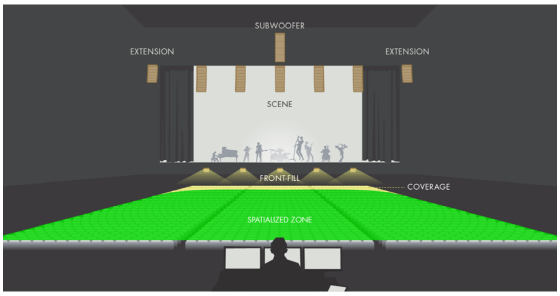
L-ISA HYPERREAL SOUND: OPTIMIZING COVERAGE AND SPATIAL RENDERING FOR FILL SYSTEMS

For loudspeaker systems with limited shared coverage (distributed fill systems like front-fills or under-balcony fills), the L-ISA technology offers the option of creating "spatial-fills". The approach first creates virtual replicates of the Scene system loudspeakers to restore cross-coverage and then uses gain-based algorithm for positioning audio objects. This improves audio object separation and audio-visual consistency while assuring coverage and level consistency.

L-ISA HYPERREAL SOUND: COVERAGE AND PANNING



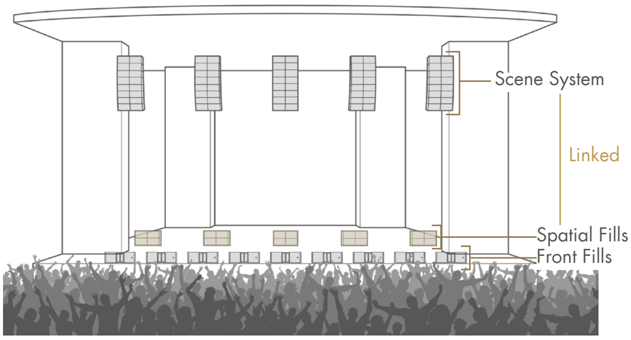
One of the core principles of the L-ISA technology is to separate coverage* of the audience from the spatial rendering step. Coverage is primarily assured by design, defining full-range sources that all address the same audience portion and may be complemented by fill systems to guarantee coverage for the entire audience. By doing so, coverage is assured independently from the chosen panning algorithm or object position. Until recently however, areas only covered by fill systems did not benefit from spatial sound reproduction.

A new spatial-fills solution is offered by the L-ISA technology, creating spatial sound in areas covered by distributed fill systems such as front-fills and under-balcony fills, which are classically fed by a mono signal. This option requires that any listener in corresponding audience area is in the coverage area of at least three loudspeakers. Based on this requirement, a first step consists in creating virtual loudspeakers to restore cross-coverage. The virtual loudspeakers are then fed with signals from the spatial rendering algorithm.

The spatial-fills solution of the L-ISA technology creates a good balance between coverage, object separation thanks to spatial unmasking*, and audio-visual consistency*, which are three essential dimensions and benefits of large-scale immersive systems.

BALANCING COVERAGE, SPATIAL UNMASKING AND AUDIO-VISUAL CONSISTENCY

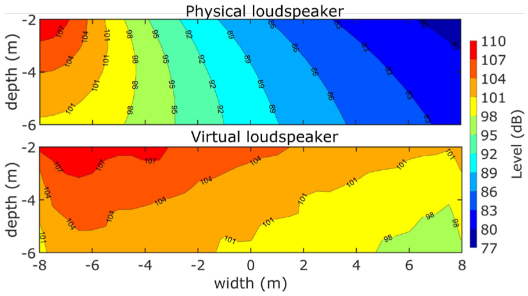
Loudspeaker system virtualization



The goal of spatial-fills is to improve coverage of the Scene system full-range sources. This is done by combining the distributed fill system sources and creating a virtual replica of the Scene system. The virtual loudspeakers are created using a delay-based algorithm with an optimized gain distribution using the following parameters:

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

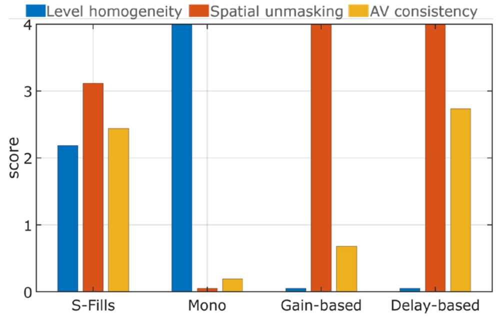
The figure on the right shows the SPL mapping (front-fills coverage area) of the outer house left physical loudspeaker of the front-fills system compared to the SPL mapping of the virtual replica of house left Scene loudspeaker. SPL differences between opposite positions in the audience is greatly reduced for spatial-fills compared to single loudspeakers, expanding the coverage area of the spatial-fills to the entire front-fills coverage area. See annex 1 for details on simulations.



Benefits for front-fills

The figure on the right compares various spatial rendering solutions for front-fills. The bars represent performance indicators (refer to annex 2, 3 and 4 for details).

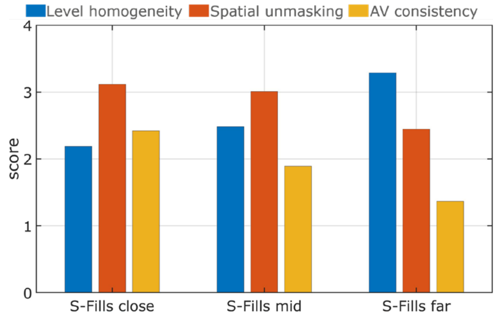
This comparison shows that only the spatial-fills solution of L-ISA (S-Fills) provides a good balance between the three criteria (level homogeneity, spatial unmasking, audio-visual consistency).



The Mono solution provides the best level homogeneity but no spatial unmasking and limited audio-visual consistency. The 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.

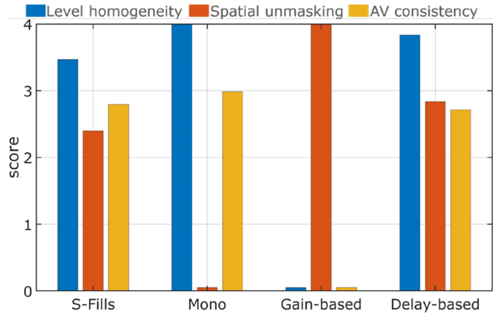
The delay-based solution (Wave Field Synthesis) performs best for spatial unmasking and audio-visual consistency but fails at providing good level homogeneity. For downstage object positions, the delay-based solution tends to concentrate all energy on a small number of loudspeakers which end up into a level homogeneity issue.

The second figure 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. Refer to annex 5 for details on audio-visual consistency.



Benefits for under-balcony

The same analysis is conducted for an under-balcony fills loudspeaker system (refer to annex 1 for information on test scenario). 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.



GLOSSARY

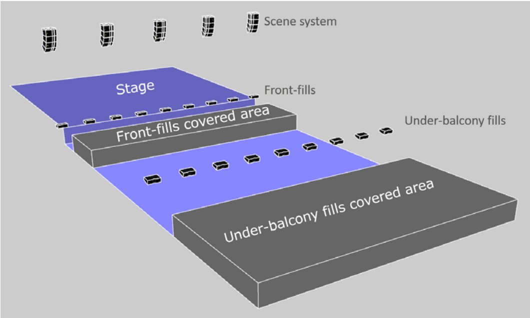
Audio object: Association of an audio input with metadata describing its properties such as spatial positioning.

Coverage: Area over which the loudspeaker system provides a direct sound in an acceptable frequency response variation.

Audio-visual consistency: Ability to perceive audio and visual cues associated to an object or a performer as originating from the same location.

Spatial unmasking: Ability of multiple sounds to unmask each other due to their separate and precisely perceived spatial origins.

ANNEX 1: TEST SCENARIO



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-fill systems consisting of 9 regularly spaced (2 m) 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:

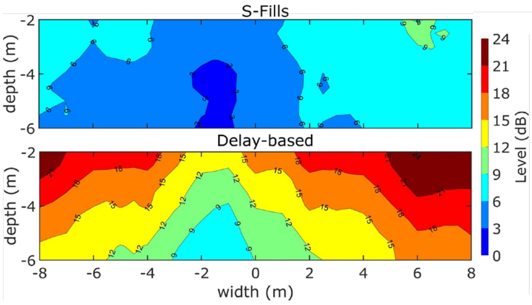
- Width: 0 (center), 2, -4, 6, -8 m,
- Depth (upstage): 2.5, 4, 5.5, 8, 12, 16 m.

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

- Front-fills: 5 m virtual distance (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 but are not represented here since they lead to similar trends and results.

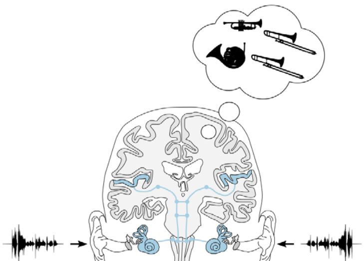
ANNEX 2: LEVEL HOMOGENEITY ESTIMATION

The level homogeneity is a criterion that is derived from the SPL generated in the 1 to 10 kHz bandwidth using Soundvision propagation model. 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.



The first figure here represents a level homogeneity mapping for spatial-fills in the tested front-fills configuration (see description in annex 1). The second figure represents level homogeneity for the delay-based algorithm, showing high level differences among the ensemble of tested audio-object positions (up to 24 dB). In this case, downstage object positions correspond to low gain values for off-axis loudspeakers thus corresponding to low rendered SPL due to the limited coverage of individual loudspeakers.

ANNEX 3: SPATIAL UNMASKING AND AUDIO-VISUAL CONSISTENCY ESTIMATION



Spatial unmasking and audio-visual consistency are evaluated with auditory models that have proven their accuracy. 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 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 are used:

- wierstorf2013_estimateazimuth for audiovisual consistency, providing an estimate of the localization error of the audio object
- jelfs2011 for spatial unmasking, providing an estimate of the increase in speech intelligibility of the target when the target and interferer are spatially separated.

---

[1] Kemar measurement from TU Berlin, 3 m distance, available here

ANNEX 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 (worst-case scenario) among all audience positions. 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 center, interferer at same depth but different left right location) and all audience positions.

Localization error is the median value (in °) among all object positions (width and depth on stage) and all audience positions.

A small localization error is associated with high audiovisual consistency.

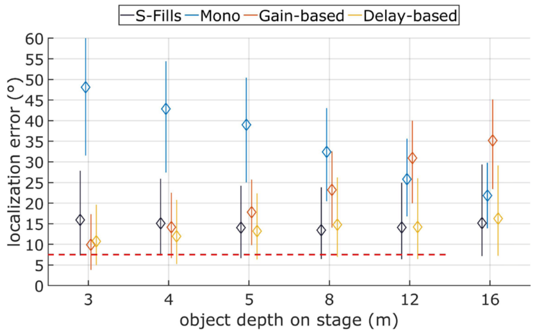
The quality score is obtained using the following thresholds (the higher the quality score, the better):

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

ANNEX 5: AUDIO-VISUAL QUALITY TRADEOFF AND VIRTUAL DISTANCE TUNING

This annex explores the type of localization error experienced by listeners within the coverage area of a front-fills system depending on the spatial rendering solution and the depth of the audio object on stage. Objects are on the outer house left positions on stage which is the most critical for localization error.

The first figure here 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).

The second figure compares multiple values of the virtual distance parameter of the spatial-fills solution (distance of the virtual loudspeaker from the front of the stage):

- S-Fills close: 5 m
- S-Fills mid: 8 m
- S-Fills far: 16 m

The close value provides the best overall results, with little dependency against the object depth on stage. Interestingly, this is also the value that allows for the best time alignment of the spatial-fills system with the scene system within their shared coverage area.