SYNTHETIC TRANSAURAL AUDIO RENDERING (STAR): A PERCEPTIVE APPROACH FOR SOUND SPATIALIZATION

Eric Méaux
L3i
University of La Rochelle, France
eric.meaux01@univ-lr.fr

Sylvain Marchand
L3i
University of La Rochelle, France
sylvain.marchand@univ-lr.fr

ABSTRACT

The principles of Synthetic Transaural Audio Rendering (STAR) were first introduced at DAFx-06. This is a perceptive approach for sound spatialization, whereas state-of-the-art methods are rather physical. With our STAR method, we focus neither on the wave field (such as HOA) nor on the sound wave (such as VBAP), but rather on the acoustic paths traveled by the sound to the listener ears. The STAR method consists in canceling the cross-talk signals between two loudspeakers and the ears of the listener (in a transaural way), with acoustic paths not measured but computed by some model (thus synthetic). Our model is based on perceptive cues, used by the human auditory system for sound localization. The aim is to give the listener the sensation of the position of each source, and not to reconstruct the corresponding acoustic wave or field. This should work with various loudspeaker configurations, with a large sweet spot, since the model is neither specialized for a specific configuration nor individualized for a specific listener. Experimental tests have been conducted in 2015 and 2019 with different rooms and audiences, for still, moving, and polyphonic musical sounds. It turns out that the proposed method is competitive with the state-of-the-art ones. However, this is a work in progress and further work is needed to improve the quality.

1. INTRODUCTION

The purpose of sound spatialization (or "3D sound") [1] is to give the listener the sensation that the sound is coming from a certain position in space, or that he/she is surrounded by sounds, etc.

This research area is not new, and several state-of-the-art methods have been proposed, such as Vector Base Amplitude Panning (VBAP) proposed by Pulkki [2], Ambisonics proposed by Gerzon [3] and generalized to higher orders (High Order Ambisonics or HOA) by Daniel [4], or Wave Field Synthesis (WFS) introduced by Berkhout [5]. However, all these methods are based on physics, and tend to reproduce either the sound wave (VBAP) or the sound field at the position of the listener (HOA) or everywhere in space (WFS). Moreover, WFS requires a lot of calibrated loudspeakers, HOA requires also calibration, less speakers but with a sound reproduction localized to a sweet spot. Since we want a system that can be used in practical situations, with a limited number of loudspeakers, in a room where the acoustics cannot be optimized, and computationally efficient, the only option seems to be VBAP, which appears to be a good choice in practice for the musical situations we are interested in [6].

Copyright: © 2018 Eric Méaux et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

However, together with Mouba we proposed in [7, 8] the principles of a perceptive approach for sound spatialization in practical conditions. We focus neither on the wave field (such as HOA) nor on the sound wave (such as VBAP), but rather on the acoustic paths traveled by the sound to the listener ears. Our method consists in canceling the cross-talk signals between two loudspeakers and the ears of the listener (in a transaural way), with acoustic paths not measured but computed by some model (thus synthetic). Like MPEG Surround [9], our model is based on perceptive cues, used by the human auditory system for sound localization. The aim is to give the listener the sensation of the position of each source, and not to reconstruct the corresponding acoustic wave of field. This should work with various loudspeaker configurations, with a large sweet spot, since the model is neither specialized for a specific configuration nor individualized for a specific listener.

The original method suffered instabilities for some configurations, as well as for extreme frequencies. Since then, the method has been enhanced and we propose now the Synthetic Transaural Audio Rendering (STAR) method.

Experimental tests have been conducted in 2015 and 2019 with different rooms and audiences, for still, moving, and polyphonic sources. The HOA, VBAP, and STAR methods have been compared.

The remainder of this paper is organized as follows. Section 2 introduces the STAR method, Section 3 describes the practical experiments used to evaluate the method, then Section 4 presents the results of experiments conducted in 2015 and 2019, before a conclusion and some perspectives in Section 5.

2. THE STAR METHOD

The principles of Synthetic Transaural Audio Rendering (STAR) we detail here were first introduced in [7, 8].

This method is suitable for spatial audio objects. Each object (or source) consists of a signal to be played at a given position. For now, we focus only on the azimuth θ in the horizontal plane.

For short, the STAR method consists in canceling the cross-talk signals between two loudspeakers and the ears of the listener (in a transaural way), with acoustic paths not measured but computed by some model (thus synthetic). Our model is based on perceptive cues, used by the human auditory system for sound localization. The aim is to give the listener the sensation of the position of each source, and not to reconstruct the corresponding acoustic wave of field. This is indeed a perceptive approach.

In a setup with many speakers such as the one illustrated on Figure 1, we use the classic pair-wise paradigm [10], consisting in choosing for a given source only the two speakers closest to it (in azimuth): one at the left of the source, the other at its right. This is the same choice as in the VBAP method in this two-dimensional case. Of course, when the source is exactly on one speaker, the source signal is directly played from this speaker and thus the spatialization process is bypassed.

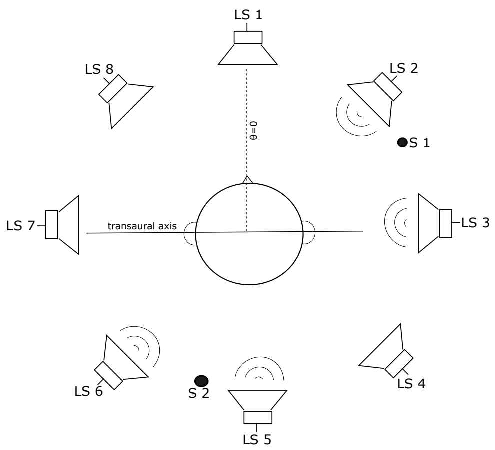


Figure 1: *Octophonic setup, thus with eight loudspeakers (LS), and with two sound sources (S). S1 is located between LS 2 and LS 3, and S2 is between LS 5 and LS 6.*

STAR operates in the spectral domain. Each source signal is passed into the frequency domain with a Short-Time Fourier Transform (STFT), using the Fast Fourier Transform (FFT), processed and distributed among the loudspeakers, then the signal for each loudspeaker is obtained from the spectral domain using the inverse FFT, see Figure 2. Thus, for n sources and m loudspeakers, we compute $n + m$ FFTs in total (i.e. 10 FFTs in the case illustrated in Figure 1). In practice, with use a Hann window and frames of 1024 samples at 44100Hz, with 50% overlap.

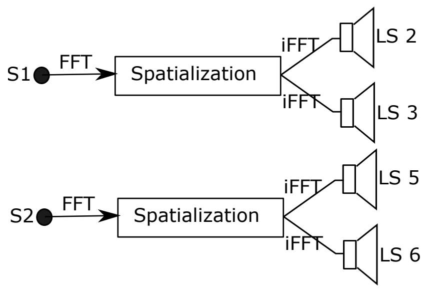


Figure 2: *General principle of STAR spatialization.*

Since low frequencies can cause problems with windowing in the STFT, such as clicks in the case of moving sources (thus changing parameters) when there are not enough periods of the signal within the window w, we filter out the frequencies below 150Hz prior to the spatialization and re-inject them equally in all the loudspeakers afterwards (we could even use a subwoofer, although we preferred not to use this possibility in our experiments). This is not problematic since human beings hardly localize such low frequencies (see [1]).

2.1. Synthetic Paths

With the STAR method, we consider the acoustic paths traveled by the sound to the listener ears. These paths are represented in the spectral domain by their transfer functions, and derived from interaural cues using a model.

The Interaural Time Difference (ITD) corresponds to the travel time difference of a sound between the two ears, while the Interaural Level Difference (ILD) corresponds to the level difference between the two ears.

These interaural cues can be derived from real Head-Related Transfer Functions (HRTFs), which are the spectral versions of the Head-Related Impulse Responses (HRIRs) that can be found, for example, in the CIPIC database [11]. More precisely, we have:

$$\begin{aligned} \mathrm{ILD}_{\mathrm{real}}(f) &= -20\log_{10}(|\mathrm{HRTF}_L(f)/\mathrm{HRTF}_R(f)|) \\ \mathrm{ITD}_{\mathrm{real}}(f) &= -\angle(\mathrm{HRTF}_L(f)/\mathrm{HRTF}_R(f))/(2\pi f) \end{aligned} \tag{1}$$

The HRTFs depend on the subject. Since we are considering only the azimuth, individualization is not really necessary and we could consider average HRTFs among subjects (see Figure 3). However, after Viste [12], in our system we use a model for each

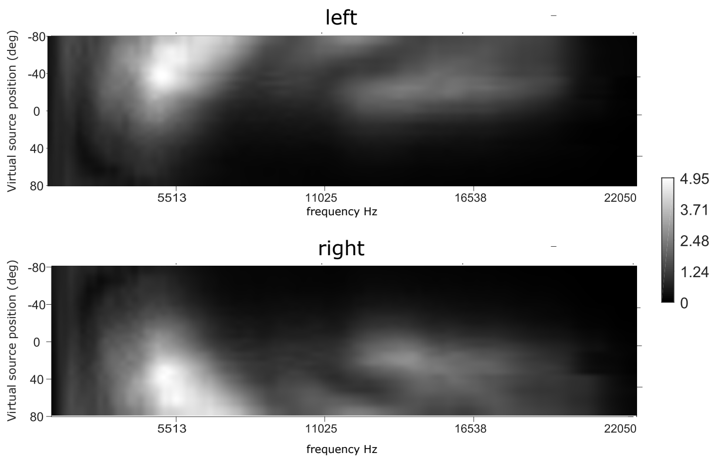


Figure 3: *HRTF magnitude as a function of azimuth and frequency.*

interaural cue. Mouba et al. proposed in [8] the following models:

$$\begin{aligned} \mathrm{ILD}(\theta, f) &= \alpha(f)\sin(\theta) \\ \mathrm{ITD}(\theta, f) &= \beta(f) r \sin(\theta)/c \end{aligned} \tag{2}$$

where α and β are scaling factors obtained from the CIPIC database [11] by matching the model to the data, in the least-square sense (see Figure 4). The overall error for all subjects, azimuths, and frequencies is of 4.29dB for the ILD and 0.052ms for the ITD.

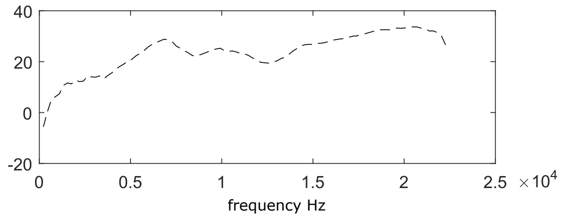
Now that we have synthetic interaural cues (see Figure 5), we can propose synthetic paths respecting these cues, first by computing

$$\begin{aligned} \Delta_a(f) &= \mathrm{ILD}(\theta, f)/20 \\ \Delta_\phi(f) &= \mathrm{ITD}(\theta, f) \cdot 2\pi f \end{aligned} \tag{3}$$

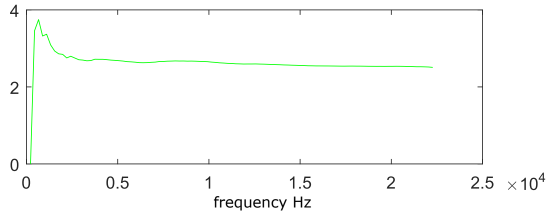
then by using the fact that the left and right HRTFs are roughly symmetric (see Figure 3), we propose

$$\begin{aligned} H_L &= 10^{+\Delta_a(f)/2} \cdot e^{+i\Delta_\phi(f)/2} \\ H_R &= 10^{-\Delta_a(f)/2} \cdot e^{-i\Delta_\phi(f)/2} \end{aligned} \tag{4}$$

where H_L and H_R are the paths going to the left and right ears, respectively.

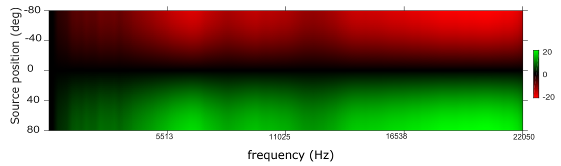


(a) α scaling factor (ILD)

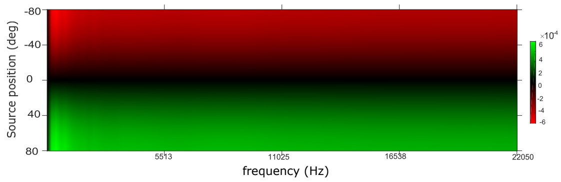


(b) β scaling factor (ITD)

Figure 4: *α and β scaling factors.*



(a) synthetic ILD (dB)



(b) synthetic ITD (s)

Figure 5: *Synthetic ILD (a) and ITD (b).*

2.2. Transaural Principle

The STAR method is largely based on the transaural principle. As shown in Figure 6, we aim at reproducing the paths H_L and H_R between the (virtual) source and the left and right ears of the listener, using the (real) acoustic path between each loudspeaker and each ear (e.g. H_{LR} denoting the path from the left loudspeaker to the right ear).

More precisely, for a given sound s (S being its spectral version), the sounds measured at the left and right should be $H_L \cdot S$ and $H_R \cdot S$, respectively. But to reproduce this virtual source, we use instead two real loudspeakers (the pair at the left and right of the source), thus we must verify the following equation system

$$\begin{aligned} H_L \cdot S &= K_L \cdot H_{LL} \cdot S + K_R \cdot H_{RL} \cdot S \\ H_R \cdot S &= K_L \cdot H_{LR} \cdot S + K_R \cdot H_{RR} \cdot S \end{aligned} \tag{5}$$

where K_L and K_R are some coefficients to be applied to the left and right loudspeakers, respectively. These two coefficients are the solutions of the preceding two-equation system, where S can be simplified.

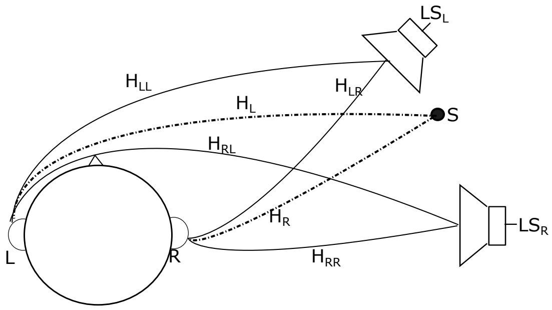


Figure 6: *Transaural principle: 4 real acoustic paths (H_{LL}, H_{RL}, H_{LR}, and H_{RR}) used to reproduce 2 virtual ones (H_L and H_R).*

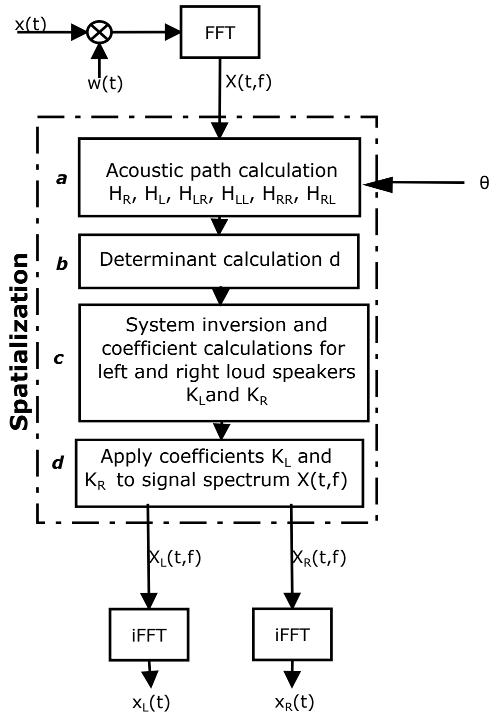


Figure 7: *STAR processing chain.*

Figure 7 then summarizes the whole STAR processing chain.

First (Part a), we start by calculating every acoustic path (as shown on Figure 6), using the procedure described above in Section 2.1, thus with Equation (4).

Second (Part b), we compute the system determinant:

$$d = H_{LR} \cdot H_{RL} - H_{LL} \cdot H_{RR} \tag{6}$$

Third (Part c), it is now possible to invert the system to find the loudspeaker coefficients, and more precisely:

$$\begin{aligned} K_L &= (H_R \cdot H_{RL} - H_L \cdot H_{RR})/d \\ K_R &= (H_L \cdot H_{LR} - H_R \cdot H_{LL})/d \end{aligned} \tag{7}$$

The fourth and last step (Part d) consists in applying the coefficients to the signal spectrum for the left and right speakers of the speaker pair:

$$
\begin{aligned}
X_L(t,f) &= K_L(f) \cdot X(t,f) \\
X_R(t,f) &= K_R(f) \cdot X(t,f)
\end{aligned} \tag{8}
$$

2.3. The Determinant

Of course, things are not so simple in practice. For example, the system determinant d (see Equation (6)) plays an essential role in the STAR method. It shall not approach the 0 value, or else the system is ill-conditioned. As shown in Equation (6), the determinant only depends on the paths of the speakers, thus on the positions of these speakers relatively to the ears of the listener. Of course, a problem would happen if the two speakers were at the same position (which is supposed to be impossible), or very close. Figure 8 shows the determinant norm as a function of the speakers spacing. We see that is necessary to have an angle between the two speakers which is greater than 2 degrees to have a determinant norm greater than 0.01 (the value we chose to guarantee the stability of the system). Hopefully, this will be the case in practice since unlike Wave Field Synthesis (WFS) [5], STAR aims at addressing sparse speaker configurations. In fact, a problem arises with the

Figure 8: *Determinant norm as a function of frequency and speakers spacing (in red where it is below* 0.01*).*

low frequencies, but hopefully recall that they are filtered out prior to the spatialization.

But if nothing is done, because of the symmetry of our synthetic paths (see Equations (4)), a null determinant problem also appears if the speakers are placed symmetrically with the transaural axis, as in our experimental setup (see Figure 15, LS 2 and 3, or LS 6 and 7). A solution could be to shift the azimuth reference (axis rotation) to break the symmetry of the paths, and move to a problem-free configuration such as the one of Figure 1. This has been done in our first experiments (prior to 2018). For this article, we made a more radical choice: placing the azimuth reference at the center of the loudspeaker pair. This way, the determinant only depends on frequency (and no more on the azimuth). Figure 9 shows the resulting determinant, which is problem-free except for very low frequencies (which are filtered). Studying the influence of this azimuth reference is part of our ongoing research.

2.4. The Coefficients

The calculation of coefficients K_L and K_R described in Equation (7) is the last step of the method.

Even if the determinant of the system is correct, we have to verify that the solutions are also correct. For example, the bigger are the coefficient values, the bigger is the risk to have a saturation on the speakers. Moreover, unlike VBAP or HOA, our coefficients are complex.

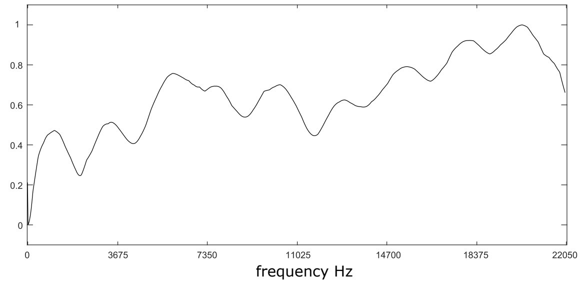


Figure 9: *Determinant norm as a function of the frequency value (here for $\theta = 0$).*

Figure 10 shows the modulus of the left and right coefficients depending of the position of the virtual source. The value is mostly between 0 and 1.4, and never exceeds 1.6, thus a risk of saturation exists, for a loud sound source.

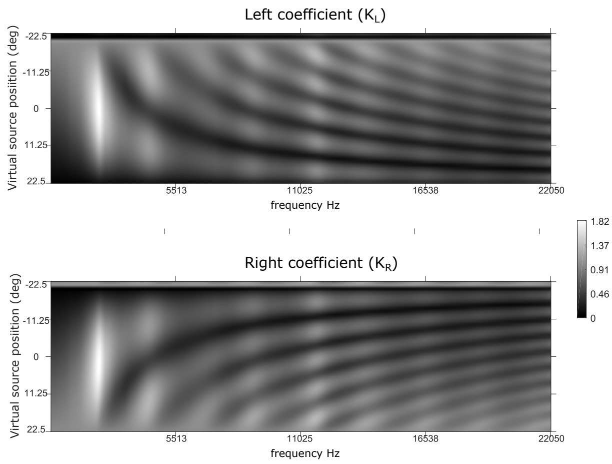


Figure 10: *Magnitude of the K_L and K_R coefficients as a function of frequency and source azimuth.*

Figure 11 compares the STAR and VBAP coefficients for a virtual source placed at 0 deg, which is in the middle of two speakers in the configuration of our experiment (see Figure 15), the two loudspeakers being then placed relatively at ± 22.5 deg. Although the STAR coefficients are complex, and spectral (i.e. dependent on the frequency), we can see that their modulus stays relatively close to the coefficients of VBAP (which are constant).

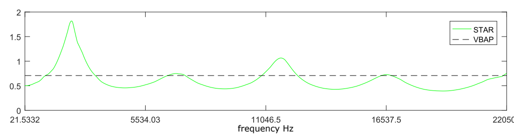


Figure 11: *VBAP (dashed black) and STAR (plain green) coefficients comparison (for $\theta = 0$, where the left and right coefficients are the same).*

Figures 11 and 10 also show the sinusoidal and symmetric aspects of the ITD and ILD used to compute the coefficients, in the first step of the method.

3. EXPERIMENTS

The experiments consist of a pre-test followed by three tests. The pre-test is used to identify people who are not sensitive to spatialization. The aim of the three others is to compare the STAR method with two state-of-the-art methods (VBAP and HOA panning) in different situations.

For each of these three tests, a sound example is played using VBAP, HOA, and STAR, together with a spatial anchor consisting of the monophonic version of the sound (with full bandwidth). The sounds are then adjusted to the same volume. The resulting 4 stimuli are randomly attributed a letter (A, B, C, or D), and the sequence (A, B, C, and D) is played and repeated.

The subjects are asked to answer for each method, possibly randomly if they do not know what to answer (for example in the case of the anchor). They also have the possibility to write free comments for each test.

3.1. Preliminary Test

A preliminary test is used for subjects who are neither sound experts nor used to spatial sound. This is a "warm up", so that the subjects listen to the sound system and focus on the spatial aspects. Moreover, this pre-test allows us to identify the subjects who do not pay attention to the spatialization, and thus have to be ruled out from the panel for the results.

For this pre-test, 4 different bird sound examples are played 4 times each, but at different positions. For each example, the first time is the reference, and the subject should retrieve this reference randomly hidden among the 3 other sounds. Thus one position is correct (the one of the reference), and two are far from the one of the reference sound, so that this exercise should be easy for non experts.

3.2. Static Test

In the static test, a musical excerpt is played by a saxophone at a fixed azimuth. The subjects are then asked two things: first, to localize this single source, by placing the letter corresponding to the method on a reference circle (see Figure 12); second, to evaluate the quality of the sound, on a MUSHRA-like [13] notation scale (see Figure 13).

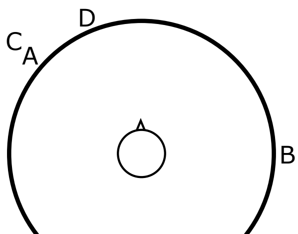


Figure 12: *Reference circle.*

3.3. Dynamic Test

The aim of the dynamic test is to compare the spatialization methods in the case of a moving source. For this purpose, we created a circular trajectory (direct orientation) on a percussive music

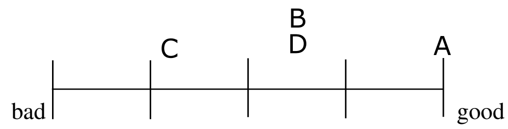


Figure 13: *Notation scale.*

(tablas). Again, the subjects are asked two things: first, to choose the trajectory they find the best among 8 possibilities (see Figure 14); second, again to evaluate the quality of the sound.

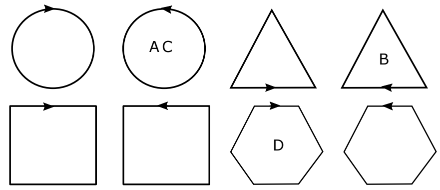


Figure 14: *A choice of* 8 *trajectories, the correct one being the second one (circle in the direct direction).*

3.4. Polyphonic Test

The third test is a polyphonic case. A pop musical song (jazz) is spatialized, with singers and instruments (drums, bass, saxophone, guitar, keyboards) as spatial audio objects, i.e. individual sources distributed in space with positions closest to the choice of the sound engineer for the artistic mix. Moreover, this musical excerpt has the advantage of presenting singing voice with different dynamics, together with various instruments, and ends with a cappella. This time we ask the listeners to evaluate three parameters, all on the notation scale of Figure 13: the quality (like in the two previous tests), the immersion, and the "intelligibility" (or clarity).

4. RESULTS

In 2015, during the Electrocution festival for electroacoustic music in Brest, France, the previous experiments were conducted by a group of Master's students, with an octophonic configuration placed in a quite reverberant concert hall (a former factory made of concrete...), with an audience constituted by composers, sound engineers, and other people with a majority of music professionals used to spatial sound (29 subjects in total). This configuration was exactly the one used for the diffusion of the concerts of the festival.

The conclusion of these experiments showed that STAR had a large sweet spot, was better for the dynamic test (VBAP was the worse, with many hexagons chosen instead of circles, i.e. jumps between loudspeakers), and preferred for the polyphonic test, although the sound timbre has sometimes been described as "nasal quality". Thus, the results were quite promising, although we had to fix this timbre problem, thought to be large coefficient values in the high frequencies (producing a high-pass filter effect).

For the present article, we decided to re-conduct these experiments in La Rochelle, again with an octophonic configuration but in a classroom with moderate reverberation. Figure 15 describes the setup. LS 1 to LS 8 are the active loudspeakers, B are 4 baits (inactive loudspeakers, to artificially increase the complexity of the setup, because it was impossible to hide the other loudspeakers), and 9 seats placed in the middle of this setup (S 1 to S 9).

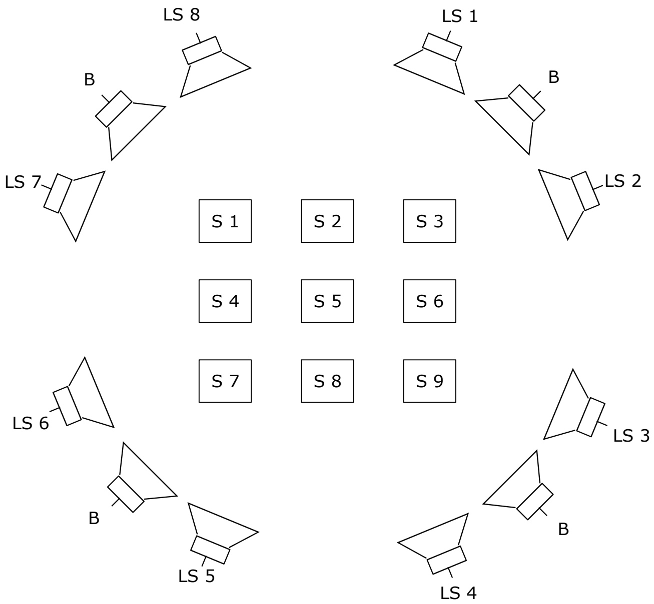


Figure 15: *Experimental setup (2019): 8 active loudspeakers (LS 1 to 8), 4 inactive ones or baits (B), and 9 seats (S 1 to 9).*

The panel consisted of 32 persons (the experiment was run 4 times), almost exclusively composed of amateurs or neophytes on music (only 1 music professional), mainly students and staff of the technical university. After the pre-test only 2 persons were eliminated (people who did not manage to find the reference for at least half of the 4 examples). Thus, the final panel consisted of 30 people (8 women and 22 men), ranging from 17 to 49 years old, with 23 of them below 25 years old.

4.1. Static Test

Figure 16 summarizes the results of the static test, where the listeners were asked to localize a static source, in terms of mean and standard deviation. It appears clearly than the (mono) anchor score is very bad, which is normal. For the three methods the mean is not very far from the real position of the virtual source, but it is clear than VBAP and STAR are better than HOA (which uses all speakers, which can be a drawback for a small room and thus some subjects seating relatively far from the sweet spot or too close to one speaker). The STAR method exhibits the best mean value but a larger standard deviation than VBAP, which seems the best choice. Principal Component Analysis (PCA) showed that the positions of the listeners and their perception of the source position are clearly correlated. More precisely, the listener tends to perceive the sound in the direction of the closest loudspeaker, which is not surprising but problematic.

Regarding the perceived quality, the results (see Figure 17) are quite surprising, because all methods show comparable results, with a mean in the middle and a large standard deviation. This might be a problem with the anchor, which is mono but with full bandwidth, thus with a probably too high quality (for an anchor).

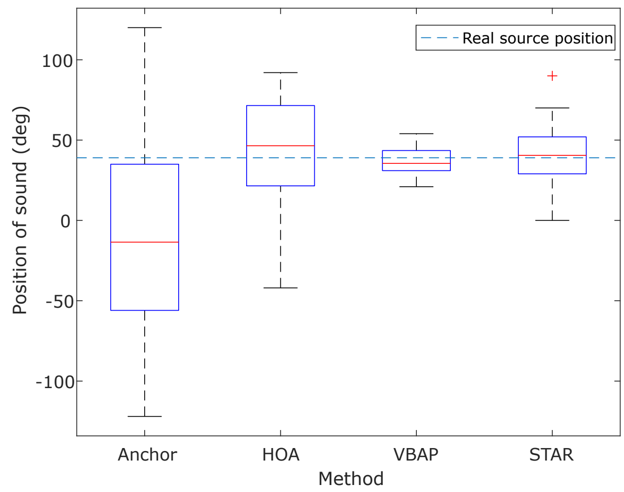


Figure 16: *Static test: perceived source localization. The correct source azimuth is materialized by the dashed line.*

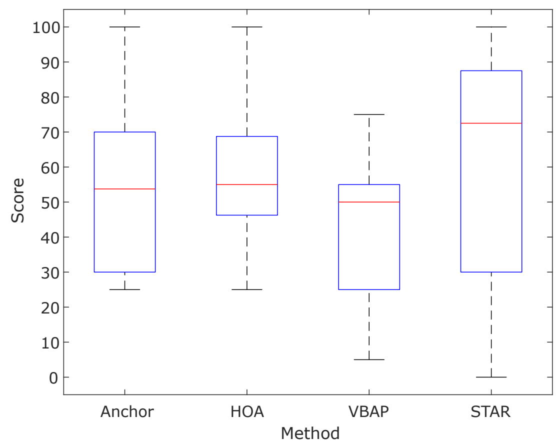


Figure 17: *Static test: perceived sound quality.*

4.2. Dynamic Test

Figure 18 summarizes the results of the dynamic test, where the listeners were asked to recognize the trajectory described in time by a moving source. The anchor exhibits a random behavior, which is normal, since there is no trajectory rendered by this mono version. Then, for all methods, there is some hesitation between the circular (correct one) and hexagonal trajectories. HOA seems to perform best, followed by VBAP then STAR. This looks surprising to us, because for the test conducted in 2015 (see Figure 19) the STAR method was first (and not last. . .). Apart from the room characteristics and audience qualification, the only change between the 2015 and 2019 tests is the fact that, because we suspected that this was the cause of the “nasal quality”, we chose to place the azimuth reference at the center of the loudspeaker pair to improve the system determinant (see Section 2.3). This might be a bad choice.

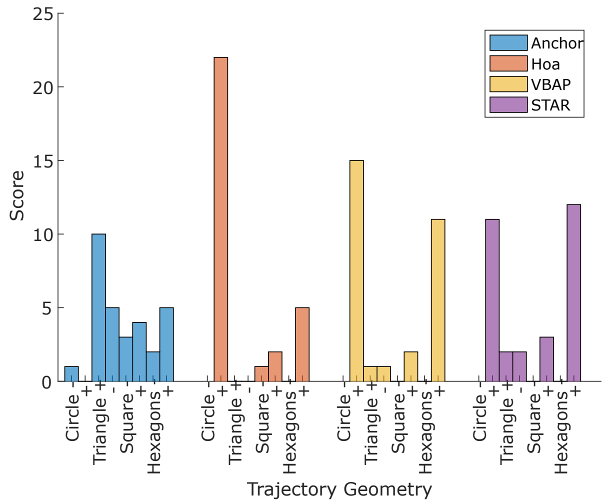


Figure 18: *Dynamic test: perceived source trajectory.*

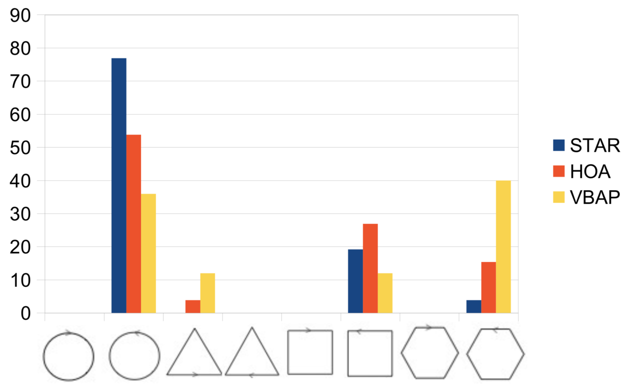


Figure 19: *Dynamic test: perceived source trajectory (values in percents), for the test conducted in 2015.*

Regarding the perceived quality, the results (see Figure 20) show that the anchor gets a lower score (which is normal since the anchor does not move), but the score of the three methods are quite similar. Again, this might be a problem of a too high quality anchor for inexperienced listeners.

4.3. Polyphonic Test

Figure 21 shows the perceived quality in the case of the polyphonic test. This time, the anchor is statistically lower, but all the three methods are judged equally good. The results are consistent between 2015 and 2019, even if in 2015 STAR was preferred, but in a non statistically significant way.

In 2019, we asked for subjective immersion and intelligibility (but not in 2015). Figure 22 shows the the perceived immersion is very similar to the perceived quality. Regarding intelligibility (see Figure 23), STAR seems to have some problems, which might be explained by the fact that unlike HOA and VBAP, the STAR coefficients are spectral and complex, thus modify also the phase in a frequency-dependent way, which might help smoothing trajectories but also might modify the timbre of the sound sources.

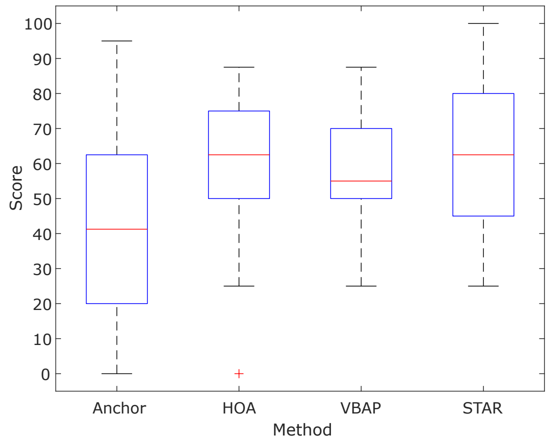


Figure 20: *Dynamic test: perceived sound quality.*

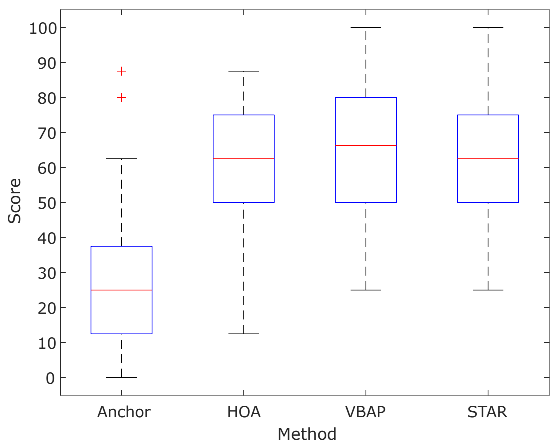


Figure 21: *Polyphonic test: perceived sound quality.*

5. CONCLUSIONS AND FUTURE WORK

In this article we proposed a perceptive approach for sound spatialization. With our STAR method, we focus neither on the wave field (such as HOA) nor on the sound wave (such as VBAP), but rather on the acoustic paths traveled by the sound to the listener ears. The STAR method consists in canceling the cross-talk signals between two loudspeakers and the ears of the listeners (in a transaural way), with acoustic paths not measured but computed by some model (thus synthetic). Our model is based on perceptive cues, used by the human auditory system for sound localization. The aim is to give the listener the sensation of the position of each source, and not to reconstruct the corresponding acoustic wave of field. This should work with various loudspeaker configurations, with a large sweet spot, since the model is neither specialized for a specific configuration nor individualized for a specific listener.

Experimental tests have been conducted in 2015 and 2019 with different rooms and audiences. The positive aspect is that the proposed method is competitive with the state-of-the-art ones. The

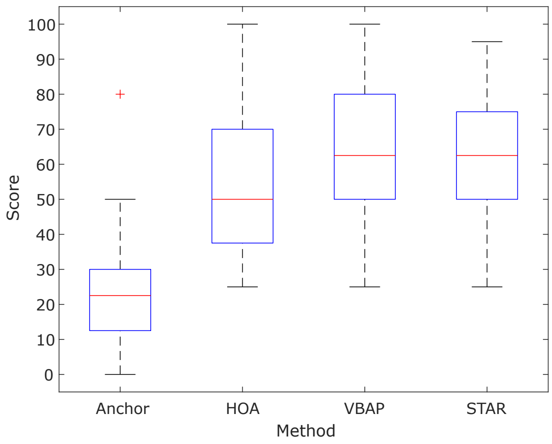


Figure 22: *Polyphonic test: perceived immersion.*

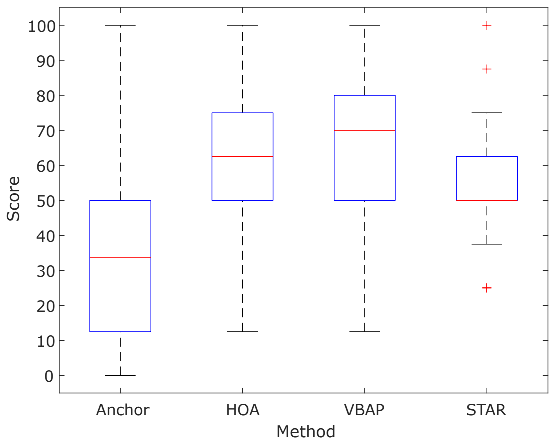


Figure 23: *Polyphonic test: perceived intelligibility.*

negative aspect is that the results are not really consistent between the 2015 and 2019 experiments. One explanation might be the fact that the anchor chosen is only spatial, with mono quality but full bandwidth, which might be too good for the non experts we had in 2019. We plan to re-conduct new tests with a low-pass filtered version, such as in standard MUSHRA tests. Another explanation is that between 2015 and 2019 we chose to place the azimuth reference at the center of the loudspeaker pair to improve the system determinant, because we suspected that this was the cause of the "nasal quality" reported by some listeners. This might be a bad choice, since the performance seems to degrade while the quality is not really improved (although it is quite good).

In the near future we plan to correct these issues, and re-conduct the experiments with a more calibrated loudspeaker configuration (a dome, that should favor HOA), and with expert listeners. Eventually, we will do some A/B testing such as in Marentakis et al. [6]. Finally, we will extend the method to distance and elevation, to generate a full 3D sound.

6. ACKNOWLEDGMENTS

The authors want to thank Aude Besnard, Anaëlle Marsollier, Clément Ruffini, and Aurianne Skybyk, the four Master's student that created the sound signals for the experimental tests and conducted them in 2015 in Brest, as well as Catherine Semal, psychoacoustician, who checked the validity of the test. Finally, thanks go to all the people who participated in the tests in 2015 and 2019.

7. REFERENCES

[1] Jens Blauert, *Spatial Hearing*, MIT Press, Cambridge, Massachusetts, revised edition, 1997, Translation by J. S. Allen.

[2] Ville Pulkki, "Virtual Sound Source Positioning using Vector Base Amplitude Panning," *Journal of the Acoustical Society of America*, vol. 45, no. 6, pp. 456–466, 1997.

[3] Michael A. Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[4] Jérôme Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Ph.D. thesis, Université Paris 6, 2001.

[5] A. J. Berkhout, "A Holographic Approach to Acoustic Control," *Journal of the Audio Engineering Society*, vol. 36, pp. 977–995, 1988.

[6] Georgios Marentakis, Franz Zotter, , and Matthias Frank, "Vector-Base and Ambisonic Amplitude Panning: A Comparison Using Pop, Classical, and Contemporary Spatial Music," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 945–955, 2014.

[7] Joan Mouba and Sylvain Marchand, "A Source Localization/Separation/Respatialization System Based on Unsupervised Classification of Interaural Cues," in *Proceedings DAFx'06*, Montreal, Quebec, Canada, September 2006, pp. 233–238.

[8] Joan Mouba, Sylvain Marchand, Boris Mansencal, and Jean-Michel Rivet, "RetroSpat: a Perception-Based System for Semi-Automatic Diffusion of Acousmatic Music," in *Proceedings of the Sound and Music Computing (SMC'08) Conference*, Berlin, Germany, July/August 2008, p. 33–40.

[9] Jeroen Breebaart, "Analysis and Synthesis of Binaural Parameters for Efficient 3D Audio Rendering in MPEG Surround," in *IEEE International Conference on Multimedia and Expo (ICME)*, Beijing, China, July 2007.

[10] John M. Chowning, "The Simulation of Moving Sound Sources," *Journal of the Acoustical Society of America*, vol. 19, no. 1, pp. 2–6, 1971.

[11] V. Ralph Algazi, Richard O. Duda, Dennis M. Thompson, and Carlos Avendano, "The CIPIC HRTF Database," in *Proceedings IEEE WASPAA*, New Paltz, New York, 2001, pp. 99–102.

[12] Harald Viste, *Binaural Localization and Separation Techniques*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 2004.

[13] "ITU-R BS.1116–3, Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," 2015.