However, to implement accurate rendering of such a perceptual cue is not without its difficulties, both in terms of development and evaluation. This study subjectively evaluates the timbral, spatial and overall quality of interior panning algorithms. Static panned auditory events are compared to a ground truth physical reference, providing novel research and insights into how far away we are in terms of quality, from the sensation of a real sound source.

*A joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits (IIS).

2 Background

2.1 Amplitude Panning

An auditory event or so-called phantom image is perceived by the listener when a mono audio signal is sent to two physical loudspeakers. By changing the relative amplitude of two adjacent loudspeaker outputs by gain factors, $g1$ and $g2$, the phantom image can be moved linearly between the positions of the loudspeakers (Figure 1a). The first panning law between two stereo loudspeakers was defined by Blumlein in 1938, as described in [3]. The reformulated version by Bauer [4] is called sine-law, from [1]. The equations are given in (1) and (2):

$$\frac{\sin\varphi}{\sin\varphi_0} = \frac{g_1 - g_2}{g_1 + g_2}, \tag{1}$$

where φ_0 (illustrated in Figure 1a) is the opening angle of the loudspeakers ($0^\circ < \varphi_0 < 90$), and φ is the angle of the virtual source ($-\varphi_0 \leq \varphi \leq \varphi_0$) and g_1, $g_2 \in [0,1]$. For the standard stereo loudspeaker setup, φ_0 is nominally 30°. The constant overall sound power is set by the constant C,

$$g_1^2 + g_2^2 = C. \tag{2}$$

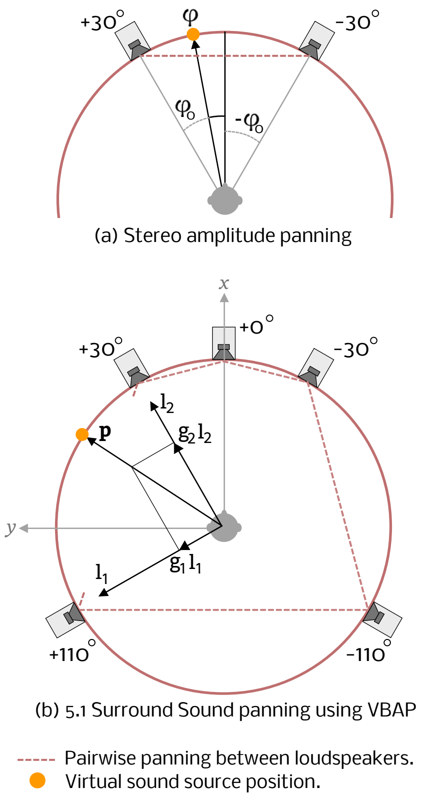
For the sine-law it is assumed that the listener is pointing to the front. For the tan-law (3) the listeners head is following the virtual source by turning around the middle axis,

$$\frac{\tan\varphi}{\tan\varphi_0} = \frac{g_1 - g_2}{g_1 + g_2}. \tag{3}$$

A generalized tangent law for horizontal pairwise amplitude panning has been defined by Zotter and Frank [5], see (4).

$$\frac{\tan\varphi}{\tan\varphi_0} = \tanh\left[\frac{ln10}{40}\gamma(L - W)\right], \tag{4}$$

where γ is the adjustable slope, L denotes the gain difference in decibels, and W a decibel shift. Different panning laws have been subjectively and objectively evaluated [6, 7, 8, 9], with varying results depending on the number of loudspeakers and panning angle.


+30°
φ
-30°
φ0
-φ0
x
+0°
+30°
-30°
l2
p
g2l2
y
l1
g1l1
+110°
-110°
---- Pairwise panning between loudspeakers.
● Virtual sound source position.

(a) Stereo amplitude panning

(b) 5.1 Surround Sound panning using VBAP

Fig. 1: Panning Pairs

2.2 Vector Base Amplitude Panning

In 1997, Pulkki reformulated the two-dimensional amplitude panning tangent law to a two- or three-dimensional Vector Base Amplitude Panning (VBAP) method [1, 10]. This allowed the auditory events to be placed anywhere around the listener with an arbitrary number of loudspeakers surrounding the listener, possibly including height or lower speakers. For two-dimensional VBAP, vector **p**, the direction of the panned sound source, can be considered a linear expression of loudspeaker unit-length vectors $\mathbf{l}_1$ and $\mathbf{l}_2$ as seen in Figure 1b,

$$\mathbf{p} = g_1\mathbf{l}_1 + g_2\mathbf{l}_2. \tag{5}$$

The respective gain factors g_1 and g_2 are described as follows, where $\mathbf{g} = [g_1\ g_2]$ and $\mathbf{L}_{12} = [l_1\ l_2]^T$,

$$\mathbf{g} = \mathbf{p}^T\mathbf{L}_{12}^{-1}. \tag{6}$$

With the inclusion of more loudspeakers, sets of pairs are selected, with each loudspeaker belonging to two pairs. In loudspeaker configurations, such as Figure 1b, the loudspeaker panning pairs would encompass the listener in a convex hull. The concept is further extended in three dimensional VBAP, where the virtual source is positioned within an active triangle of three loudspeakers. Further three-dimensional panning algorithms have since been developed and refined [11, 12, 13, 14, 15]. VBAP still remains a robust and efficient method of panning and as such, is implemented even in new standards as a method of gain calculations in simple point source panners [16]. However, while constant sound power C (Eq. 2) can be seen as a way to prompt some distance cue based on intensity, other psychoacoustic aspects should also be considered [1].

2.3 Interior Panning

Many studies have been conducted, focusing on the perception of distance of sound sources within rooms [17, 18, 19, 20, 21, 22]. Attributes such as spectral content, direct-to-reverberation ratio (DRR), near-field effects, and sound intensity all play a role in our perception of distance. Placing auditory events to be perceived at distances beyond the loudspeaker configuration can be achieved with relatively convincing psychoacoustic results by altering the perceived DRR via artificial reverberation. For 'interior panning', a more complex approach is required with various signal processing methods [23, 24] to cue additional psychoacoustic effects of distance changes. As such, numerous panning algorithms have been developed such as, Distance-Based Amplitude Panning [25], Complex Near-Field Imaging [26], Dual-Balance Panning [27, 28] and Auditory Distance Rendering [29]. These have the *additional* aim of being able to place an auditory event inside the convex hull of the loudspeaker configuration (see green shaded region in Figure 2). In a move towards next generation audio codecs, the ITU Radiocommunication sector has developed the new standard ITU-R BS.2127-0 [16] describing a renderer for advanced sound systems compliant with the Audio Definition Model [30]. However, to the authors' knowledge, no study yet has observed what levels of quality are achieved with interior panning algorithms in comparison to an actual physical audio source placed in the interior space.

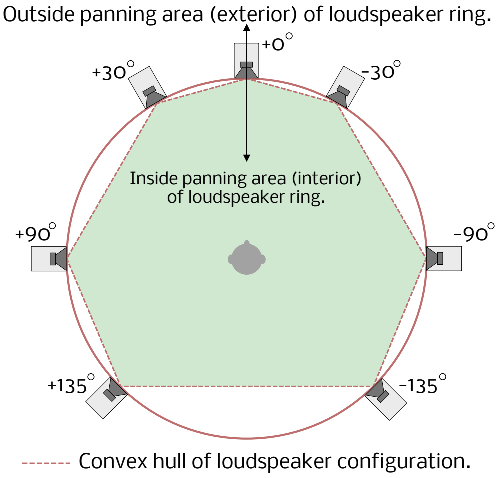

Outside panning area (exterior) of loudspeaker ring.
+0°
+30°
-30°
Inside panning area (interior) of loudspeaker ring.
+90°
-90°
+135°
-135°
Convex hull of loudspeaker configuration.

Fig. 2: Visualization of the three panning regions: (1) between the loudspeakers (red circle), (2) interior panning area (green shaded) and (3) outside panning area (outside the loudspeakers).

2.4 Evaluation of Interior Panning

Previous evaluations of panning algorithms often focus on sources positioned along the convex hull of the loudspeaker configuration [10, 31, 32, 33, 14, 34]. However, very few investigations exist on the subjective quality assessment of *interior* panning approaches.

In the context of theatre reproduction, Tsingos et al. [27] conducted a 2-interval forced choice subjective experiment. Listeners were asked to indicate which audio playback, (A) or (B), best represented a 2D illustrated trajectory of a moving sound source. The timbre of (A) and (B) were also compared to a monophonic reference and listeners were asked to select which was most timbrally faithful. The conditions under test included directional pair-wise amplitude panning, Dual-Balance panning and Distance-Based Amplitude Panning (DBAP). The results show that all algorithms perform significantly different regarding position, content type and seating position of the user. In this experiment, the use of a 2D illustrated reference requires an additional cognitive load on subjects to translate the illustrated trajectory, orientated in a particular way, and perceive it as a movement in physical space. This reference eliminates, to some degree, the need for a subjective 'internal' reference. However, additional errors may be introduced due to individual subjective interpretation of the reference and consequently, the comparison of this interpretation against the presented conditions.

Thomas and Robinson [35] conducted a study using physical loudspeakers as *anchors* in a localization task assessing dual-balance panned, auditory events. Five loudspeakers were configured at positions +90°, +45°, 0°, -45° and -90°. Loudspeakers +45° and -45° were in 'corner positions' and not on an equidistant radius, but time and amplitude aligned. Sixteen static positions were tested, all positioned on a square grid towards the front-left of the listener. An acoustically transparent curtain was used to blind the setup. However, subjects were informed that conditions would *only* come from this quadrant. To aid in localization, subjects could listen to the stimuli individually through loudspeakers +90°, 0° and -90° at any time as 'acoustic anchors'. Additionally, no restrictions were imposed on subjects to return to previously selected positions and thus allowing all positions to be compared against one another. To record the results, subjects marked the perceived (x,y) position using a graphical panner for each of the 16 items. Results indicate that positions closer to the listener yield the highest degree of error but overall, interior panned positions can be consistently perceived by the listener. In this study, the 16 positions tested were confined to a quarter of the interior space. Although this investigation assessed a high number of positions, it cannot be concluded that such results hold throughout the whole interior space. While it is plausible to assume left-to-right symmetry in localization tasks, front-to-back confusion limits these observations, where perception in the back behaves differently than the front. Furthermore, giving subjects prior knowledge that all positions were confined to a particular space potentially introduces bias. Lastly, subjects were permitted to test all 16 rendered positions with no restriction to return to previously heard conditions. This means that the results per position are not based solely on the perception of their own individual accuracy, but influenced by the perception of other rendered positions as well.

Assessing interior panning algorithms is particularly difficult due to the implementation of a reference. Real loudspeakers have been used as indicators of distances in some interior panning testing and development [29]. However, to the authors' knowledge, no experiment has investigated how far away current algorithms and techniques are from the perception of a true sound source in terms of *subjective quality*. This study focuses on providing novel data on the quality of auditory events rendered by interior panning algorithms when compared against a true physical loudspeaker reference.

3 METHOD

3.1 Evaluation Method

The evaluation methodology chosen was MUSHRA (Multi Stimulus test with Hidden Reference and Anchor) as described in ITU-R BS.1534-3 [36] for subjective assessment of intermediate audio quality. For the reference (and hidden reference) a loudspeaker was employed. This provides a true physical reference as ground truth. To measure the quality of the static auditory events, the study was broken down into three MUSHRA test sessions: Spatial Quality (S_Q), Timbral Quality (T_Q), and Overall Quality (O_Q).

For session Spatial Quality (S_Q), listeners were asked: "*Please evaluate the difference between the reference and the conditions regarding sound localization (this includes; azimuth, elevation, distance) and the extension/spread of the auditory event.*" Timbral Quality (T_Q) was assessed by asking: "*Please evaluate the difference between the reference and the conditions regarding sound coloration. This should encompass any changes of treble, mid-range or low frequency content.*" For Overall Quality (O_Q), listeners were asked: "*Please evaluate the overall quality (including localization, distance, timbre, spread/width, artifacts etc.) of each condition against the reference.*"

In addition to Overall Quality (which corresponds to the well-known 'Basic Audio Quality' attribute assessing any, and all detected differences [37], Timbral and Spatial Quality were chosen due to their known importance in subjective audio evaluations. Particularly for interior panning, these two qualities may vary in an interesting way. For example, algorithms that involve many loudspeakers for better localization, may produce a difference in timbre compared to algorithms that use fewer loudspeakers. The combinations of these specific attributes to the overall quality are of particular interest.

3.2 Sound Source Selection and Positioning

Five static positions, distributed across three general directions (*front, front-side, and rear-side*) were chosen for evaluation. Figure 3 shows the right hemisphere

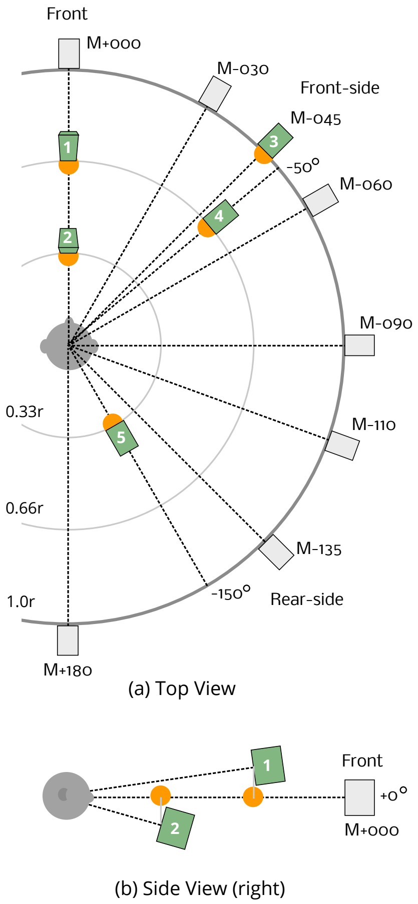
Table 1: Reference and rendered positions of auditory events corresponding to Figure 3.

Position ID	Reference Position	Rendered Position
1	(0.66r, 0°, +8°)	(0.66r, 0°, +0°)
2	(0.33r, 0°, -16°)	(0.33r, 0°, +0°)
3	(1.0r, -45°, +0°)	(1.0r, -45°, +0°)
4	(0.66r, -50°, +0°)	(0.66r, -50°, +0°)
5	(0.33r, -150°, +0°)	(0.33r, -150°, +0°)

of the setup, with positions of reference loudspeakers (green) and respective phantom images under test (orange). The positions of interior reference loudspeakers were carefully selected as to not impede the direct sound of any other loudspeakers. With the exception of two loudspeakers 1 and 2, all loudspeakers were positioned with the acoustical axis at 0° elevation. For loudspeakers 1 and 2, a vertical offset and tilt was applied while the elevation of the rendered position remained at 0° (Figure 3b). This reduces possible subjective bias introduced by all interior loudspeakers being positioned at 0° elevation. Three relative radii were chosen for evaluation: $r = 1.00$, $r = 0.66$ and $r = 0.33$, where the nominal distance 1.00 corresponds to a distance of 3 m. Position 3 is the only position placed at the boundary. This helps establish a quality 'check point' given a phantom image rendered using standard pairwise panning by all algorithms.

3.3 Reproduction Configuration and Calibration

The listening tests were conducted in an ITU-R BS.1116 [37] compliant listening room at Fraunhofer IIS [38]. Dynaudio BM6 MKII loudspeakers were used (40 Hz - 21 kHz frequency range) with a 22.2 (ITU-R BS.2051-1 System 10+9+3) loudspeaker configuration. Although reference loudspeakers were all positioned to allow for clear transmission of direct sound, additional calibration was conducted. An equalization patch developed in Max/MSP was applied to all loudspeakers to assure that interior loudspeakers do not affect the timbre of outer loudspeakers. The patch applied individual level, delay and equalization settings. Up to nine IIR peak filters with a $Q \leq 3$ and a high shelf filter were used to calibrate all loudspeakers within the tolerance curves for the operational room response (±3 dB) as defined in ITU-R BS.1116.



(a) Top View

(b) Side View (right)

Fig. 3: (a) Top view of the loudspeaker configuration. (b) Side view of Loudspeakers 1 and 2 with vertical offset and rotational tilt. In both figures, every box represents a real loudspeaker. Light green boxes indicate reference loudspeaker positions. 'M' denotes loudspeakers in the 'middle' layer of a 22.2 configuration.

The loudness calibration was done in two stages. Stage one consisted of pink noise measurements for initial estimation. Stage two included a group of expert listeners who perceptually evaluated the loudness and made fine adjustments for each individual stimuli. Stage two was iterated over separate sessions to further reduce subtle loudness differences. Finally, binaural room responses were taken using a Cortex MKI Head and Torso Simulator to check that no strong comb filtering effects were induced by reflections from internally placed loudspeakers.

Table 2: Conditions under test.

Renderer Label	Type of Renderer
REF	Ground truth loudspeaker
A	Standard VBAP algorithm
B, C, D	VBAP + extensions for radius
E	'Dual-balance' panner
Lp3.5	Anchor

3.4 Conditions

Five different rendering algorithms were selected for testing, four of which specifically designed to manipulate radius information to perform interior panning. The four interior panning algorithms were selected based on availability. The other condition, included for comparison, was a standard VBAP renderer (without processing of radius information, and using only gain cues). The low anchor employed was a 3.5 kHz low-pass filtered version of the test signal, with a minimum attenuation of 25 dB at 4 kHz and of 50 dB at 4.5 kHz. The low anchor was rendered by standard VBAP and spatially impaired by inverting the polarity of the channel with the highest signal output. The reference (and hidden reference) were the same Dynaudio BM6 MKII loudspeakers used in reproduction configuration. Table 2 provides an overview of all conditions. This study focuses on quality ratings of perceived interior panning against a true reference and not specifically which interior rendering technique performs best. Therefore, only a brief overview of condition types in the following sections are given.

3.4.1 VBAP (A)

Briefly described in Section 2.2, VBAP is an amplitude panning technique using speaker pairs or triplets to position a phantom image. This condition employs the point source panner published in ITU-R BS.2127-0 [16]. Its purpose within this study is to provide a clear comparison against algorithms that attempt to produce an interior phantom image. The distance cues are provided by simply altering the gain factor C (eq. 2.2).

3.4.2 VBAP: Extensions and Variations (B, C, D)

Three VBAP extensions were used for conditions B, C, and D. Condition B is a development version of the algorithm published in [39]. The main approach is to split transient and continuous components of the signal into two streams. The transient stream is rendered as two virtual sources with a width inversely proportional to the distance. The transient signal virtual sources are also mirrored, rendering a phantom image approaching from the opposite side to the intended position. The continuous stream is transformed into STFT domain, where frequency bins are spatially rendered with an extent double that of the transient stream. Condition C employs the same point source panning as A with modifications to source extent. As radius decreases, the angular extent of the phantom image increases via a non-linear function similar to that seen in [16] (Figure 9). Condition D builds on the edge-fade amplitude panning from [14] which utilizes polygons instead of triangles for geometry discretization. For perception of an interior position, gain weights are computed along the X (left/right) and Y (front/back) axis based on the distance of the panned source to the discretized polygons along the convex hull. These weights are then used as loudspeaker gains for polygon pairs (left/right, and front/back).

3.4.3 Dual-Balance Panning (E)

Dual-balance panning makes use of a set of up to eight speakers enclosing the desired 2D object position. The maximum eight loudspeakers are comprised of four pairs, two for left/right and two for front/rear. The gains for each balance pair are computed using sine/cosine functions [35]. The Cartesian point source panner is a 3D extension of the dual-balance panner, as described in [16].

3.5 Stimulus Selection

Three different stimuli were chosen to render as sound sources, see Table 3. A selection of noise, music and male speech material was selected to analyse any effect of content type. After pre-tests, it was decided to filter all signals with an 80 Hz 4th-order Butterworth high-pass. This prevents the boost of low frequency content while reproduced over two or more loudspeakers at the same time, found to be very noticeable when compared against the single reference loudspeaker. This high-pass also prevents excitement of particular room modes induced by internally placed reference loudspeakers. The strength of these room modes may vary depending on loudspeaker placement and can therefore be better controlled with careful stimuli equalization rather than individual loudspeaker equalization.

Table 3: Programme material rendered for the test.

Stimuli	Length (s)	Description
Pink Noise	4.0	500ms Pulses
Male speech [40]	4.75	SQAM Track
Conga drums	3.32	SQAM Track

4 EVALUATION PROCEDURE

4.1 Subjects

For the O_Q test, there was a total of 9 listeners, with an age range of 26 to 55, all of which were male. The S_Q and T_Q tests had 12 participants, with an age range of 21 to 42, two of which were female. The different number of participants across sessions was due to not all listeners being available for testing. Eight subjects conducted all three test sessions. All test sessions were conducted on different days with sufficient time apart to counter any learning effects. The participants were either expert listeners or people who have had previous experience in listening tests. Subjects did not receive specific additional training for interior panning tests.

4.2 Test Procedure

Subjects were instructed both verbally and in written form via an instruction sheet. The instruction sheet contained the question, task, and definitions of all comprising attributes for the different quality metrics. Natural head movements were allowed within a tolerance of ±30° azimuth rotation, analogous to head movements when consuming film content. No physical restriction was imposed on these tolerances, rather subjects were observed throughout the test to ensure no exaggerated movements were used. Subjects could record their MUSHRA response by using either a mouse or keyboard short-cuts. The monitor was placed in the center behind the loudspeaker ring to avoid any acoustical interference. Volume adjustments were allowed within ±3 dB at the first item, but volume was then fixed for the remainder of the test. The average test time for each of the sessions (S_Q, T_Q and O_Q) was ≈30 minutes.

5 Results

The results for all three tests are presented together in Figure 4. Plots (a), (b), and (c) categorize the same data differently using the independent variables *Distance*, *Stimuli*, and *Direction* respectively. Different renderer conditions are separated along the x-axis. For every renderer, the mean values together with their 95% confidence intervals are clustered along the y-axis in groups of three, representing Overall (O_Q), Timbral (T_Q) and Spatial quality (S_Q). For each plot, the different distances, stimuli, and directions are separated in three columns and indicated by the label above.

Table 4: Two way repeated measures ANOVA for each independent quality listening session. (Significant effects are denoted by *.)

Overall Quality (O_Q)		$p < 0.05$
*Distance**	$F_{(2,16)} = 20.8286$	$p < 0.001$
*Condition**	$F_{(6,48)} = 191.018$	$p < 0.001$
Distance × *Condition**	$F_{(12,96)} = 14.423$	$p < 0.001$
Timbral Quality (T_Q)		$p < 0.05$
Distance	$F_{(2,22)} = 2.997$	$p = 0.097$
*Condition**	$F_{(6,66)} = 298.917$	$p < 0.001$
Distance × *Condition**	$F_{(12,132)} = 2.666$	$p = 0.003$
Spatial Quality (S_Q)		$p < 0.05$
*Distance**	$F_{(2,22)} = 3.492$	$p < 0.001$
*Condition**	$F_{(6,66)} = 204.098$	$p < 0.001$
Distance × *Condition**	$F_{(12,132)} = 6.849$	$p < 0.001$

The data represented in Figure 4a show the main independent variables of interest. As such, a two-way repeated measured analysis of variance was conducted considering *Distance* and *Conditions* for each quality type (O_Q, T_Q, and S_Q). The F-statistic and degrees of freedom, and associated p-value are reported in Table 4. Results show that significant differences were present for all except one main effect, as indicated by (*). For T_Q, *Distance* alone had no effect on quality ratings. This is reflected in Figure 4a, where 95% confidence intervals display no significant difference across distances.

The effect of *Stimuli* (seen in Figure 4b) had some significant effect on O_Q and T_Q ratings. On average, pink noise was rated lower than conga and speech samples. This is likely due to the critical nature of the test signal. The same general trend for S_Q ratings can be observed across all three stimuli. Finally regarding *Direction* (Figure 4c), no significant difference was observed for T_Q. Most noticeable are the ratings for S_Q, where some renderers have a significant difference depending on the general direction.

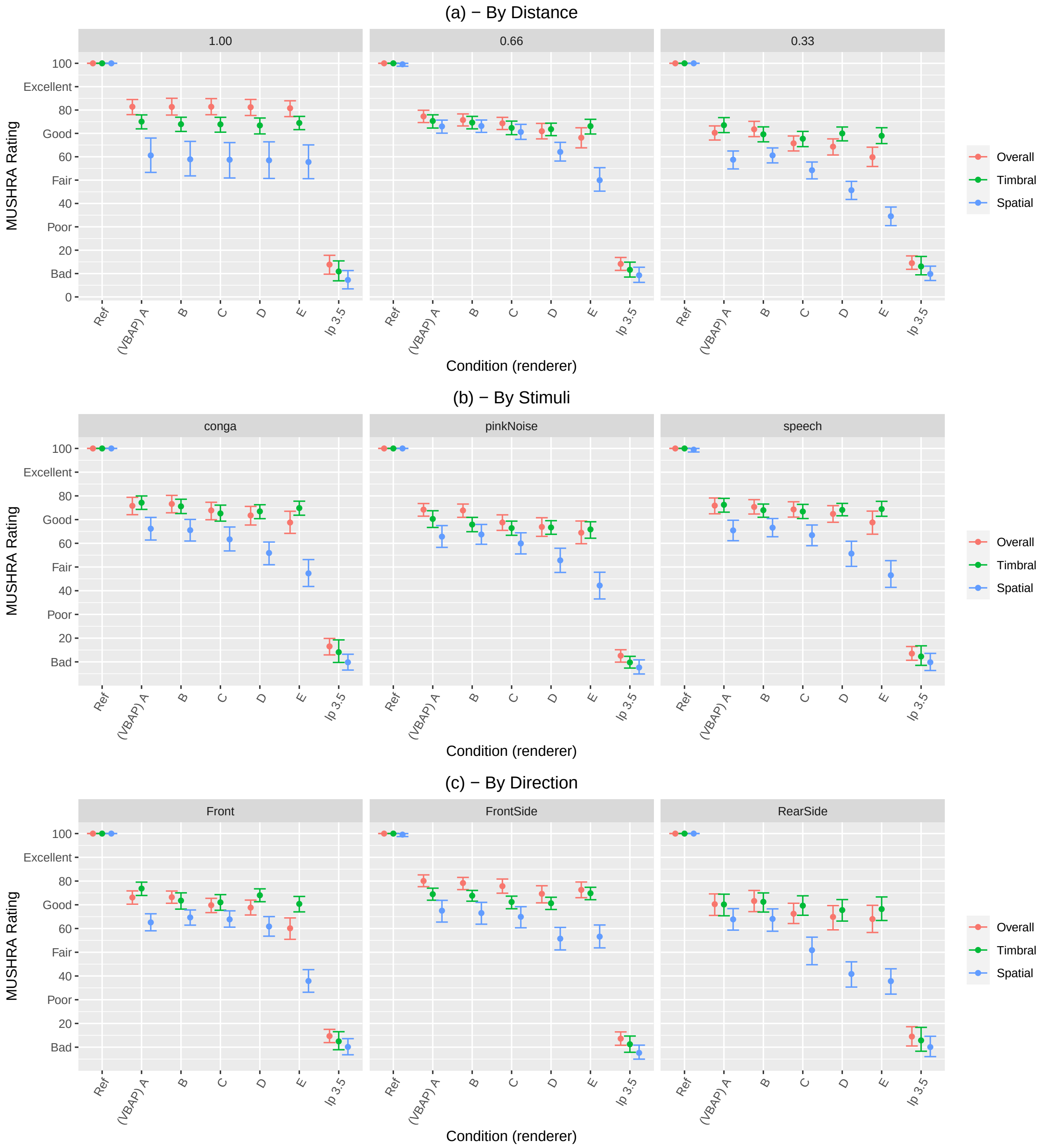


Fig. 4: Results for MUSHRA ratings of interior panning algorithms. Plots **(a)**, **(b)**, and **(c)** group the results into independent variables **(Distance)**, **(Stimuli)** and **(Direction)** respectively. Conditions under test are displayed along the x-axis, with the mean ratings and 95% confidence interval for each quality along the y-axis. (Overall Quality $N = 9$, Timbral Quality $N = 12$, and Spatial Quality $N = 12$).

6 Discussion

6.1 Quality and Distance

At the nominal distance $r = 1.0$, all renderers (A - E) were rated equal for all three quality measures (Figure 4a). This demonstrates that these conditions could perform traditional pairwise panning with no relative disadvantage. In general, it can be observed that the auditory event produced from simple pairwise panning is perceived to be clearly different to that of a true loudspeaker. This confirms former published results [41]. Furthermore, it can be reasonably inferred that O_Q is predominantly determined by T_Q, with S_Q having minimal influence on O_Q judgements.

At a reduced radius $r = 0.66$, the relationship between all quality ratings have changed. Results for O_Q and T_Q are equal for all conditions but the ratings for S_Q in conditions A - C seem to have increased compared to $r = 1.00$. One potential reason for this is the strength of directional cues given by VBAP, or certain VBAP-based renderers. For example, Position 1 has the rendered position and reference loudspeaker in exact azimuth alignment with the reproduction loudspeaker M+000. As such, the auditory event is rendered solely by loudspeaker M+000 and degradations for this position do *not* induce azimuth error. As localization cue in the vertical median plane is not as strong as azimuth, it is likely that the degradations observed here are mostly due to changes in distance. Conversely, renderers that include more loudspeakers (e.g., Conditions D and E) have a higher probability of azimuth error or localization blur. Such renderers remain statistically the same in S_Q compared to $r = 1.00$, but are relatively lower in comparison to Conditions A - C, at $r = 0.66$.

Finally, the trends observed at $r = 0.66$ are also present at the minimum radius $r = 0.33$. Both O_Q and T_Q show similar fluctuations and some minor reductions. Ratings for S_Q have significantly decreased, and differences between conditions magnified.

These results show that distance does indeed have a significant effect on S_Q for all renderers, in some cases resulting in *'Poor'* S_Q at the closest distance. However, even if no attempt to accommodate distance is made, a high S_Q can still be achieved if localization error on the azimuth plane is kept perceptually minimal, potentially by utilizing fewer loudspeakers. As such, none of the renderers that utilized radius information were able to perform better than standard VBAP and distance cues using gain (condition A) in any quality measure tested, highlighted by Figure 5. For T_Q, no significant difference was observed with reduction of distance and remained *'Good'* throughout. Whilst O_Q did degrade with respect to distance, ratings were still *'Good'* at reduced distances. This implies that as long as renderers can maintain high T_Q in comparison to an audible reference, O_Q may not be as drastically affected.

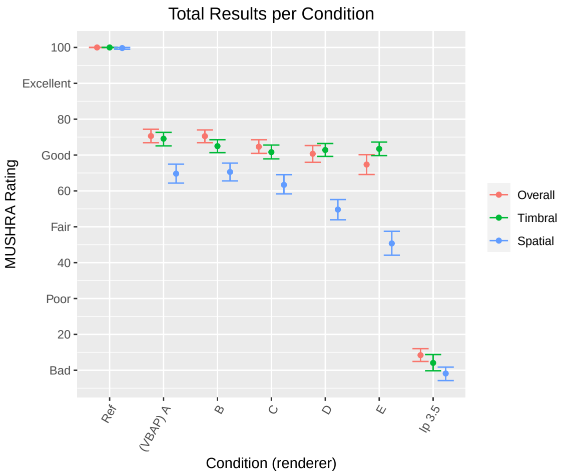


Fig. 5: Mean and 95% confidence interval for each condition and test session, for all data.

6.2 Static Ground Truth Reference

This study focused on the use of a true physical loudspeaker reference, carefully calibrated for comparison against static panned auditory events. While this eliminates the need for any subjective internal reference, the scope of this reference should be recognized.

Firstly, it is understood that interior panning is not limited to static auditory events. Common cinematic effects (e.g., plane/helicopter fly-overs) frequently feature dynamic objects that move in position over time. It is reasonable to suggest that this movement and variance of an effect over time, introduces further psychoacoustic cues to give the perception of a smooth transition, closer and further away from listeners. However, perception of such an effect is in essence, based on renderings of static auditory events. Therefore, the findings in this study focus specifically on fixed interior panning psychoacoustic capabilities. Here, only the static cues are available to the human auditory system for quality assessment and any additional perceptual tolerance from dynamically panned sources is detached from the given rating. Furthermore, the advent of services such as Atmos Music and Sony 360 also open up immersive content to the music industry. In this domain, static objects are much more prevalent and effects such as dynamic instruments are seldom used by content creators. [42].

Secondly, the role of direct-to-reverberation energy ratio (DRR) within the room. Auditory distance perception is largely a property of the intensity cues and reflections arriving at the listener [22]. In this study, it is unavoidable that loudspeakers closer to the receiver (listener) will produce a higher DRR than those further away. The repercussions of this can be highlighted by Condition A (VBAP) at $r = 0.66$ and $r = 0.33$ in Figure 4a. For Condition A, the reference speakers are compared to the auditory event produced strictly by the M+000 loudspeaker. As the conditions are all loudness aligned, and thus no significant intensity differences are present, it is reasonable to suggest that DRR is a large contributor to this S_Q degradation. Additional factors such as incorrect source extent may also contribute to a degraded rating. However, extended VBAP algorithms that apply source extent processing were also rated either equal or lower than Condition A. This would suggest that even if additional processing is applied to prompt more distance cues, it would still not compensate for the degradation due to low DRR for static auditory events.

7 Conclusion

This study provides results of a novel experiment using a MUSHRA comparison methodology of static audio events rendered by interior panning algorithms, against a true physical loudspeaker reference. Great care was taken to eliminate any unwanted contributing factors, and yield results truly representative of interior panning algorithms. Four static interior positions were tested, and one position at the nominal distance rendered using pairwise panning. The results show that for static interior panning;

- All renderers specifically optimized or designed to manipulate radius information, performed equal or worse than VBAP and gain changes when compared to a physical loudspeaker reference.
- A reduction in distance had no significant effect on Timbral quality for all conditions.
- A reduction in distance had a significant effect on Spatial quality, with conditions ranging from *'Good'* to *'Poor'* at a reduced distance of $r = 0.33$.
- Overall quality is predominantly influenced by Timbral Quality than Spatial Quality.

References

[1] Pulkki, V., "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, 45(6), pp. 456–466, 1997.

[2] Herre, J., Hilpert, J., Kuntz, A., and Plogsties, J., "MPEG-H Audio - The New Standard for Universal Spatial/3D Audio Coding," *J. Audio Eng. Soc.*, 62(12), pp. 821–830, 2014.

[3] Blumlein, A. D., "British Patent Specification 394,325," *J. Audio Eng. Soc.*, 6(2), pp. 91–98, 1958.

[4] Bauer, B. B., "Phasor Analysis of Some Stereophonic Phenomena," *J. Acoust. Soc. Am.*, 33, pp. 1536–1539, 1961.

[5] Zotter, F. and Frank, M., "Generalized Tangent Law for Horizontal Pairwise Amplitude Panning," in *International Conference on Spatial Audio*, pp. 1–7, Graz, Austria, 2015.

[6] Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, London, England, 2nd edition, 1996, ISBN 9780262024136.

[7] Gerzon, M. A., "Panpot Laws for Multispeaker Stereo," in *92nd AES Convention*, pp. 1–32, Vienna, Austria, 1992.

[8] Neoran, I. M., "Surround Sound Mixing Using Rotation, Stereo Width, and Distance Pan-Pots," in *109th AES Convention*, pp. 1–14, Los Angeles, CA, USA, 2000, ISBN 9723510766.

[9] Griesinger, D., "Stereo and Surround Panning in Practice," in *112th AES Convention*, pp. 1–6, Munich, Germany, 2002.

[10] Pulkki, V. and Karjalainen, M., "Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning," *J. Audio Eng. Soc.*, 49(9), pp. 739–752, 2001, ISSN 00047554.

[11] Dickins, G., Flax, M., McKeag, A., and McGrath, D., "Optimal 3D Speaker Panning," in *AES 16th International Conference*, pp. 421–426, Rovaniemi, Finland, 1999.

[12] Ando, A. and Hamasaki, K., "Sound Intensity Based Three-Dimensional Panning," in *126th AES Convention*, pp. 1–9, Munich, Germany, 2009.

[13] Pulkki, V., "Uniform Spreading of Amplitude Panned Virtual Sources," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct, pp. 1–4, New Paltz, New York, NY, USA, 1999, ISBN 9512283417.

[14] Borß, C., "A Polygon-Based Panning Method for 3D Loudspeaker Setups," in *137th AES Convention*, pp. 1–10, Los Angeles, CA, USA, 2014.

[15] Lee, H., Johnson, D., and Mironovs, M., "Virtual Hemispherical Amplitude Panning (VHAP): A Method for 3D Panning without Elevated Loudspeakers," in *144th AES Convention*, pp. 1–6, Milan, Italy, 2018.

[16] International Telecommunication Union Radiocommunication Sector, *ITU-R BS.2127-0: Audio Definition Model Renderer for Advanced Sound Systems*, volume BS Series, Geneva, Switzerland, 2019.

[17] Little, A. D., Mershon, D. H., and Cox, P. H., "Spectral Content as a Cue to Perceived Auditory Distance." *Perception*, 21(3), pp. 405–416, 1992.

[18] Bronkhorst, A. W. and Houtgast, T., "Auditory Distance Perception in Rooms," *Letters To Nature*, 397(6719), pp. 517–520, 1999.

[19] Larsen, E., Iyer, N., Lansing, C. R., and Feng, A. S., "On the Minimum Audible Difference in Direct-to-Reverberant Energy Ratio," *J. Acoust. Soc. Am.*, 124(1), pp. 450–461, 2008.

[20] Zahorik, P., "Direct-to-Reverberant Energy Ratio Sensitivity," *J. Acoust. Soc. Am.*, 112(5), pp. 2110–2117, 2002.

[21] Kearney, G., Liu, X., Manns, A., and Gorzel, M., "Auditory Distance Perception with Static and Dynamic Binaural Rendering," in *AES 57th International Conference*, pp. 1–8, Hollywood, CA, USA, 2015.

[22] Zahorik, P., Brungart, D. S., and Bronkhorst, A. W., "Auditory Distance Perception in Humans: A Summary of Past and Present Research," *Acta Acustica united with Acustica*, 91(3), pp. 409–420, 2005.

[23] Gerzon, M. A., "The Design of Distance Panpots," in *92nd AES Convention*, pp. 1–33, Vienna, Austria, 1992.

[24] Füg, S., Plogsties, J., Fleischmann, F., and Laitinen, M.-v., "Evaluation eines Algorithmus zum Rendern von Distanz und Nähe bei binauraler Kopfhörerwiedergabe. [Evaluation of an algorithm to render distance and proximity for binaural headphone reproduction]," in *Proc. DAGA*, pp. 1–4, Aachen, Germany, 2016.

[25] Lossius, T. and Baltazar, P., "DBAP - Distance-Based Amplitude Panning," in *Proceedings of the 2009 International Computer Music Conference*, June, pp. 1–4, Monteal, Canada, 2009, ISBN 9780971319271.

[26] Menzies, D. and Fazi, F. M., "A Complex Panning Method for Near-Field Imaging," *IEEE Transactions on Audio, Speech, and Language Processing*, 26(9), pp. 1539–1548, 2018.

[27] Tsingos, N., Robinson, C. Q., Darcy, D. P., and Crum, P. A. C., "Evaluation of Panning Algorithms for Theatrical Applications," in *2nd International Conference on Spatial Audio*, pp. 1–8, Erlangen, Germany, 2014.

[28] Mikami, T., Nakahara, M., Someya, K., and Omoto, A., "3D Sound Intensity Measurement of 1241 Sound Objects on Fine Panning Grids by Using a Virtual Source Visualizer," in *AES 144th Convention*, pp. 1–5, Milan, Italy, 2018.

[29] Laitinen, M.-V., Walther, A., Plogsties, J., and Pulkki, V., "Auditory Distance Rendering Using a Standard 5.1 Loudspeaker Layout," in *139th AES Convention*, pp. 1–7, New York, NY, USA, 2018.

[30] International Telecommunication Union Radiocommunication Sector, *ITU-R BS.2076-2: Audio Definition Model*, volume BS Series, Geneva, Switzerland, 2019.

[31] Pulkki, V., "Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning," *J. Audio Eng. Soc.*, 49(9), pp. 753–767, 2001.

[32] Pulkki, V., "Coloration of Amplitude-Panned Virtual Sources," in *110th AES Convention*, pp. 1–12, Amsterdam, The Netherlands, 2001.

[33] Gretzki, R. and Silzle, A., "A New Method for Elevation Panning Reducing the Size of the Resulting Auditory Events," in *Proc. DAGA*, 1, pp. 443–444, Strasbourg, 2004.

[34] Marentakis, G., Zotter, F., and Frank, M., "Vector-Base and Ambisonic Amplitude Panning: A Comparison Using Pop, Classical, and Contemporary Spatial Music," *Acta Acustica united with Acustica*, 100, pp. 945–955, 2014.

[35] Thomas, M. R. P. and Robinson, C. Q., "Amplitude Panning and the Interior Pan," in *143rd AES Convention*, pp. 1–8, New York, NY, USA, 2017.

[36] International Telecommunication Union Radiocommunication Sector, *ITU-R BS.1534-3: Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems*, volume BS Series, Geneva, Switzerland, 2015.

[37] International Telecommunication Union Radiocommunication Sector, *ITU-R BS.1116-3: Methods for the Subjective Assessment of Small Impairments in Audio Systems*, volume BS Series, Geneva, Switzerland, 2015.

[38] Silzle, A., Geyersberger, S., Brohasga, G., Weninger, D., and Leistner, M., "Vision and Technique Behind the New Studios and Listening Rooms of the Fraunhofer IIS Audio Laboratory," in *126th AES Convention*, pp. 1–15, Munich, Germany, 2009.

[39] Juhani, P. and Pulkki, V., "Short-range rendering of virtual sources for multichannel loudspeaker setups," in *149th AES Convention*, pp. 1–10, Online Convention, 2020.

[40] European Broadcasting Union, *3253-E, Sound Quality Assessment Material Recordings for Subjective Tests (SQAM)*, Brussels, Belgium, 1988.

[41] Silzle, A. and Theile, G., "HDTV-Mehrkanalton: Untersuchungen zur Abbildungsqualitaet beim Einsatz zusaetzlicher Mittenlautsprecher. [HDTV Multichannel Audio: Investigation About the Perceptual Quality Using an Additional Center Loudspeaker.]," in *16th Tonmeistertagung*, Karlsruhe, Germany, 1990.

[42] Kenny, T., "Three Years, Three Players: Dolby, UMG/Capitol Studios, PMC and the Launch of Atmos Music," *Mix*, pp. 22–25, 2020, doi:dolbylabs.co/Mix-Mag.