



L-ISA HYPERREAL SOUND: WHY AMPLITUDE-BASED ALGORITHMS ARE BEST SUITED TO FRONTAL LOUDSPEAKER CONFIGURATIONS?

For loudspeaker systems with shared coverage, the L-ISA technology uses an **amplitude-based** algorithm for positioning audio objects.

This maximizes natural separation of audio objects compared to delay-based algorithms, while maintaining audio-visual consistency.

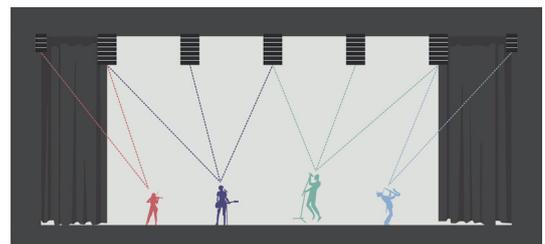
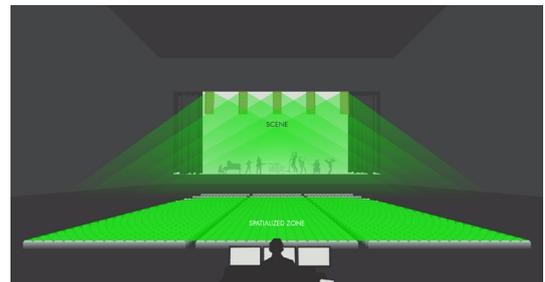
L-ISA HYPERREAL SOUND: BENEFITS AND REQUIREMENTS

L-ISA hyperreal sound provides high spatial definition for live:

- increasing the separation of audio objects* (spatial unmasking), limiting the use of equalization* and compression*,
- allowing the sound of performers to be perceived as coming from their location on stage (audio-visual consistency),
- offering new creative possibilities for mixing engineers.

The L-ISA design guidelines impose a minimum number of 5 loudspeakers located above the stage sharing the same coverage*.

The L-ISA panning algorithm* enables to position audio objects at their intended location, spreading the signal of audio objects over one or several loudspeakers using an amplitude-based approach. This algorithm provides accurate positioning information to the entire audience, allowing for audio-visual consistency and outperforming delay-based algorithms in terms of spatial unmasking and coloration*.



PERCEPTION OF SPATIAL SOUND REINFORCEMENT

Audio-visual consistency

Considering here a frontal stage, performers are positioned along the width and the depth of the stage, on or above the stage, depending on the scenography of the show or the venue. The audio-visual consistency should therefore be envisaged in all three dimensions.

In the vertical dimension, audio-visual consistency is entirely related to the loudspeakers' height. The height should be set to create less than 30 degrees of vertical separation between performers and the loudspeaker system, to allow for audio-visual consistency in the vertical dimension.



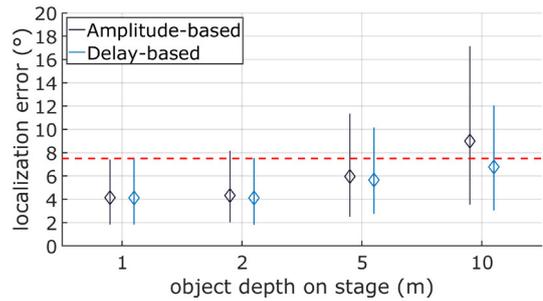
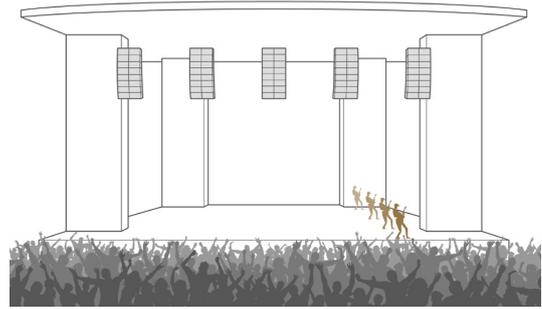
In the horizontal dimension, an error of 7.5° between target and actual auditory positioning is considered as negligible to guarantee audio-visual consistency. Angular localization* depends on the position of the performer along the width but also along the depth of the stage, particularly for offside listeners.

Audio-visual consistency in the horizontal dimension therefore depends on:

- performer (resp. listener) position along the width and depth of the stage (resp. venue),
- the loudspeaker system resolution (i.e. number of loudspeakers),
- the panning algorithm (amplitude-based or delay-based).

A system of 5 loudspeakers regularly spaced along a 16 m wide stage is considered here. A virtual listening test was created, computing responses at listeners' ears and using an auditory model to estimate localization (see annex 2 for details).

Results show that the localization error is less than 7.5° in the horizontal dimension for most performer and listener locations. Amplitude-based (LISA) and delay-based algorithms perform similarly in terms of localization error, even for objects located upstage (see annex 3 for details).

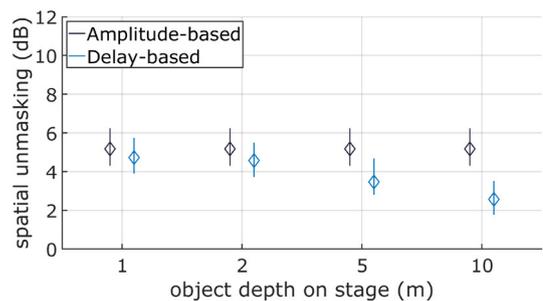
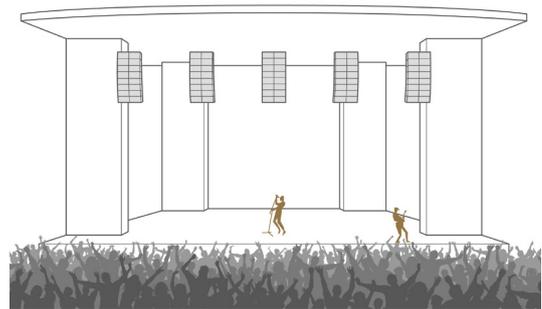


Improving the auditory separation of multiple audio objects

Spatial unmasking is the ability to identify what a specific performer plays, amongst concurrent performers (for example, lead singer against background vocal or keyboard). This typically reduces the need of compression or EQ on the background objects to maintain the intelligibility of the lead vocal. The spatial unmasking is estimated here in dB (see annex 2 for details).

In this example, the lead vocal is centered, and the background object is located at 6 m house-right. Both objects are tested at the same depth onstage (1, 2, 5, 10 m behind the loudspeaker system).

The spatial unmasking is always larger for amplitude-based than for delay-based algorithms. In delay-based algorithms, the apparent angular separation of performers decreases and the localization blur increases when performers move upstage, which in turns reduces by half the spatial unmasking (see annex 3 for details). Amplitude-based algorithms maintain a large spatial separation of performers and minimize the localization blur.



GLOSSARY

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

Coloration: Audible alteration of the natural tonal characteristics of a sound.

Compression: Reduction of the dynamic range of a signal.

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

Equalization: Tool or process aimed at electronically adjusting the frequency response of an audio system.

Localization: Perceived spatial origin of a sound.

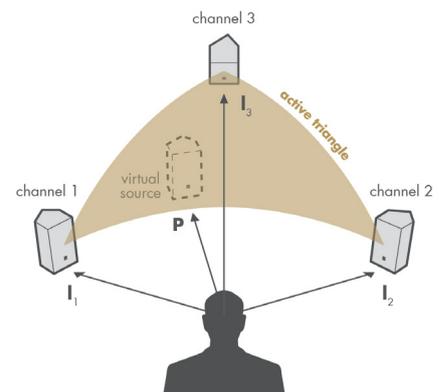
Panning algorithm: Process that modifies audio signals sent to the loudspeakers to control the spatial origin of the associated audio object.

ANNEX 1: SPATIAL PANNING ALGORITHMS

Amplitude-based: L-ISA panning algorithm

The spatial panning in L-ISA is based on amplitude panning, using two (resp. three) loudspeakers for 2D (resp. 3D) panning. The algorithm is based on Vector Based Amplitude Panning (VBAP) and Multiple Directions Amplitude Panning (MDAP). The L-ISA algorithm brings two major enhancements to the VBAP/MDAP approaches:

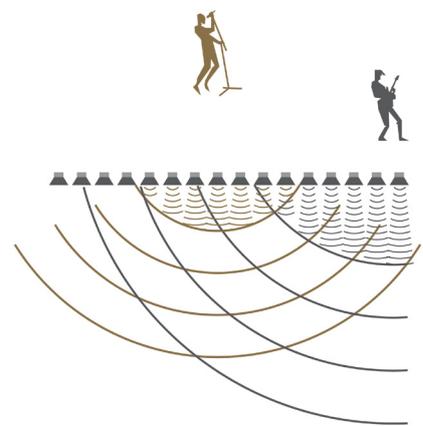
- a low-frequency build-up compensation algorithm for objects positioned between loudspeakers,
- an original decorrelation algorithm for the control of width of objects that minimizes tonal coloration and temporal artefacts when combining multiple loudspeakers.



Delay-based: Wave Field Synthesis

Wave Field Synthesis is a delay-based panning technique that theoretically reproduces the physical sound field emitted by an audio object within the entire audience. 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 restricts physically accurate reproduction below 100 Hz for typical large-scale systems (loudspeaker spacing >4 m). In this case, there is no physical wave front reconstruction and the localization is driven by the first loudspeaker contribution that reaches the listener (precedence-based localization). WFS with such a small number of loudspeakers is therefore analogous to delta stereophony.



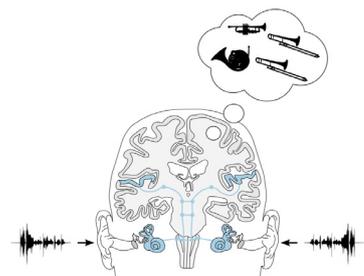
L-ISA vs. delay-based algorithms

In addition to localization, we can also compare L-ISA and delay-based algorithms on the following aspects:

- Coverage: delay-based algorithms can extend coverage by multiplying the number of active sources, but such coverage extension is not needed in L-ISA loudspeaker system designs.
- Coloration depending on object position: delay-based algorithms modify time-alignment between loudspeakers when an object changes position. This is not the case with amplitude-based algorithms.
- Coloration artefacts with moving objects, trajectories: moving delays in real-time for delay-based algorithms can easily introduce artefacts, either clicks for fast moving objects or lagging to allow enough time for cross fading, as well as revealing large coloration change between object positions.

ANNEX 2: PANNING ALGORITHMS EVALUATION FRAMEWORK

Auditory models have proven their accuracy for localization estimate of loudspeaker-based spatialization systems and are therefore used here instead of conducting perceptual experiments. 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 environment¹ are used to simulate a free field situation, not considering the reverberation of the environment but concentrating on the direct sound only.



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

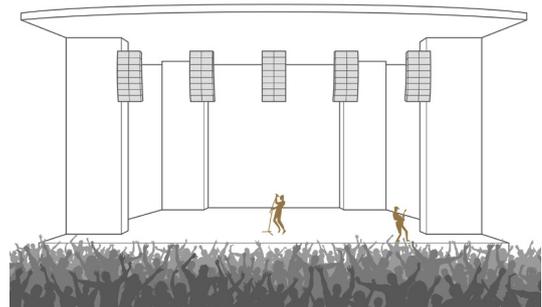
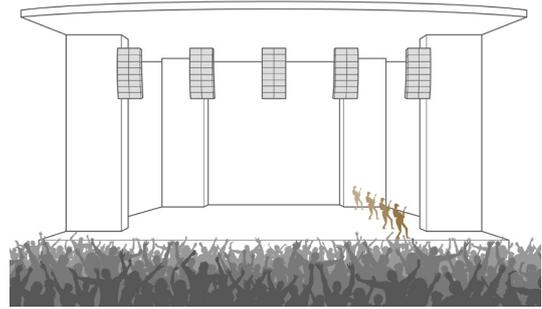
The following parameters are considered:

- 5 loudspeakers regularly spaced (4 m) along a 16 m wide stage, located at 6 m above the stage,
- 81 positions in the audience (16*16 m, every 2 m),
- 4 object positions at 6 m house-right: 1, 2, 5, 10 m behind the loudspeaker system.

More stage width (12 and 20 m), object lateral positions (centered, and off-centered), audience sizes were tested but not reproduced here as results are very similar and show the same tendency.

Two models of the [Auditory Modeling Toolbox](#) are used:

- wierstorff2013_estimateazimuth for localization estimation, providing an estimate of the horizontal localization and the associated standard deviation (see annex 3 for details).
- jelfs2011 for spatial unmasking, estimating the increase in speech intelligibility of the target when the target and interferer are spatially separated. The test situation here corresponds to a lead singer in the center and an harmonic instrument located house-right having a similar frequency range as speech.



ANNEX 3: LOCALIZATION PRECISION: ACCURACY VS. BLUR

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 analyzed 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.

When assessing the localization error, the spread of participants responses is simulated so that the standard deviation of all responses corresponds to the estimated localization blur. The localization error is calculated as the absolute difference between simulated participants responses and the target localization for each test listening position. The values in the graph represent the median localization error (diamond) and the 25th and 75th percentile (lower and upper end of the vertical bar).

The amplitude-based algorithm only makes use of the right outer two loudspeakers with about equal level to create the **house-right** audio object, creating a localization blur of about 5 degrees. A delay-based algorithm induces a localization blur that increases when the objects gets further upstage and can typically go up to 10 degrees (median, diamond in graph at 10m upstage). Indeed, in delay-based algorithms, the further the source moves upstage, the more loudspeakers are active with significant level, thus increasing localization blur.

The localization blur can be minimized by placing the source directly on-axis to a loudspeaker, for example in the **center** of the stage. In L-ISA, this reduces the blur to 2 degrees, which is the lowest threshold of the auditory system. In delay-based algorithm, the blur increases gradually with object depth, reaching 8 degrees at 10 m distance, in the same range as the house-right position.

