SPATIAL RENDERING OVER DISTRIBUTED FILL SYSTEMS IN IMMERSIVE LIVE SOUND REPRODUCTION

E Corteel L-Acoustics, Marcoussis, France
F Roskam L-Acoustics UK, London, United Kingdom
S Moulin L-Acoustics, Marcoussis, France

1 INTRODUCTION

Immersive live sound reproduction is an emerging field that aims at providing high quality spatial sound for large audiences in live shows. This entails three major shifts from common practices:

1. transition from channel-based to object-based audio description,
2. use of dedicated immersive audio algorithm to create loudspeaker signals based on the audio objects input channels, the audio objects metadata and the loudspeaker positioning information,
3. adapted loudspeaker system design methodologies.

The principal objective of the loudspeaker system design process is that the loudspeaker system covers the entire audience. The coverage is the area over which the loudspeaker system provides a direct sound in an acceptable frequency response variation. In an immersive system, this should be achieved for any possible audio object location, independently of the number of loudspeakers being used.

Like classical approaches, the loudspeaker system can be broken into several components:

- a principal system that delivers most of the coverage, most of the frequency bandwidth, and most of the spatial rendering capabilities,
- a subwoofer system that extends the low frequency capabilities of the principal system in the same coverage area,
- additional fill systems that aim primarily at complementing the coverage of the principal system in specific audience areas.

Fill systems are classically fed with mono signals, consisting of the sum of all audio object channels, therefore providing limited spatial information for audience members located within their coverage area.

This article presents options to restore spatial information within the corresponding area using dedicated loudspeaker system design approaches and immersive audio rendering techniques.

First, loudspeaker system design for immersive systems is presented with a specific focus on fill systems that use distributed loudspeakers over a wide audience area (front-fills, under-balcony fills). Second, various spatial rendering algorithms are introduced. Third, an evaluation framework to compare multiple approaches along key quality criteria is presented. Fourth, results are presented and finally discussed, presenting additional benefits of the proposed “spatial fills” approach in terms of time alignment.

2 LOUDSPEAKER SYSTEM DESIGN FOR IMMERSIVE LIVE SOUND REPRODUCTION

2.1 Frontal loudspeaker system design

The minimum loudspeaker system for spatial sound reinforcement typically consists in 5 full-range sources located above the stage and spanning its full width. Full-range sources can be individual loudspeakers or line sources, depending on the depth of the audience area, required Sound Pressure Level (SPL) average and distribution, as well as frequency response.

In the L-Acoustics approach for immersive sound reproduction, all full-range sources of the so-called **scene** system should be configured to maximize their shared coverage area[1], as can be seen in Figure 1. This shared coverage area defines the area of the audience where audio objects can be precisely localized (spatialized zone). Like so, coverage is assured in a well-defined audience area independently of the position of the audio object.

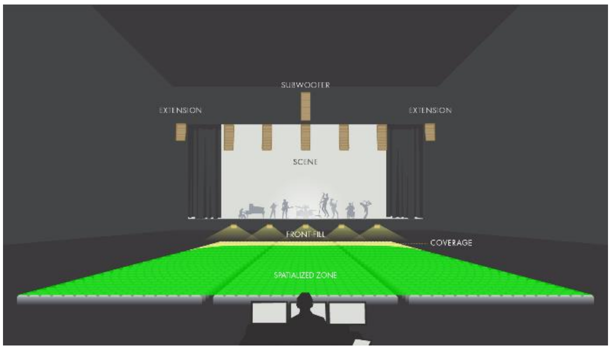

Figure 1: frontal system layout, spatialized zone (cross coverage area of all full range sources of the scene system), and coverage of a typical front-fill system

An optional **extension** system can be used to extend the perceived total width of the sound scene (panorama). It consists of one or multiple full-range sources located on either side of the scene system, configured so that their coverage maximally overlaps with the spatialized zone.

The scene and extension systems form an essential component of the **principal** loudspeaker system which receives the main contributions in a live frontal act. This is where the main performers are located and where most of the acoustical energy needs to be created. All loudspeakers of the principal system should be fed with discrete signals from the spatial sound processing unit to provide the intended spatial information to the audience.

2.2 Fills systems

The **fill** systems are dedicated to complement the principal system and are designated according to the portion of the audience they need to cover and where they are implemented. Different types of fill systems are described in Table 1.

Fill system name	Audience area	Loudspeaker positioning
Front-fills, near-fills, lip-fills	directly in front of the stage	along the stage
In-fills	directly in front of the stage	on either side of stage, pointing inwards
Out-fills	on either side of the stage	on either side of stage, pointing outwards
Under-balcony fills	below a balcony where the frontal system is physically masked	along the edge of the balcony
Delay-fills	where the principal system cannot provide enough intelligibility or SPL	large distance from the stage, separate sources

Table 1: fill system types, audience area and loudspeaker positions

Full-range sources that comprise In-fills, out-fills or delay-fills systems are covering audience area that have little or no overlap. There is therefore little to no potential to offer spatial rendering, which require combining multiple full-range sources.

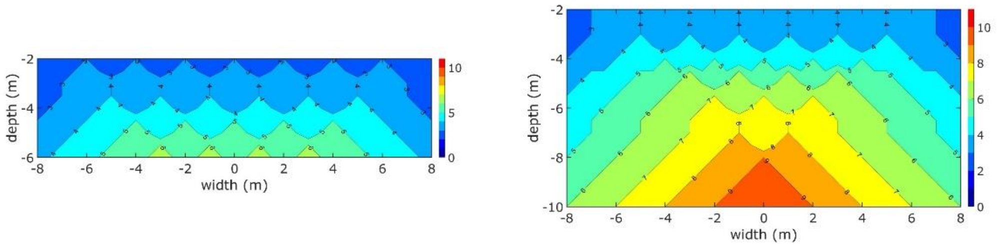

Figure 2: number of loudspeakers within 6 dB level difference at each audience position, front-fills (left) and under-balcony fills (right), loudspeakers located at 0m depth, details can be found in section 4.1.

On the contrary, distributed fills systems, such as front-fills or under-balcony fills, comprise a distribution of full-range sources over a relatively wide area. In contrast to the principal system, all loudspeakers are necessary to create coverage in the corresponding audience area. As can be seen in Figure 2, some audience members can be in the coverage of only some loudspeakers, but most are in the coverage of, at least, half of the deployed full-range sources.

3 SPATIAL RENDERING FOR DISTRIBUTED FILLS SYSTEMS

3.1 Gain-based: Vector Based Amplitude Panning

Vector Based Amplitude Panning (VBAP) is an extension of amplitude based stereo panning to an arbitrary loudspeaker layout[2, 3]. The algorithm typically selects loudspeakers that are closest to the audio object direction. It creates gains depending on the angular difference between the selected loudspeakers and the audio object directions. The smaller the angular difference, the higher the gain. An object exactly in the direction of one loudspeaker is played on this single loudspeaker.

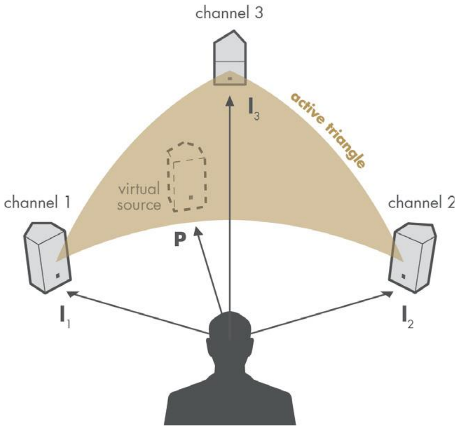

Figure 3: basic principle of VBAP

3.2 Delay-based: Wave Field Synthesis

Wave Field Synthesis (WFS)[4] is a delay-based panning technique that theoretically reproduces the physical sound field emitted by an audio object within the entire audience (see Figure 4). This however requires an infinite number of loudspeakers that can be individually controlled and amplified.

Reducing the number of loudspeakers restricts the positioning of audio objects, most often to two dimensions, and limits the accuracy of the reproduction to low frequencies only. Large loudspeaker spacing typically used for frontal systems restrict the physically accurate reproduction below 100 Hz (loudspeaker spacing >4 m). Above this so-called aliasing frequency[5], there is no physical wave

front reconstruction, and the localization is driven by the first loudspeaker contribution that reaches the listener (precedence-based localization[6]).
With WFS, the number of active full-range sources depends on the distance of the object to the declared full-range sources positions. For front-fills, objects located downstage may only activate a small number of full-range sources, which may cause coverage issues for some part of the target audience area.

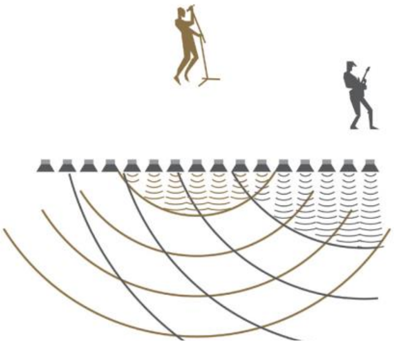

Figure 4: basic principles of Wave Field Synthesis

3.3 L-ISA spatial fills algorithm

A new spatial-fills solution is offered by the L-ISA technology by L-Acoustics, creating spatial sound in areas covered by distributed fill systems. This option requires that any listener in corresponding audience area is in the coverage area of at least three loudspeakers (front-fills and under-balcony fills). Based on this requirement, a first step consists in creating virtual loudspeakers to restore cross-coverage. The virtual loudspeakers (referred to as spatial fills in Figure 5) are then fed with the same signals as the scene system.

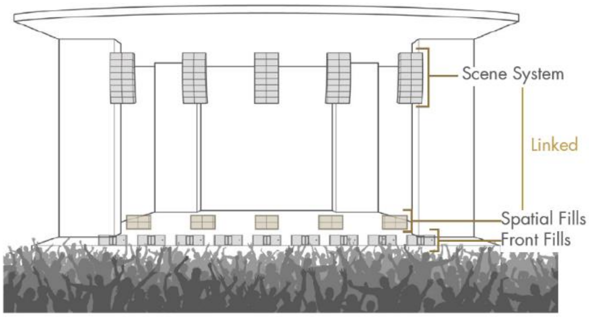

Figure 5: L-ISA spatial fills principle

The goal of spatial-fills is to improve coverage of the Scene system full-range sources. This is done by combining the distributed system sources and creating a virtual replica of the Scene system. Doing so, the coverage of the spatial-fills can expand to the coverage area of the entire distributed fill system. This is shown in Figure 6 comparing the Sound Pressure Level produced in an area close to the stage in the case of front-fills. The single house left front-fill loudspeaker creates level differences in the audience of up to 30 dB (top figure) whereas the virtual replica of the house left Scene loudspeaker reduces that difference to 8 dB for positions closest to the stage (depth = -2 m).

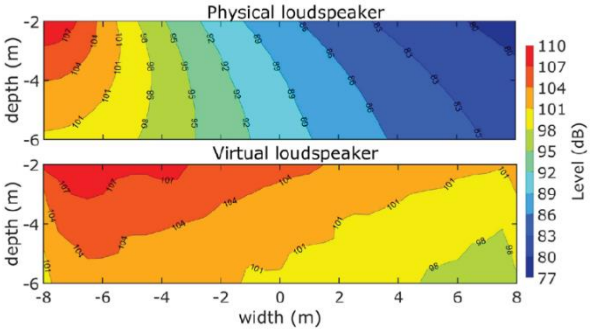

Figure 6: Simulated SPL over the coverage area of a front-fill system, based on L-Acoustics Soundvision propagation model, comparing a single physical loudspeaker (top) and a virtual loudspeaker created with the L-ISA spatial-fills algorithm (bottom)

The virtual loudspeakers are created using a delay-based algorithm with an optimized gain distribution. Two parameters are available to the end-user:

- Virtual distance (in m): distance of the virtual loudspeakers from the physical fill system,
- Gain gradient (in dB): difference between the highest and the lowest gain among all created virtual loudspeakers.

4 EVALUATION FRAMEWORK

The aim of this section is to define a test configuration and performance metrics for the evaluation of the three algorithms described in section 3 and the baseline mono solution.

4.1 Test configuration

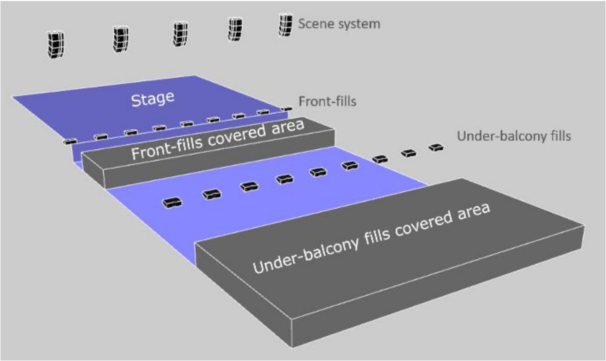

Figure 7: test configuration for evaluation (tested audience areas, loudspeaker positioning and type)

The test scenario considers a typical 16 m wide shoe-box venue with a 16*16 m stage. The scene system comprises 5 full-range sources spanning the width of the stage. It is complemented by a front-fills system consisting of 9 regularly spaced (2m) L-Acoustics Kara II loudspeakers and a similar under-balcony system located 16 m away from the stage, each covering their own specific area.

The tested audio object positions create a grid over the width and depth of the stage:

- Width: 0 (centre), 2, -4, 6, -8 m,
- Depth (upstage): 2.5, 4, 5.5, 8, 12, 16 m.

The following tuning values are used for the L-ISA spatial-fills solution:

- Front-fills: 5 m virtual distance (approx. 1/3rd of the stage width) and 8 dB gain gradient,
- Under-balcony fills: 16 m virtual distance (distance from scene to under-balcony fills) and 4 dB gain gradient.

More configurations were tested in terms of dimensions, number of loudspeakers, virtual distance and gain gradient. They are not represented here since they lead to similar trends and results under the defined conditions (each audience position should be in the coverage at least three full range sources).

4.2 Level homogeneity estimation

The level homogeneity corresponds to the ability of the sound system to limit the SPL variation at a listening position depending on the audio object position. Level homogeneity can be derived from the SPL estimated in the 1 to 10 kHz bandwidth. This bandwidth is classically used to evaluate coverage and associated intelligibility criteria[7].

For each audience position and spatial rendering solution, the level homogeneity is calculated as the difference between the highest and the lowest SPL among all rendered audio object positions. The smaller the better to maintain consistency of the mix within the covered area.

Level homogeneity estimation is shown in Figure 8 in the case of front-fills. High level differences can be observed in the case of the delay-based algorithm among the ensemble of tested audio-object positions (up to 24 dB). In delay-based algorithms, downstage object positions create low gain values for loudspeakers located at opposite side of the stage. This creates low rendered SPL due to the limited coverage of individual loudspeakers.

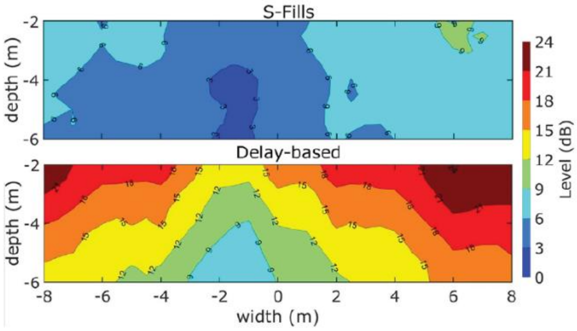

Figure 8: level homogeneity estimation for front-fills for two spatial rendering algorithms: L-ISA spatial-fills (top) and delay-based (bottom)

4.3 Spatial rendering quality estimation

Two essential qualities of spatial reproduction are evaluated:

- Localization
- Spatial unmasking

Localization determines the ability of listeners to localize the audio objects according to their intended position on stage. It provides an indication of the ability of the sound system to create consistent audio and visual information that fuse naturally into a unique perceptual object. In the horizontal dimension, an error of 7.5° between target and actual auditory positioning is considered as acceptable to guarantee audio-visual consistency.

Spatial unmasking evaluates the auditory benefits in terms of release from masking due to the spatial separation of objects. This allows listeners to focus more easily on one of the audio objects in a complex mix. Spatial unmasking also improves the intelligibility of the main performer while limiting the requirement of equalization and compression on the remaining components of the mix.

Auditory models have proven their accuracy for localization estimate of loudspeaker based spatialization systems and are therefore used here instead of conducting perceptual experiments[8]. Auditory models are fed with binaural signals corresponding to sound waves arriving at both ears of a listener in a concert situation. Head Related Transfer Functions[9] measured in an anechoic environment are used to simulate a free field situation, not considering the reverberation of the environment but concentrating on the direct sound only.

Two models of the Auditory Modeling Toolbox[10] are used:

- wierstorf2013_estimateazimuth for localization estimation, providing an estimate of the horizontal localization and the associated standard deviation[8].
- jelfs2011 for spatial unmasking, estimating the increase in speech intelligibility of the target when the target and interferer are spatially separated[11, 12]. The test situation here corresponds to a lead singer in the centre and an harmonic instrument located house-right having a similar frequency range as speech.

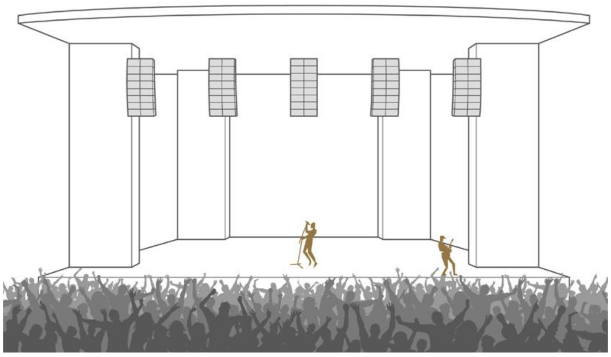

Figure 9: target and interferer locations for spatial unmasking estimation

When assessing localization precision, participants are asked to perform a localization task, indicating where they perceive a given sound stimulus. The results of localization tests are typically analysed according to two dimensions, both being available through the auditory model:

- Accuracy: average response among participants,
- Blur: uncertainty in localization, corresponding to the spread of participants answers.

The localization error is calculated as the absolute difference between simulated participants responses and the target localization for each test listening position. The spread of participants responses is simulated so that the standard deviation of all responses corresponds to the estimated localization blur.

4.4 Quality scores

A quality score system is proposed for each criterion to facilitate the comparison between different solutions on criteria that do not share the same scale and units.

- **Level homogeneity** is calculated as the 95th percentile (in dB) of SPL differential among all audience positions (worst-case scenario). The worst-case scenario must be accounted for as there must not be positions where objects cannot be heard.
- **Spatial unmasking** is the median value (in dB) among all object pairs (target centre, interferer at same depth but different left-right location) and all audience positions.
- **Audio-visual consistency**, related to localization error is the median value (in °) among all object positions (width and depth on stage) and all audience positions.

Quality score	4*	3*	2*	1*
Level homogeneity (in dB)	<3	>6	>9	>12
Spatial unmasking (in dB)	>4	>3	>2	>1
Localization error (in °)	<5	<10	<15	<20

Table 2: correspondence between quality scores and evaluation metrics

The quality score is obtained using the thresholds presented in Table 2. The higher the quality score, the better.

5 EVALUATION RESULTS

5.1 Front-fills

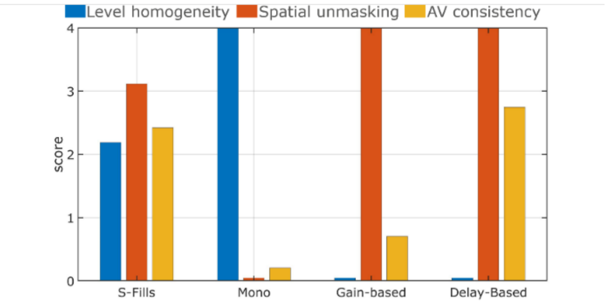

Figure 10: quality score for front-fills

This comparison shows that only the spatial-fills solution of L-ISA (S-Fills) provides a good balance between the three criteria (level homogeneity, spatial unmasking, audio-visual consistency).
The Mono solution provides the best level homogeneity but no spatial unmasking and limited audio-visual consistency. The Mono solution is a mono downmix of all objects sent to each physical loudspeaker of the distributed fills system.
The gain-based solution is improving spatial unmasking at the expense of level homogeneity and audio-visual consistency. This is due to the lack of shared coverage of the physical full-range sources used as front-fill loudspeakers.
The delay-based solution (WFS) performs best for spatial unmasking and audio-visual consistency but fails at providing good level homogeneity. Indeed, for downstage object positions, the delay-based solution tends to concentrate all energy on a small number of loudspeakers which end up into a level homogeneity issue.

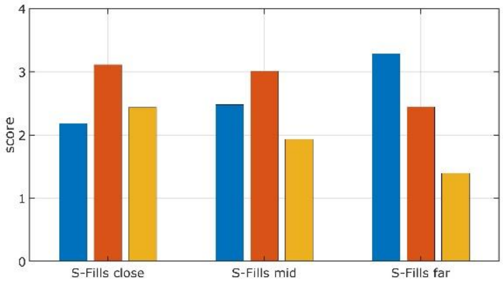

Figure 11: L-ISA spatial-fills quality scores for different virtual distance and gain gradient combinations (front-fills scenario)

Figure 11 presents the influence of the spatial-fills solution parameters on the three criteria. The S-Fills close settings (5 m virtual distance and 8 dB gain gradient) provide the best balance between coverage, spatial unmasking, and audio-visual consistency. As virtual distance increases (8 m for S-Fills mid and 16 m for S-Fills far) and gain gradient decreases (7 dB for S-Fills mid and 5 dB for

S-Fills far) coverage improves at the expense of, first, audio-visual consistency and then, spatial unmasking.

5.2 Under-balcony fills

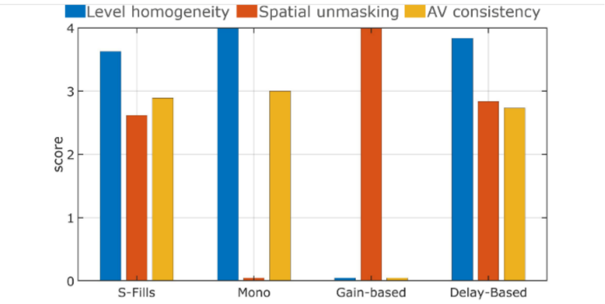

Figure 12: quality scores for under-balcony fills

The same analysis is conducted for an under-balcony fills loudspeaker system (refer to section 5 for information on test scenario). As shown on Figure 12, spatial-fills and delay-based outperform the mono and gain-based solutions. Mono ensures good level homogeneity and audio-visual consistency for this covered area (16 m away from stage and 8 m deep area). Gain-based provides good spatial unmasking but fails at level homogeneity and audio-visual consistency.

6 DISCUSSION

6.1 AUDIO-VISUAL QUALITY TRADE-OFF

This section provides a more focused discussion on localization error and the associated trade-offs in terms of audio-visual quality for a front-fills system. We considered audio objects are on the outer house left positions on stage which produce the largest localization errors. Figure 13 compares the distribution of localization errors (diamond: median, vertical bar: 25th and 75th percentile) among spatial rendering solutions.

The mono and gain-based solutions exhibits very large localization errors respectively at small and large audio object depth on stage. Both spatial-fills and delay-based solutions have similar performances, which are mostly independent from object depth on stage. Their median localization error is above the target threshold of 7.5 degrees but is better than the baseline solution (mono).

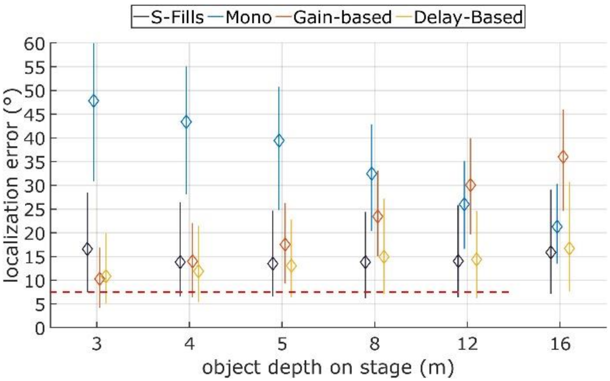

Figure 13: localization error distribution (median: diamond, 25th and 75th percentiles: vertical bar) for front-fills, depending on object depth on stage and spatial rendering solutions for audio objects in the outer house left position

Figure 14 compares multiple values of the virtual distance and gain gradient parameters of the L-ISA spatial-fills solution (see section 5.1 for details). The close setting provides the best overall results, with little dependency against the object depth on stage.

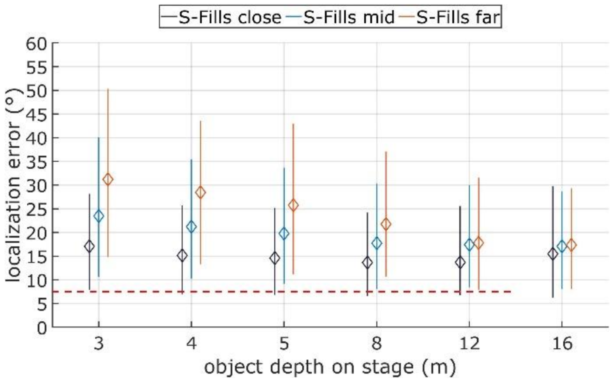

Figure 14: localization error distribution for front-fills, L-ISA spatial fills solution with various settings of virtual distance and gain gradient

6.2 Time alignment benefits

Fills systems need to be time aligned in their shared coverage area with full-range sources of the principal system. Time alignment guarantees that no serious artefacts (echoes) may be perceived by audience members. For immersive live sound systems, the challenge is to time align the fills systems with each full-range source of the scene system.

In Figure 15, the mono and the spatial fills solutions are compared for time alignment of front-fills with the scene system full-range sources.
Full-range sources are considered time aligned if all full-ranges sources that are within ±6 dB (pink noise, unweighted) of a reference source, are arriving not more than 10 ms after the reference source (light yellow colour in Figure 15). The reference source is the loudest source at the considered audience position. Any source that is within 6 dB in level of the reference source but arriving more than 10 ms after the reference source may cause an echo (red in Figure 15). If all sources are more than 6 dB below the level to the reference source, the reference source can be considered in isolation (dark blue in Figure 15).

In the time alignment process, the fills system is often aligned against of the centre full-range source of the fills system. To do so, a delay is applied either to the scene system or the fills system entirely.
In the mono solution (left side of Figure 15), the delay is applied to the front-fills system. As can be seen in the top left corner of Figure 15, the obtained alignment with the centre full-range source of the scene system is near optimum. However, this alignment choice leaves a large portion of the audience area unaligned with respect to other components of the scene system such as the house left full-range source (bottom left corner of Figure 15).
With spatial-fills however, the obtained alignment is optimum against both the centre and the house left full-range sources of the scene (right side of Figure 15), effectively with all full-range sources of the scene system.
It should be noted as well that time alignment of the fills system would be guaranteed against all full-range sources of the scene system even if the loudspeakers of the fills system are not distributed along a straight line. One only needs to align the centre full-range source of the scene system with the fills system for which the spatial-fills algorithm is used.

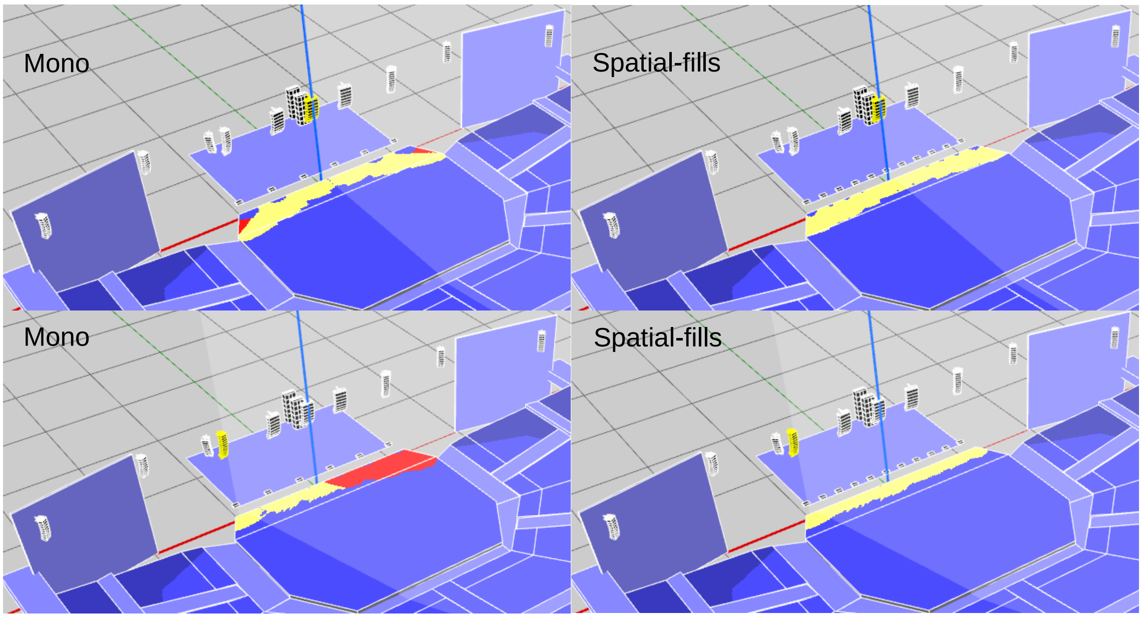

Figure 15: time alignment quality mapping of center (top) and house left (bottom) scene full-range sources (highlighted in dark yellow) with front-fills with the mono (left) and spatial-fills (right) solutions. Dark blue color represents the coverage area of the selected scene source, light yellow indicates that all active sources are in time, red indicates that some active sources are out of time while having a significant level and may cause an echo. L-ISA reference loudspeaker system design, Zenith de Paris, France.

7 CONCLUSION

Fills systems are indispensable components of the loudspeaker system that guarantees coverage can be achieved for the entire audience. In immersive live sound reproduction, it is important that the coverage area of the fills system is well defined and independent of the position of audio objects and the number of loudspeakers being used by the spatial audio algorithm. This way, the target audience area of each fills system can be well defined and fills systems properly designed to create coverage in the corresponding audience area.

In this article, we present a novel spatial-fills algorithm that enables spatial rendering on so-called distributed fills systems such as front-fills and under-balcony fills. This spatial-fills algorithm is shown to provide an optimum balance in terms of coverage, spatial unmasking and audio-visual consistency, which are three essential qualities for effective immersive live sound. The algorithm is also shown to provide optimum time alignment of the fills system with all full-range sources of the scene system.

8 REFERENCES

1. E. Corteel, G. Le Nost, F. Roskam, “3D audio for live”, in 3D audio, edited by J Paterson, H Lee, 1st edition (Routledge), 2021.

2. V. Pulkki. "Virtual sound source positioning using vector base amplitude panning." Journal of the audio engineering society 45.6 (1997): 456-466.

3. V. Pulkki. "Uniform spreading of amplitude panned virtual sources." Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452). IEEE, 1999.

4. A. J. Berkhout, D. de Vries, and P. Vogel. "Acoustic control by wave field synthesis." The Journal of the Acoustical Society of America 93.5 (1993): 2764-2778.

5. E. Corteel. "On the use of irregularly spaced loudspeaker arrays for wave field synthesis, potential impact on spatial aliasing frequency." Proceedings of the 9th International Conference on Digital Audio Effects (DAFx'06). 2006.

6. Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. The Journal of the Acoustical Society of America, 106(4), 1633-1654.

7. “Variable Curvature Line Source”, skills development course, L-Acoustics, 2019.

8. H. Wierstorf, A. Raake, and S. Spors. "Binaural assessment of multichannel reproduction." The technology of binaural listening. Springer, Berlin, Heidelberg, 2013. 255-278.

9. H. Wierstorf, M. Geier, and S. Spors. "A free database of head related impulse response measurements in the horizontal plane with multiple distances." Audio Engineering Society Convention 130. Audio Engineering Society, 2011.

10. P. Søndergaard and P. Majdak "The Auditory Modeling Toolbox," in The Technology of Binaural Listening, edited by J. Blauert (Springer, Berlin, Heidelberg), pp. 33-56 (2013).

11. M. L. Hawley, R. Y. Litovsky, and J. F. Culling. "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer." The Journal of the Acoustical Society of America 115.2 : 833-843, 2004.

12. S. Jelfs, J. F. Culling, and M. Lavandier. "Revision and validation of a binaural model for speech intelligibility in noise." Hearing research 275.1-2: 96-104, 2011.

13. E. Hendrickx, M. Paquier, V. Koehl, and J. Palacino. “Ventriloquism effect with sound stimuli varying in both azimuth and elevation”, The Journal of the Acoustical Society of America. 138. 10.1121/1.4937758, 2013.