Comparison of Transaural Configurations Inside Usual Rooms

Adrien Vidal, Philippe Herzog, Christophe Lambourg, Jacques Chatron

Comparison of Transaural Configurations Inside Usual Rooms

ADRIEN VIDAL, [1,2,3] **PHILIPPE HERZOG,** [1,4] *AES Member*,
(vidal@prism.cnrs.fr) (philippe.herzog@arteac-lab.fr)
CHRISTOPHE LAMBOURG, [2,4] *AES Member* **AND JACQUES CHATRON** [1]
(christophe.lambourg@arteac-lab.fr) (chatron@lma.cnrs-mrs.fr)

[1]*Aix Marseille University, CNRS, Centrale Marseille, LMA, Marseille, France*
[2]*ANSYS-OPTIS, Aix-En-Provence, France*
[3]*Aix Marseille University, CNRS, PRISM, Marseille, France*
[4]*ARTEAC-LAB SAS, Marseille, France*

This paper deals with the design of transaural systems in usual rooms, whose response has a strong influence on sound reproduction. The paper proposes to select configurations for the best perceptual rendering. However, realistic perceptive experiments cannot deal with the many possible room and loudspeaker configurations. Therefore, the authors propose to assess them using objective scores that are extrapolated from the results of perceptive tests assessing a suitable selection of rooms and loudspeaker configurations. This extrapolation then allows comparison of a much larger set of combinations, leading to the conclusion that close-to-ears configurations allow reduction of the room influence, leading to a good perceived fidelity—even inside usual rooms. Closer loudspeakers are, however, likely to be more sensitive to listener position, so the robustness of loudspeaker configurations to listener misplacement were investigated. A suitable objective score, again based on a perceptive test, led to the surprising conclusion that some close-to-ears configurations are also robust to listener position.

0 INTRODUCTION

During the last few decades, many 3D sound systems have been developed—leading to various technologies, differing in performances and complexity [1–3]. In this study, transaural systems, involving only two loudspeakers, are considered, seeking for best performances inside usual rooms. They have existed since the late 1960s, and various loudspeaker configurations were proposed [4–8], including systems with more than two loudspeakers [9–13]. Improving transaural systems is still a current issue: some recent studies aimed to improve the robustness to listener placement [14, 15] or reduce the timbre coloration [16, 17]. The present work focuses on the optimization of a transaural system inside usual rooms, i.e., with significant reflections on the walls. To the authors' knowledge, this aspect was marginally addressed in previous studies. Previous work from the authors [18] dealt with simulations of many configurations for transaural systems, assessing them through several objective indicators. In addition to usual free-field accuracy of the transaural decoding, these indicators attempted to take into account the effect of the diffuse field inside a usual room and listener placement relative to the sweet spot. Other previous works dealt with the influence of a transaural system equalization in several rooms, using recordings for listening tests [19, 20]. The main result from these previous works was that the distance between the loudspeakers and listener has a major influence: at short enough distances, the room influence might be significantly reduced.

The authors' aim is thus to check the previously published results in a way that is as realistic as possible. The comparison of numerous source configurations is, however, not an easy task: direct perceptive evaluation of a large panel of stimuli is barely possible, especially for stimuli that involve different rooms. Indeed, a large number of rooms may be considered though simulations: it has been shown that simulations may give quite realistic results, at least for speech signals [21]. Still, the perceptual comparison of numerous simulations is a difficult problem. The authors' approach is then to split the process into two steps: Perceptive evaluation is performed over a realistic number of configurations. An objective score is then determined and used for the evaluation of a much larger number of configurations. This process is described in Sec. 1.

An alternative assessment method is also considered, based on an objective criterion: the system "efficiency,"

*To whom correspondence should be addressed Tel: +33-4-9116-42-84; e-mail: vidal@prism.cnrs.fr

which may be computed directly from the geometry of the loudspeaker system [18, 20]. Although it is somewhat related to the optimization criterion proposed in an early study [10], it indirectly takes into account some room effect. It is described in Sec. 2. A perceptive evaluation is then performed considering a few configurations, and its results are compared to the two previous criteria in Sec. 3. Indeed, these results emphasize that loudspeakers should be relatively close to the listener.

Such configurations might, however, be less robust to listener misplacement. A suitable criterion is thus proposed, again as an objective score resulting from a limited number of perceptive evaluations. It is presented and discussed in Sec. 4, before a general conclusion about the considered system configurations proposed in Sec. 5.

1 ASSESSMENT OF ROOM INFLUENCE

As stated above, a perceptive experiment was conducted to build an objective indicator for assessing the room influence on transaural systems. The authors considered several rooms of various sizes and used two distances between the loudspeakers and listener.

Comparison of such configurations by physically moving the listener between rooms would have been very complicated and was thus not considered. Because the authors' aim is to build an objective score, they assumed that a comparison of recordings through headphones would be valid enough while leading to a much simpler series of test. Moreover, a diotic listening test was used for this first test, although the authors are interested in binaural perception. This was also chosen as a simpler mean to build an objective criterion, although it should be kept in mind that perceptive evaluation of timbre could be affected by spatial quality [22].

1.1 Perceptive Test Protocol

For this perceptive experiment, the MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) protocol was used [23] because it allows evaluation of a relatively high number of stimuli in a short period of time. The listener had to rate the similarity between an anechoic sound and the same sound reproduced in a listening room, both recorded and reproduced through headphones (monophonic recordings). Similarity scores ranged between 0 and 100 (100 meaning no perceived differences between sounds).

The recordings were performed using a Tannoy System 600 coaxial loudspeaker with blocked vents. This source was characterized in an anechoic chamber and equalized using a minimal phase Finite Impulse Response (FIR) filter in such a way that its frequency response was almost flat (± 1 dB) from 80 Hz to 13 kHz. Recordings were realized inside five "small" rooms of surface S and volume V: a Large Office ("LOf"), recording Studio ("Stu"), Medium Office ("MOf"), Small Office ("SOf"), and audiometric Cabin ("Cab"). Dimensions and Reverberation Time for the octave band centered on 1 kHz (RT_{1kHz}) of these rooms are reported in Table 1.

Table 1. Surface (S), Volume (V) and Reverberation Time for the octave band centered on 1 kHz (RT_{1kHz}) for each room used during the test on room influence.

	LOf	Stu	MOf	SOf	Cab
S (m^2)	19	18	16	8	4
V (m^3)	48	45	40	19	10
RT_{1kHz} (s)	0.71	0.16	0.66	0.93	0.04

Cab = audiometric Cabin; LOf = Large Office; MOf = Medium Office; SOf = Small Office; Stu = recording Studio.

During the recordings, the loudspeaker and microphone were placed on the room diagonal at 120 cm above the ground. The loudspeaker was placed at one third of this diagonal, and the microphone was 40 or 80 cm away from the loudspeaker—except for the audiometric cabin (too small for the 80-cm distance). Nine configurations are thus considered for the experiment. A single label designates a "configuration," i.e., a room and the measurement distance. For example, a measurement 40 cm away from the loudspeaker in the Studio is labeled "Stu40."

For each configuration, two measurements were realized: for the first one, the microphone was placed on axis in front of the loudspeaker and for the second one, the microphone was laterally shifted by 10 cm, keeping the measuring distance constant. Moreover, four equalization techniques were tested—they were presented in a previous paper showing little influence on the perceptual ratings of the different rooms [19]; they are thus not detailed here. For each configuration, ten recordings were thus available (four equalized and one non-equalized, each at two locations).

It was not considered possible to compare the 90 available recordings to each others. Even the comparison of the 18 non-equalized recordings was not an option. The authors decided to average over all possible recordings for a given configuration; it was therefore chosen to assess each configuration through a separate MUSHRA panel. To allow the comparison between configurations, three anchors were added to each panel: a high anchor (the hidden reference), mid anchor (non equalized sound of MOf80), and low anchor (non-equalized sound of SOf80, with microphone shifted by 20 cm). This required a post-processing, which is described in Sec. 1.2. Following this protocol, the listeners assessed nine series of 13 stimuli, rating at least one of them at 0 and another at 100.

The sound stimulus was a burst of pink noise with a total duration of 1 s. Stimuli were diffused through a Beyerdynamic DT990 Pro, with loudness equalized at 89 phons [24]. Twenty-two listeners took part to the experiment. Their auditions were not tested, but attributed responses to the hidden reference allowed to evaluate their listening abilities: according to standard [23], a listener should be excluded if he attributed the hidden reference a

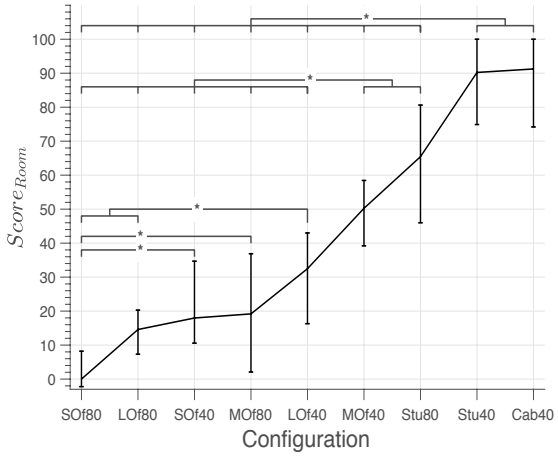

Fig.1.Median $Score_{Room}$ values, with respect to the configurations. * means a significant difference at 0.05 level. Error bars represent the interquartile range.

score lower than 90 for 15% of series. One listener was in this case: his answers have not been taken into account.

1.2 Perceptive Test Results

Comparison of the results from the nine MUSHRA tests involves an analysis of the scores for their three shared anchors. The high and low anchors have both been well identified: their mean scores were 99 and 2, respectively. For the mid anchor, the mean score was 27 but with wide variations with the configuration (mean score was 13 for Stu80 and 47 for LOf80). As could be expected, the assessment of this stimulus depends on the other stimuli for each configuration. A post-processing step has therefore been applied to all results. For each listener and configuration, a score transformation has been defined by two linear segments: the first one for scores between the low and mid anchors and the other one between the mid and high anchors. This transformed all the attributed score of the high, mid, and low anchors to their mean values, respectively 100, 27, and 0. Post-processed scores are denoted $Score_{Room}$ in the following; they correspond to the median values for all the test results for each configuration (21 listeners and ten recordings).

A repeated measures Analysis of Variance (ANOVA) using Statistica has been applied to the post-processed data, considering the factors "Configuration," "Equalization," and "Shifting." The ANOVA yielded a significant effect for all factors and all interactions at the level of 0.01. In this paper, the effect of the configuration is mainly investigated, and post-hoc tests were conducted applying the Bonferroni procedure at the significance level of 0.05. Results of tests and median $Score_{Room}$ are plotted in Fig. 1.

The median $Score_{Room}$ ranged over almost the entire scale from 0 to 91, meaning the effect of the configuration was quite significant. The Studio and Cabin were rooms with the shortest reverberation time, and were the one with the highest $Score_{Room}$. Effect of distance is also noticeable: for all room, $Score_{Room}$ attributed to a configuration at 40-cm distance were significantly superior to the one at 80 cm. The most important difference was for the Medium Office, for which the median $Score_{Room}$ at 40 cm was 50, whereas it was only 20 at 80 cm.

The listening position or the equalization have also been analyzed and showed much less influence on the test results. They are not shown here but are available in a previous publication [20]. The main result of this test is that the room and listening distance have a dominant influence on the perceived fidelity of reproduction systems. A room with good acoustic performances seems suitable for a wide range of listening distances (for example, the Studio has a median $Score_{Room}$ higher than 60 at both distances), whereas a shorter listening distance allows good transaural performances in usual rooms (such as Medium Office) and led to the best performance of the test in the small audiometric Cabin.

1.3 Predictive Score

The previous results allowed assessment of a reproduction system installed in a few existing rooms at two listening distances. The authors' goal is, however, to explore a much larger set of configurations. They therefore try to link these previous results with objective data, which could be determined for any room, already existing or not. Because the high anchor of the test is the anechoic response, it is expected that a good candidate criterion could deal with the importance of direct sound compared to the room response. Several room indicators can be computed from the impulse response (IR): Clarity C50 and C80, Definition D50, and Central Time Ct [25]. To encompass early reflection in the room response, two modified indicators may also be computed: C20 and D20, which are respectively the Clarity and Definition with integration time of 20 ms. These classical room indicators, however, require an IR, which could be difficult to determine at an early design step.

Because the authors want to define a simple criterion, the room effect may be roughly characterized as a diffuse field energy $E_{diffuse}$, which may be related to the direct field energy E_{direct} by the equivalent absorption area of the room walls [25]:

$$E_{diffuse} = \frac{16\pi r^2 E_{direct}}{A_{eq}}, \qquad (1)$$

where r is the distance to the source, E_{direct} is the energy of the source, and A_{eq} is the equivalent absorption area. Actual room effect also includes early reflections, especially on nearby walls. A few simulations assuming a single nearby plane (mirror source) did not significantly change the trends presented below, so the authors decided to stick with the simplest possible model. An alternative indicator can then be based on a simple model of the IR, allowing computation of an approximate Clarity for each octave band from 125 Hz to 8 kHz. Indeed, the IR may be modeled in two parts: the first one with a constant energy during a very short period of time t (corresponding to the direct energy) and the second one with decreasing energy related to room damping. This second part of the IR may be

Table 2. Correlation coefficient R and rank correlation coefficient ρ for each indicator.

	C_{20}	C_{50}	C_{80}	D_{20}	D_{50}	C_t	C_{mod}
R	0.97	0.96	0.95	0.88	0.87	–0.93	0.96
ρ_{Spear}	0.95	0.95	0.95	0.95	0.95	–0.85	0.93

modeled from the reverberation time needed for the energy decrease by 60 dB (RT_{60}). The IR model $ir(t)$ is thus defined as

$$\begin{cases} ir^2(t) = \frac{E_{Direct}}{\Delta t} & for\ t < \Delta t \\ ir^2(t) = K e^{\frac{-13.8}{RT_{60}} t} & for\ > \Delta t \end{cases}. \quad (2)$$

Because the considered criteria is based on energy ratio, E_{Direct} may be set to 1. Moreover t is arbitrarily set to 0.4 ms—a value that has little influence, chosen because it is much lower than 20 ms (this approximates a dirac-like direct response, without propagation delay). According to definition of $E_{diffuse}$ given by Eq. (1), K is obtained by

$$K = \frac{E_{Diffuse}}{\sum_{t=0}^{T_{max}} e^{\frac{-13.8}{RT_{60}} t}}. \quad (3)$$

T_{max} is the length of the modeled impulse response. It has also little influence if long enough, so the authors set it at $1.5RT_{60}$. A clarity C_{oct} is then computed for each octave band:

$$C_{oct} = 10log_{10} \frac{\sum_{t=0}^{0.02} ir^2(t)}{\sum_{t=20}^{T_{max}} ir^2(t)}. \quad (4)$$

A global C_{mod} is then defined as the mean of C_{oct} for octave bands from 125 kHz to 8 kHz.

To determine which objective indicator is the best suitable to describe perceptive scores, Bravais-Pearson correlation coefficient $R$and Spearman rank correlation coefficient ρ_{Spear} were computed between room indicators and scores attributed during the test. Correlation coefficients are presented in Table 2. All correlation coefficients are high, the best one being C_{20} with R = 0.97 and ρ_{Spear} = 0.95. C_{mod} got close values (R = 0.96 and ρ_{Spear} = 0.93), and was chosen because its computation only requires a very basic modelization of the room acoustics. The modelized score $\hat{Score}_{Room}$ is thus defined as an affine transform of C_{mod}, determined for best fit with listening test results:

$$\hat{Score}_{Room} = 3.78C_{mod} + 20.5. \quad (5)$$

Post-processed $Score_{Room}$ attributed by listeners are drawn with respect to computed values of $\hat{Score}_{Room}$ in Fig. 2. Their relation is close to linear: most predictive scores are included in the interquartile range of $Score_{Room}$, except extreme values. The proposed $\hat{Score}_{Room}$ score therefore gives a relatively good estimation of the median score $Score_{Room}$ attributed by listeners, with the advantage that it may easily be computed from basic room properties (RT_{60} reverberation times).

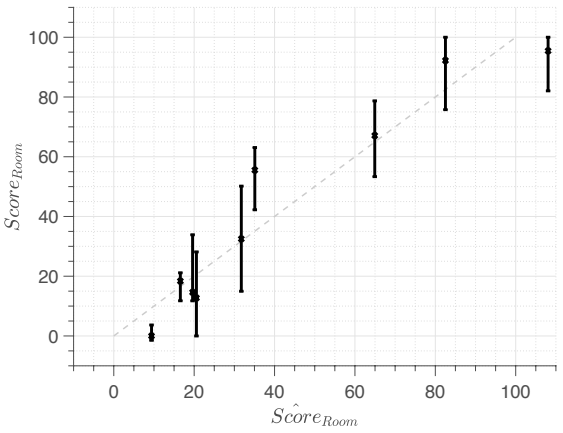

Fig. 2. Median $Score_{Room}$ with respect to $\hat{Score}_{Room}$.

Table 3. Reverberation Time RT_{60} and Equivalent absorption area A_{eq} for the medium size room and central frequency of octave bands from 125 Hz to 8 kHz.

f_c (Hz)	125	250	500	1,000	2,000	4,000	8,000
RT_{60} (s)	0.34	0.39	0.35	0.39	0.39	0.37	0.32
A_{eq} (m2)	19.3	16.8	18.6	16.8	16.7	17.9	20.3

1.4 Optimal System Configuration

The proposed $\hat{Score}_{Room}$ indicator is now used to compare the rendering of numerous transaural systems, in order to select the best-scored ones. This indicator is implicitly based on a diffuse to direct field ratio, which is estimated from the acoustic power radiated by the transaural system outside the listening area, in free-field. These simulations use the method presented in APPENDIX A.1.

For practical purposes, the pressure P_{rad} radiated by each reproduction system is computed over a sphere with radius 2 m centered on the listener's head (this distance is far enough from all loudspeakers positions studied here and coherent with the distances from the walls of listening rooms).

Discretizing the sphere in N_{sphere} = 1,000 almost equally spaced points, the quadratic mean pressure over the sphere P_{rad} is

$$P_{rad} = \sqrt{\frac{1}{N_{sphere}} \sum_{\theta_s, \phi_s}^{N_{sphere}} [P_1(r_s, \theta_s, \phi_s) + P_2(r_s, \theta_s, \phi_s)]^2}, \quad (6)$$

where P_1 and P_2 are the pressure from the two loudspeakers at the point of coordinate (r_s, θ_s, φ_s) of the sphere. This simple expression for the radiated pressure allows to estimate $E_{diffuse}$, using Eq. (1) with $E_{direct} = P_{rad2}$.

As an example, simulations are performed in the case of a medium-sized room of surface 16 m^2 and volume 41 m^3; its properties are described by the equivalent absorption surfaces A_{eq} [25], whose estimated values are reported in Table 3. Note that this room is not the MOf used during the perceptive test.

Simulation results at several distances in the horizontal plane are presented by Fig. 3. Without surprise, closer

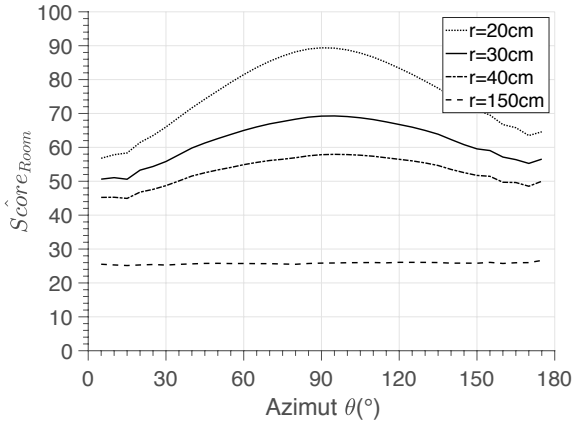

Fig. 3. $\hat{Score}_{Room}$ for $\phi = 0^\circ$ with respect to the azimuth (horizontal axis) and selected distances (line patterns).

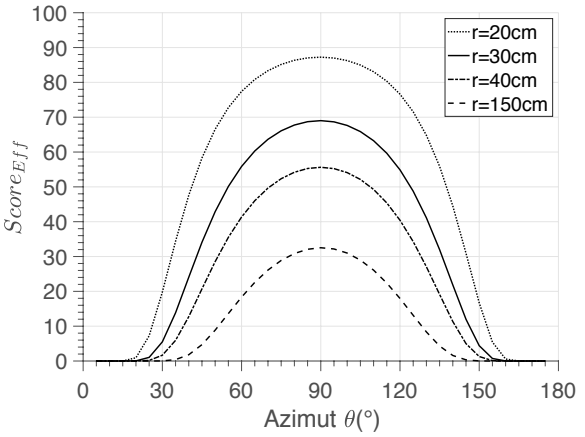

Fig. 4. $Score_{Eff}$ in the horizontal plane ($\varphi = 0^\circ$) with respect to the azimuth (horizontal axis) and selected distances (line patterns).

configurations lead to better scores because they favor the direct sound: the $\hat{Score}_{Room}$ is 89 at 20 cm at 90°, whereas it is always below 27 at 150 cm. The closer a configuration is, the higher the effect of azimuth is. At 150 cm, $\hat{Score}_{Room}$ are almost constant with azimuth, whereas at lower distance, widely spaced configurations get higher $\hat{Score}_{Room}$ than closely spaced configurations.

The effect of elevation is less significant and is not presented here. As an example, at a distance of 30 cm all $\hat{Score}_{Room}$ values are between 50 and 70, and its best value is reached for $\theta = 90^\circ$ in the horizontal plane.

2 SCORE FOR CONFIGURATION EFFICIENCY

In Sec.1.3, the room effect is estimated from simple room properties but also involves the far-field radiation pattern of the reproduction system. The present section investigates an alternative score based solely on the reproduction system, similarly to the "loss of dynamic range" criterion, which led to the "Optimal Source Distribution" [10]. This score is based solely on the reproduction system, implicitly modeled within an anechoic environment.

The underlying idea is that transaural filters cancel out part of the pressure radiated by each source. For a given sound level, they thus tend to increase the sources' stress. This increased stress may induce two effects that could corrupt the rendering at listeners' ears. First, the linear behavior of loudspeakers is limited to a drive level range, which could be exceeded. Second, usual loudspeakers generally have a wide directivity so increasing the loudspeaker drive level increases the pressure radiated all around the loudspeaker and thus the room influence at the listener's ears. For both reason, it is better to limit the sources' stress.

To evaluate the increase of sources' stress, transaural filters are again computed using the method presented in APPENDIX A.1. A target pressure $P = OUT_L$ at the left listener ear thus leads to drive signals S_L and S_R. They are compared to a monophonic drive signal S_m fed to both sources, when the same pressure $P = OUT_L$ results from the addition in module of the free-field contributions of the two sources (considered as monopoles).

An "efficiency ratio" R_{VV} is thus defined as

$$R_{VV} = < \frac{|S_L| + |S_R|}{2|S_m|} >_{freq} . \quad (7)$$

In Eq. (7), $<>_{freq}$ denotes an average value over frequency: the ratio is averaged over frequency bins of each octave band from 125 Hz to 8 kHz. The octave values are again averaged with a weight equal to the inverse of their central frequency squared. This weighting is a coarse mean to take into account the physical limitations of usual loudspeakers and damping trend of usual rooms.

For practical purposes, the efficiency ratio R_{VV} is computed in the case of the reproduction of sound toward a single ear (arbitrarily chosen as the left one, because this has no importance). This ratio is then converted into a score between 0 and 100, for which 100 is the best possible score, corresponding to $R_{VV} = 1$ (no increase in source stress). A simple law based on a Gaussian function is used so that the score decreases toward 0 when R_{VV} increases and reaches 0 when $R_{VV} = 1$:

$$Score_{Eff} = 100 e^{\frac{-(R_{VV}-1)^2}{2\sigma^2}} . \quad (8)$$

The choice of $\sigma = 0.5$ is arbitrary; it allows assignment of a low score value (≈13%) for an efficiency ratio of 2 (i.e., the required power is four times higher than without transaural processing).

$Score_{Eff}$ values may then be computed for various configurations. Fig. 4 shows results in the horizontal plane. Azimuth has a clear effect on $Score_{Eff}$: widely spaced configurations have higher $Score_{Eff}$ than closely spaced configurations. In the horizontal plane at 20-cm distance, $Score_{Eff}$ reaches 88 for $\theta = \pm 90^\circ$, whereas it is null for $\theta < \pm 20^\circ$ and for $\theta > \pm 160^\circ$. Distance has also a clear effect: closest configurations have the best $Score_{Eff}$. In the horizontal plane for a loudspeaker azimuthal span angle of $\theta = \pm 90^\circ$, $Score_{Eff}$ is 70 at 30 cm and 33 at 150 cm.

Elevation angle has also an effect on $Score_{Eff}$ values: non-elevated configurations get the best ones by far. As an example, for a 30-cm distance, $Score_{Eff}$ never exceeds 6 at $\varphi = 60^{\circ}$ elevation.

The authors' $Score_{Eff}$ indicator is linked to the one proposed for Optimal Source Distribution [10], in which widely spaced configurations were found to be better at lower frequencies. Because the $Score_{Eff}$ weighting is higher at lower frequencies, widely spaced configurations also get the best $Score_{Eff}$ values.

$Score_{Eff}$ variations shown in Fig. 4 also have similarities with those of $Scôre_{Room}$, shown in Fig. 3. Indeed, widely spaced and close-to-the-ears configurations get the highest scores, whereas both scores decrease when the distance increases and/or the loudspeaker azimuthal span angle decreases. For a loudspeaker azimuthal angle of 90°, score values are almost the same at both distances. More generally, ranking of configurations should be similar, using either of these two scores.

However $Score_{Eff}$ and $Scôre_{Room}$ values differ when the loudspeaker azimuthal span angle decreases: the curves' slope between minima and maxima is less pronounced for $Scôre_{Room}$ than for $Score_{Eff}$. Therefore, $Score_{Eff}$ values are much lower for closely spaced configurations. This is consistent with the high level of interference between left and right channels for such configurations, leading to much increased stress on the sources.

3 COMPARISON FOR SELECTED CONFIGURATIONS

Previous sections presented two objective scores to assess transaural source configurations. These scores are very basic and cannot fully assess all aspects of the reproduction system. Their validity must therefore be estimated for representative configurations. Some configurations have thus been selected and compared inside a medium-size room through a listening test. They are then compared using the proposed scores.

During the test, each listener was sitting within a frame supporting various loudspeakers. This setup was installed in the middle of the medium-size room (see Fig. 5). The seat was adjusted in order to put the listener's ears at the right positions. A screen and mouse allowed dialing the test sliders. Because all speaker systems were simultaneously available, this setup thus allowed switching quickly between them, leading to efficient comparisons and a reasonable test duration (about half an hour for each listener).

3.1 Selection of Configurations

Seven configurations were selected for comparison. Among "standard" configurations, the authors selected two at a usual distance (150 cm): the classical stereo configuration ($\theta = \pm 30^{\circ}$) labeled "FarEquiEars" (FEE) and a narrow stereo dipole ($\theta = \pm 2^{\circ}$) labeled "FarDipoleEars" (FDE). The authors added an elevated nearfield stereo dipole, labeled "NearDipoleUp" (NDU), studied recently [11].

Fig. 5. Experimental setup. On the right of the seat, a mouse allows usage of the test interface displayed on the screen placed in front of the listener.

Table 4. Loudspeakers position and scores of selected configurations.

Name	r (cm)	θ (°)	φ (°)	$Scôre_{Room}$	$Score_{Eff}$
FEE	150	30	0	25	0
FDE	150	2	0	26	0
NDU	32	7	60	53	0
NFM	34	57.5	15	59	42
NMH	31	87.5	30	65	54
NME	32	102.5	0	66	64
NRH	32	137.5	30	59	12

FDE = FarDipoleEars; FEE = FarEquiEars; NDU = NearDipoleUp; NFM = NearFrontMid; NME = NearMidEars; NMH = NearMidHigh; NRH = NearRearHigh.

Four new configurations at short distance involved increasing angles between sources, going around the listener: "NearFrontMid" (NFM), "NearMidHigh" (NMH), "NearMidEars" (NME), and "NearRearHigh" (NRH). The seven configurations are briefly described in Table 4 with the objective scores expected from their geometry and the acoustic properties of the medium-sized office where they were characterized (see Table 3).

Five configurations involved sources at about 30 cm from the listener head center, because it is the shortest distance the authors consider acceptable for the listener. They involved small homemade loudspeakers (Visaton FRS 8M drivers loaded by a closed box). The other configurations (at 150-cm distance) involved more powerful sources (MeyerSound MMX4P amplified speakers) to avoid overloading them and thus biasing the test.

3.2 Protocol

These seven configurations were tuned as described by APPENDIX A.1 and used for comparisons with reference sources, inside the medium-size room described in Table 3. This way, it was possible to switch quickly between any transaural system and any reference source.

Indeed, the reference sources were loudspeakers placed in the same room. Three references were used: Ref1 (r = 90 cm, θ = 80° left, φ = –13°), Ref2 (r = 250 cm, θ = 9° right, φ = 11°), and Ref3 (r = 150 cm, θ = 30° right, φ= 0°). Loudspeakers used for these references were MeyerSound MMX4P. A photo of the setup is presented in Fig. 5, showing the test configuration and speakers.

Binaural impulse responses of these references were measured for each listeners using binaural microphones (DPA 4060). Targets for the transaural systems were then obtained by convolution of the stimuli with these impulse responses.

Three stimuli were used for this experiment: the same noise burst as the one used in Sec. 1, a male voice saying the French sentence "Le coq réveille le village," and a short music excerpt with drums and bass.

As for the previous test, a MUSHRA protocol [23] was chosen for this test. For each MUSHRA test, the listeners had to assess the similarity between a reference sound source and its rendering through seven transaural systems. This comparison was performed for the three stimuli and three references. Twenty-two listeners took part to the experiment, and none were excluded.

Listeners thus assessed nine series of eight sounds, all series dealing withthe same candidate systems: their results could thus be averaged, leading to a global score for each transaural system.

3.3 Results

Attributed scores are denoted $Score_{Overall}$ in the following. The median $Score_{Overall}$ for the hidden reference was 100 with a null interquartile range, meaning that listeners were all able to identify it.

The median $Score_{Overall}$ of the FDE system was 0 with a null interquartile range, meaning that this system was unanimously considered the most different from the reference (implicit low anchor). $Score_{Overall}$ attributed to the reference and FDE were thus not taken into account for the following analysis.

A repeated-measures ANOVA was conducted considering the factors "Configuration," "Virtual Source Incidence," and "Stimulus." The ANOVA yielded a significant effect of the Configuration [$F(5,105) = 151.56$, $p < 0.001$] and Stimulus [$F(2,42) = 17.39$, $p < 0.001$] but no effect of the Virtual Source Incidence [$F(2,42) = 0.62$, $p = 0.54$]. Moreover, the interaction of the Configuration and Stimulus was significant [$F(10,210) = 3.13$, $p < 0.001$], and so was the interaction of the Configuration and Virtual Source Incidence [$F(10,210) = 3.57$, $p < 0.001$].

Post-hoc tests were applied using the Bonferroni procedure at the significance level of 0.05. Results of these post-hoc tests and median $Score_{Overall}$ with respect to the configuration are shown by Fig. 6.

$Score_{Overall}$ of configuration FEE is significantly lower than the five other configurations, with a median $Score_{Overall}$ of 20. The three configurations NFM, NMH, and NME scores are not significantly different from each other but are significantly higher than the three other configurations.

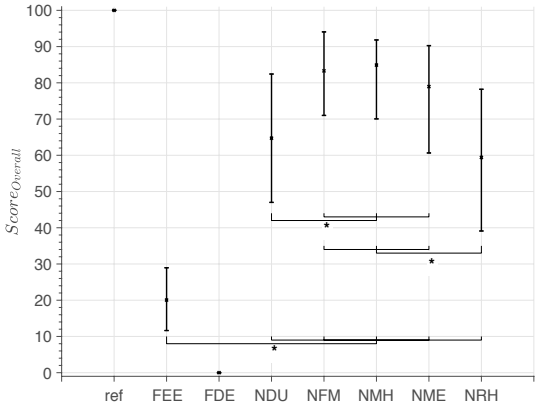

Fig. 6. Median $Score_{Overall}$ with respect to the configuration. * means a significant difference at 0.05 level. Error bars represent the interquartile range. FDE = FarDipoleEars; FEE = FarEquiEars; NDU = NearDipoleUp; NFM = NearFrontMid; NME = NearMidEars; NMH = NearMidHigh; NRH = NearRearHigh.

Especially, median scores of NMH reach 85. Configurations NDU and NRH get intermediate $Score_{Overall}$ values (65 and 59, respectively).

Post-hoc tests also revealed that stimuli are significantly differentiated. A high median value (79) of $Score_{Overall}$ is reached for the speech stimuli, whereas the music excerpt gets 69 and the burst noise only 60. The burst noise is thus the most discriminant stimulus: this explains that it was used to assess differences during the preliminary test.

Conversely, the speech stimulus is the least demanding one, especially for the NFM configuration: its median $Score_{Overall}$ is 90 for this system. According to the MUSHRA recommendation [23], it means that most of the listeners were not able to distinguish the rendering of speech reproduced by a physical loudspeaker and its reproduction through the NFM system.

Further analysis also shows that $Score_{Overall}$ is higher when the reference source is in the same direction as one of the loudspeakers of the tested configuration. The choice of the reference sources' positions may thus have slightly biased the system comparison, although the reference sources were fairly even distributed in front of the listener. The authors could not test a larger number of reference sources to keep a reasonable test duration; especially, the authors did not consider rear reference sources because their rendition using binaural technology is generally considered better than for front source [26].

3.4 Comparison With Objective Scores

Results of the final perceptive evaluation may be compared to the two objective scores $\hat{Score}_{Room}$ and $Score_{Eff}$ proposed in Secs. 1 and 2. A third one $\hat{Score}_{Predict}$ is also built as their linear combination, with coefficients obtained for best fit with $Score_{Overall}$:

$$\hat{Score}_{predict} = 0.36 Score_{Eff} + 0.97 \hat{Score}_{Room}. \quad (9)$$

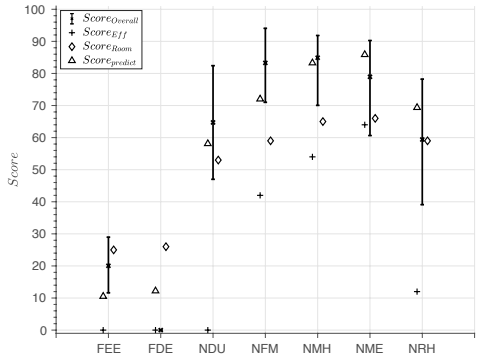

Fig. 7. Median $Score_{Overall}$, $Score_{Eff}$, $\widehat{Score}_{Room}$, $\widehat{Score}_{predict}$ with respect to the configuration. FDE = FarDipoleEars; FEE = FarEquiEars; NDU = NearDipoleUp; NFM = NearFrontMid; NME = NearMidEars; NMH = NearMidHigh; NRH = NearRearHigh.

Fig. 7 represents scores of the perceptive test ($Score_{Overall}$) and the three above-mentioned objective scores. It highlights that none of the individual scores are able to fully predict the perceptive results:

- $\widehat{Score}_{Room}$ is often close to $Score_{Overall}$ but does not discriminate enough configurations NRH, NDU, NME NFM, and NMH (values between 53 and 66, whereas they are between 59 and 85 for $Score_{Overall}$).
- $Score_{Eff}$ is systematically lower than $Score_{Overall}$, especially for configurations NRH and NDU. Tuning the value of σ in Eq. (8) might slightly improve the similarity with the higher values of $Score_{Overall}$.
- $\widehat{Score}_{Predict}$ is relatively close to $Score_{Overall}$, almost always within its interquartile range—except for the lowest scores (FDE and FEE).

The main contributor of $\widehat{Score}_{predict}$ is clearly $\widehat{Score}_{Room}$, as shown by its near-unity coefficient. $Score_{Eff}$ appears to add a smaller but significant corrective factor. Remaining discrepancies are of the same order of magnitude as the uncertainties on the perceptive results, so the simple scores proposed seem sufficient for a first analysis.

Objective scores and perceptive evaluation both indicate that configurations close-to-ears (widely spaced and short distance) lead to better transaural reproduction. These configurations lead to a low cross-talk between ears, and thus, simple transaural filters involving minimal cancellation between channels. This reduces the stress on the sources and limits the room influence.

4 LISTENER MISPLACEMENT

Although close-to-ears configurations seem interesting to deal with the room influence, one may expect that the short distance between the sources and listener ears leads to a lack of robustness to listener movements. This is now investigated, again through a listening test allowing for definition of an objective criterion.

4.1 Perceptive Test Protocol

A major difficulty when perceptively assessing the influence of listener movements is that they cannot be reliably controlled without the listener knowledge. Moreover, there are many possible movements; testing all situations would lead to an unrealistic test protocol.

Controlled static misplacements were therefore simulated, again using the method described in APPENDIX A.1 and considering only two situations: a lateral shift toward the left (+5 cm *y*) and a shift toward the front (+5 cm *x*). These shifts were chosen from preliminary listening tests because they seemed to have the highest influence on binaural rendering. These simulations allowed assessment of the perceived differences through headphones (Beyerdynamic DT990 Pro), again using a MUSHRA protocol to compare a binaural sound and its reproduction through a transaural system with the listener position shifted.

The stimuli were the same burst of pink noise than for the first test (Sec. 1), simulated as monophonic and static sources. Change in sound loudness corresponding to a misplacement was considered part of this misplacement effect and was thus not compensated.

Three loudspeaker distances (20, 40, and 80 cm) and four loudspeaker azimuthal span angles (±5°, ±30°, ±60°, and ±90°) were combined, leading to 12 system configurations. For each configuration, three virtual sources were simulated, characterized by different incidences: 0°, 45°, and 90°. Both transaural systems and virtual sources were placed in the horizontal plane. Lastly, for each virtual source, two listener misplacements were simulated. This resulted in a large number of configurations (72), requiring multiple MUSHRA tests like in Sec. 1.

The authors chose to associate each MUSHRA test with a virtual source incidence and a loudspeaker azimuthal span angle. Each test thus compared three loudspeaker distances and two listener shifts. Four anchors were added to each tests: a hidden reference, high anchor (monophonic sound of the reference reproduced in diotic conditions with loudness corrected to match the virtual source at 0° incidence), mid anchor (non-equalized on-axis sound in Stu80 room), and low anchor (non-equalized on-axis sound in MOf80 room). The two last anchors (also diotic) were common with the perceptive test of Sec. 1.

There were thus 12 series of ten stimuli each, each reference being the binaural sound of the virtual source, for a well-placed listener. Eighteen listeners took part to the experiment, but one was discarded because he did not correctly identify the hidden reference.

4.2 Perceptive Test Results

The scores granted by the listeners, denoted by $Score_{Mispl}$, were respectively 100 and 0 for the hidden reference and low anchor, with a null interquartile range. Both were clearly identified.

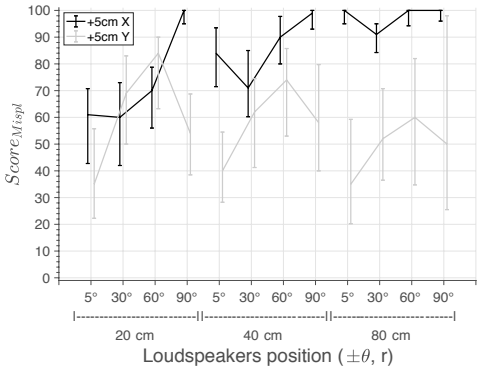

Fig. 8. Median $Score_{Mispl}$ attributed during the test on sensitivity to listener placement, with respect to the loudspeakers' position (horizontal axis) and for the two misplacements (line patterns). * means a significant difference at 0.05 level. Error bars represent the interquartile range.

The high anchor got a median $Score_{Mispl}$ of 77.5 with a large interquartile range (45). This resulted from the multiple virtual source incidences; indeed, median $Score_{Mispl}$ was 100 for the 0° incidence (diotic sound perceived as a front source), whereas it was 55.5 for the 90° incidence.

The intermediate anchor got a median $Score_{Mispl}$ of 19, which was quite a low value. This stimulus is shared with the test of Sec. 1 for which it got a $Score_{Room}$ of 67; $Score_{Mispl}$ seems more sensitive than $Score_{Room}$. However, note that this diotic anchor was compared with diotic stimuli in Sec. 1, whereas it is now compared with transaural stimuli.

A repeated-measures ANOVA using Statistica has been applied on $Score_{Mispl}$ for all sounds except the anchors, considering the factors "Loudspeakers distance," "Loudspeakers angle," "Virtual source incidence," and "Misplacement." All of these factors and two-by-two interactions had significant effect on result at the level of 0.01.

Median $Score_{Mispl}$ values are shown by Fig. 8, with respect to the loudspeakers' positions and for the two misplacements. The virtual source influence is not detailed because it seems less significant.

$Score_{Mispl}$ values for frontal shifts (+5 cm *x*) are always higher than the ones for lateral shifts (+5 cm *y*), except for two configurations (30° and 60° at 20 cm). For lateral shifts(+5 cm *y*), $Score_{Mispl}$ values increase with loudspeaker azimuthal span angle until 60° but then decrease for 90°. For frontal shifts (+5 cm *x*), $Score_{Mispl}$ values increase with loudspeaker azimuthal span angles higher than 5°.

Configurations at 80 cm are very robust to a frontal shift (+5 cm *x*) but not to a lateral one (+5 cm *y*). Moreover, the interquartile range for the lateral shift is very large (this results from the virtual source incidence being or not being coincident with physical sources).

Conversely, configurations at 20 cm exhibit much less difference between the shift direction, except for the 90° configuration (high $Score_{Mispl}$ value for the +5 cm *x* misplacement and medium $Score_{Mispl}$ value for the +5 cm *y* misplacement).

Globally, close-to-ears configurations do not seem to be particularly sensitive to listener shifts. Especially, configurations with speakers at 20 and 40 cm and ±60° loudspeaker azimuthal span angle seem quite robust to the two misplacements tested here: median $Score_{Mispl}$ values exceed 70, despite the high sensitivity of this criterion.

4.3 Objective Score

Following the same approach as in Sec. 1, the authors now propose an objective score $\hat{Score}_{Mispl}$ allowing assessment of a much larger number of configurations than the one used for the listening test. As a first step, the authors considered combinations of usual objective criterions, based on remarks collected from the listeners at the end of the test; they mentioned differences in level, timbre, and localization, which should thus be assessed. Seven objective indicators were considered: one indicator for binaural loudness dissimilarity [27], four indicators for timbre dissimilarity [28, 29], and two indicators for localization dissimilarity based on interaural time difference or interaural level difference estimations.

All linear combinations of these seven indicators were fitted to the listening test results, and the best candidate was a combination of the binaural loudness difference and the total difference of specific loudness [28]. The correlation coefficient between $Score_{Mispl}$ and this simple two-terms combination reached 0.94; it could not be significantly improved considering more complex combinations.

The next step is to build an objective score $\hat{Score}_{Mispl}$ based only on frequency response simulations, instead of using sounds excerpts as the ones listened to during the tests. To approximate the test listening conditions, the responses are filtered by a –3-dB–octave slope and weighted by a B weighting (because the reproduced level was 89 phons). A binaural level *Niv* may be computed as [27]

$$Niv = 3.5 log_2(2^{\frac{Niv_L}{g}} + 2^{\frac{Niv_R}{g}}), \quad (10)$$

where Niv_L and Niv_R are the RMS levels, expressed in decibels (unscaled), of the left and right responses after filtering and weighting. The level dissimilarity D_{level} is then defined as

$$D_{level} = |Niv_{ref} - Niv_{sig}|, \quad (11)$$

where the *ref* subscript corresponds to the simulation with the head centered and the *sig* subscript corresponds to the simulations with a head misplacement.

Similarly, a timbre difference D_{timbre} is defined from third-octave deviations:

$$D_{timbre} = \sum_{i=1}^{i_{max}} |[Niv_{ref}(i) - Niv_{ref}] - [Niv_{sig}(i) - Niv_{sig}]|, \quad (12)$$

where index *i* designates a third octave band and the value without index is the average over all bands. Computation is performed over frequency bands from 100 Hz to 12.5 kHz. Finally, the proposed objective score $\hat{Score}_{Mispl}$ is

$$\hat{Score}_{Mispl} = 100 - 5.78 D_{level} - 1.24 D_{timbre}. \quad (13)$$

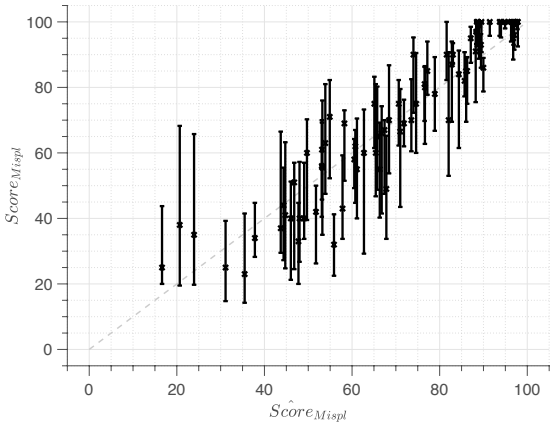

Fig. 9. Median $Score_{Mispl}$ with respect to $\hat{Score}_{Mispl}$.

The correlation coefficient R between the score $Score_{Mispl}$ resulting from the listening test and proposed objective score $\hat{Score}_{Mispl}$ is again found to be 0.94. Fig. 9 represents median post-processed scores $Score_{Mispl}$ attributed by listeners, with respect to the values of $\hat{Score}_{Mispl}$ computed from simulations. Relation between them seems indeed linear, because most $\hat{Score}_{Mispl}$ values are within in the interquartile range of $Score_{Mispl}$.

4.4 Robustness of System Configurations

The proposed objective score $\hat{Score}_{Mispl}$ is now used to assess the robustness of the same set of configurations as for the previous indicators. Moreover, it is simulated for 12 misplacements and three virtual source incidences. The 12 misplacements consist of two head rotations (±10° around the z axis), six axial translations (±5 cm along x, y, and z axes), and four diagonal translations (±5 cm along diagonals of the x and y axes). The virtual source incidences are the same as for the perceptive evaluation (0°, 45°, and 90°). $\hat{Score}_{Mispl}$ values are computed for each misplacement and virtual source incidence, but only the worst value (lowest score) is kept.

Fig. 10 shows the simulation results for configurations in the horizontal plane. There is no major difference between distances 20, 30, and 40 cm, but at 150 cm, $\hat{Score}_{Mispl}$ is lower than at shorter distances. The azimuth effect is more pronounced at 150 cm than at other distances, with two maxima around 35° and120°. From 20 to 40cm, $\hat{Score}_{Mispl}$ is almost constant between 30° and 150° with values between 50 and 75. Whatever the distance, $\hat{Score}_{Mispl}$ is very low for closely spaced configuration ($\hat{Score}_{Mispl}$ < 30 for θ < 15° and θ > 165°). This somewhat contradicts previous results that found the stereo dipole robust to listener head movements [30, 10]. This may result from the natural contrast obtained for wider source angles.

Results for various elevations are shown by Fig. 11. For closely spaced configuration (θ = 5°, and θ = 175°), elevated configurations (φ = 45°, and φ = 60°) get high $\hat{Score}_{Mispl}$ (>70). For spaced configurations, elevation has no significant effect on results, except at φ= 60°, for which $\hat{Score}_{Mispl}$ is lower than for other elevations.

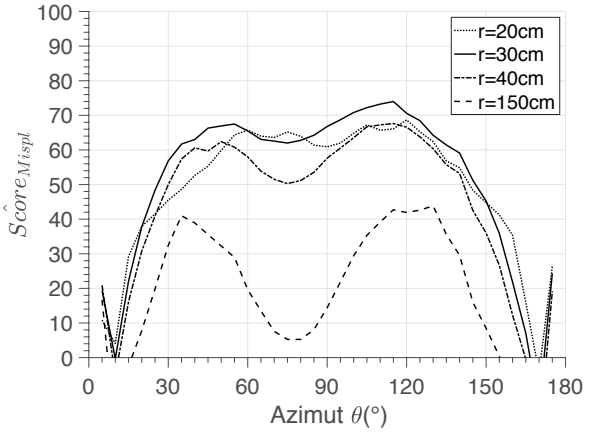

Fig. 10. $\hat{Score}_{Mispl}$ for ϕ= 0deg with respect to the azimuth (horizontal axis) and selected distances (line patterns).

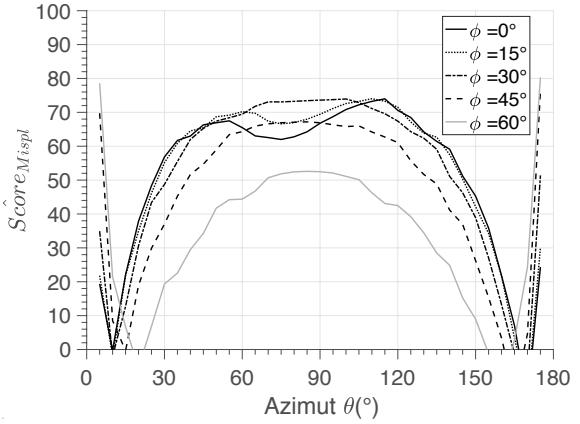

Fig. 11. $\hat{Score}_{Mispl}$ at r = 30 cm with respect to the azimuth (horizontal axis) and selected distances (line pattern).

This result is in accordance with [11], in which elevated stereo dipoles were considered to be more robust to the listener misplacement.

Globally, these exhaustive simulations confirm the results from the listening test: close-to-ears configurations are quite robust to listener misplacement, at least for the distances considered here.

Note that the low values of $\hat{Score}_{Mispl}$ computed for configurations at usual distances may be misleading: this score mimics the results of the listening test, which was shown to be significantly more sensitive than $Score_{Room}$. Part of the difference might result from the use of diotic recordings (which emphasizes the timbre changes) for the $Score_{Room}$ test, but this artifact is unlikely to switch the results. It thus seems that the robustness to listener misplacement is less a concern than the room effect, according to listeners' evaluation.

5 CONCLUSION

This paper presents a method to assess various transaural configurations inside usual rooms. Because a direct comparison of multiple configurations is not possible, the authors used listening tests over headphones to assess recordings of a few representative configurations. Objective scores are then established from these perceptive results and applied to a much larger set of configurations.

Two different objective scores have been proposed, as a simple mean to assess the perceived quality of any transaural system. A linear combination of these two scores fits very well with the result of perceptive tests, allowing a quantitative comparison of loudspeaker configurations at a design stage.

Assessment of a large number of system configurations then led to the conclusion that close-to-ears configurations permit a very realistic reproduction in a usual room like a medium-sized office. Indeed, most listeners were not able to distinguish the rendering of speech through a real loudspeaker and its reproduction through the NFM system, even for a frontal incidence (which is usually problematic in binaural reproduction). It is, of course, not surprising that close-to-ears configurations reduce the room influence. The main interest of the proposed quantitative score is that it allows to search for a trade-off between room and system designs.

Especially, a close-to-ears configuration can be installed inside a small room with limited acoustic treatment. Such a facility is much cheaper than a large room with a specific acoustic treatment as recommended by standards [31, 32].

The authors expected, however, that such close-to-ears configurations would not be robust to listener misplacement. A third objective score is thus proposed, defined from a specific listening test. It allows assessment of the robustness to listener misplacement for many system configurations, an important criterion for system design.

An unexpected output of this work is that close-to-ears configurations can also be quite robust, even more than distant ones. The considered loudspeaker distances allow a good sound restitution, even for reasonable listener misplacements.

The proposed objective scores allow comparison of various configurations and rooms, based on quick simulations that can be performed at the design stage of a room or system. However, the proposed method is based on the extrapolation of listening test results and may thus be biased. Future work should therefore compare the proposed objective scores to perceptive tests for many different configurations, in different rooms. Other 3D sound systems could also be compared after suitable adaptation of the method.

6 ACKNOWLEDGMENT

The authors would like to thank Patrick Boussard for initiating this work.

7 REFERENCES

[1] H. Møller, "Fundamentals of Binaural Technology," *Appl. Acoust.*, vol. 36, no. 3–4, pp. 171–218 (1992 Mar.). https://doi.org/10.1016/0003-682X(92)90046-U.

[2] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466 (1997 Jun.).

[3] J. Daniel, S. Moreau, and R. Nicol, "Further Investigations of High-Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," presented at the *114th Convention of the Audio Engineering Society* (2003 Mar.), paper 5788.

[4] M. Schroeder, "Digital Simulation of Sound Transmission in Reverberant Spaces," *J. Acoust. Soc. Am.*, vol. 47, no. 2, pp. 424–431 (1970 Feb.). https://doi.org/10.1121/1.1911541.

[5] P. Damaske, "Head-Related Two-Channel Stereophony With Loudspeaker Reproduction," *J. Acoust. Soc. Am.*, vol. 50, no. 4B, pp. 1109–1115 (1971 Oct.). https://doi.org/10.1121/1.1912742.

[6] O. Kirkeby, P. A. Nelson, and H. Hamada, "The 'Stereo Dipole'—A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers," *J. Audio Eng. Soc.*, vol. 46, no. 5, pp. 387–395 (1998 May).

[7] M. R. Bai and C.-C. Lee, "Objective and Subjective Analysis of Effects of Listening Angle on Crosstalk Cancellation in Spatial Sound Reproduction," *J. Acoust. Soc. Am.*, vol. 120, no. 4, pp. 1976–1989 (2006 Oct.). https://doi.org/10.1121/1.2257986.

[8] F. Kaiser, *Transaural Audio – The Reproduction of Binaural Signals Over Loudspeakers*, Diploma thesis, Graz University, Graz, Austria (2011 Mar.).

[9] J. Bauck and D. H. Cooper, "Generalized Transaural Stereo and Applications," *J. Audio Eng. Soc.*, vol. 44, no. 9, pp. 683–705 (1996 Sep.).

[10] T. Takeuchi and P. A. Nelson, "Optimal Source Distribution for Binaural Synthesis Over Loudspeakers," *J. Acoust. Soc. Am.*, vol. 112, no. 6, pp. 2786–2797 (2002 Dec.). https://doi.org/10.1121/1.1513363.

[11] Y. L. Parodi and P. Rubak, "Objective Evaluation of the Sweet Spot Size in Spatial Sound Reproduction Using Elevated Loudspeakers," *J. Acoust. Soc. Am.*, vol. 128, no. 3, pp. 1045–1055 (2010 Sep.). https://doi.org/10.1121/1.3467763.

[12] M. F. Simon Galvez, M. Blanco Galindo, and F. M. Fazi, "A Study on the Effect of Reflections and Reverberation for Low-Channel-Count Transaural Systems," in *Proceedings of the International Congress and Exposition on Noise Control Engineering (InterNoise)*, paper1945 (Madrid, Spain) (2019 Jun.).

[13] X. Ma, C. Hohnerlein, and J. Ahrens, "Concept and Perceptual Validation of Listener-Position Adaptive Superdirective Crosstalk Cancellation Using a Linear Loudspeaker Array," *J. Audio Eng. Soc.*, vol. 67, no. 11, pp. 871–881 (2019 Nov.).

[14] R. Matsuda, M. Otani, and H. Okumura, "Evaluation of Robustness of Dynamic Crosstalk Cancellation for Binaural Reproduction," in *Proceedings*

of the AES International Conference on Spatial Reproduction - Aesthetics and Science (2018 Jul.), paper P1-1.

[15] M.F. Simon Galvez, E.Hamdan, D.Menzies, and F.M. Fazi, "A Study of the Effect of Head Rotation on Cross Talk Cancellation," presented at the *145th Convention of the Audio Engineering Society* (2019 Oct.), paper 10125.

[16] K. Young, G. Kearney, and A. I. Tew, "Loudspeaker Positions With Sufficient Natural Channel Separation for Binaural Reproduction," in *Proceedings of the AES International Conference on Spatial Reproduction - Aesthetics and Science* (2018 Jul.), e-Brief 71.

[17] L. Liu and B. Xie, "A High-Frequency-Band Timbre Equalization Method for Transaural Reproduction With Two Frontal Loudspeakers, "presented at the *148th Convention of the Audio Engineering Society* (2020 May), paper 10331.

[18] A. Vidal, P. Herzog, C. Lambourg, P. Boussard, and L. Husnik, "Binaural Rendering Using Near-Field Loudspeakers," in *Proceedings of the 3rd International Conference on Spatial Audio*, pp. 1–6 (Graz, Austria) (2015 Sep.).

[19] A. Vidal, P. Herzog, and C. Lambourg, "Comparaison de Methodes d'Egalisation pour une Restitution en Salle d'Ecoute," in′ *Proceedings of the Congres Francais d'Acoustique*, pp. 201–207 (Le Mans, France) (2016 Apr.).

[20] A. Vidal, *Diffusion de son 3D par Synthese de Champs Acoustiques Binauraux*, Ph.D. thesis, Aix Marseille University, Marseille, France (2017 Feb.).

[21] M. Blau, A. Budnik, M. Fallahil, et al., "Toward Realistic Binaural Auralizations - Perceptual Comparison Between Measurement and Simulation-Based Auralizations and the Real Room for a Classroom Scenario," *Acta Acust. united Acust.*, vol. 5, paper 8 (2021 Jan.). https://doi.org/10.1051/aacus/2020034.

[22] S. Le Bagousse, M. Paquier, S. Moulin, and C. Colomes, "Determination of a Relevant Spatial Anchor for Audio Quality Evaluation of Codecs," *Acta Acust. united Acust.*, vol. 102, no. 2, pp. 383–388 (2016 Mar./Apr.). https://doi.org/10.3813/AAA.918954.

[23] ITU-R, "Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems, "*Recommendation ITU-R BS.1534-3* (2015 Oct.).

[24] ISO, "Acoustique — Methode de Calcul du Niveau D'isosonie, "*ISO Standard 532B : 1975 Standard 532B:1975* (1975 Jul.).

[25] A. C. Gade, "Acoustics in Halls for Speech and Music, " in T. D. Rossing (Ed.), *Springer Handbook of Acoustics*, Springer Handbooks, pp. 301–349 (Springer, New York, NY, 2007).

[26] F. Volk, F. Heinemann, and H. Fastl, "Externaliza-¨ tion in Binaural Synthesis: Effects of Recording Environment and Measurement Procedure," *J. Acoust. Soc. Am.*, vol. 123, no. 5, p. 3935 (2008 May).

[27] V. P. Sivonen and W. Ellermeier, "Binaural Loudness for Artificial-Head Measurements in Directional Sound Fields," *J. Audio Eng. Soc.*, vol. 56, no. 6, pp. 452–461 (2008 Jun.).

[28] P.-Y. Michaud, M. Lavandier, S. Meunier, and P. Herzog, "Objective Characterization of Perceptual Dimensions Underlying the Sound Reproduction of 37 Single Loudspeakers in a Room," *Acta Acust. united Acust.*, vol. 101, no. 3, pp. 603–615 (2015 May/Jun.). https://doi.org/10.3813/AAA.918856.

[29] G. Peeters, "A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project," http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf (2004 Apr.).

[30] J. J. Lopez, A. Gonzalez, and F. Orduna, "Modeling and Measurement of Cross-Talk Cancellation Zones for Small Displacements of the Listener in Transaural Sound Reproduction With Different Loudspeaker Arrangements," presented at the *109th Convention of the Audio Engineering Society* (2000 Sep.), paper 5267.

[31] AES, "AES Recommended Practice for Professional Audio - Subjective Evaluation of Loudspeakers," *AES Standard AES20-1996* (1996 Jan.).

[32] ITU-R, "Methods for the Subjective Assessment of Small Impairments in Audio Systems," *Recommendation ITU-R BS.1116-3* (2015 Feb.).

[33] R. V. L. Hartley and T. C. Fry, "The Binaural Localization of Pure Tones," *Phys. Rev.*, vol. 18, no. 6, pp. 431– 442 (1921 Dec.).
https://doi.org/10.1103/PhysRev.18.431.

[34] R. O. Duda and W. L. Martens, "Range Dependence of the Response of a Spherical Head Model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058 (1998 Nov.). https://doi.org/10.1121/1.423886.

[35] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz, "Auditory Localization of Nearby Sources. II. Localization of a Brandband Source," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1956–1968 (1999 Oct.). https://doi.org/10.1121/1.427943.

[36] R. V. Algazi, C. Avendano, and R. O. Duda, "Estimation of a Spherical-Head Model From Anthropometry," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 472–479 (2001 Jun.).

[37] A. Vidal, P. Herzog, C. Lambourg, and J. Chatron, "HRTF Measurements of Five Dummy Heads at Two Distances," in *Proceedings of the International Conference on Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–8 (Bologna, Italy) (2021 Sep.).
https://doi.org/10.1109/I3DA48870.2021.9610914.

[38] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1996).

[39] S. K. Mitra, *Digital Signal Processing: A Computer Based Approach* (McGraw-Hill, New York, NY, 1998).

A.1 SIMULATION METHOD

To simulate the transaural rendering, diffraction by a spherical head model is used because its analytic solution is known and involves a reasonable computational cost. The spherical head model is widely used [33–36] but with various parameters (sphere radius and ear locations). A recent study showed that differences between mannequin heads should be of the same order for a spherical head model [37]. In this paper, the radius is 8.75 cm and ears are placed at 100° from the frontal axis and on the horizontal plane, as used in [33]. The center of the coordinate system of axis is the point in the middle between the two ears.

A basic transaural system is made of two loudspeakers placed at spherical coordinates (r, $\pm\theta$, φ). The coordinate system is described by radius r, azimuth θ, and elevation φ with a null elevation on the equator, following notations by Blauert [38]. The matrix $C(f)$ contains the four transfer functions between loudspeakers and ears. The pressure OUT at the listener ears is defined with the following relation:

$$OUT = [C][H]IN. \tag{14}$$

IN is the binaural signal to reproduce, composed of IN_L and IN_R. $[H]$ is matrix of transaural filters computed as the Moore-Penrose pseudo inverse of $[C]$, regularized with a Tikhonov matrix [6]:

$$[H] = ([C]^T[C] + \beta[Id])^{-1}[C]^T A. \tag{15}$$

Superscript T denotes theHermitianoperator, and βisthe regularization parameter, chosen according to a dynamic Dyn: $\beta = \max(C^2).10^{-\frac{Dyn}{20}}$. Dyn is set to 80 dB in this paper. Id is the identity matrix, and A is a target response corresponding to a FIR filter of 85-ms length delayed by its half length. To simulate the rendering at the listener's ears, signals are built in the temporal domain:

$$s_L(t) = in_L(t) * h_{LL}(t) + in_R(t) * h_{RL}(t), \tag{16}$$

where $s_L(t)$ is the signal played on the left loudspeaker and asterisks denote the convolution operator. The signal $s_R(t)$ played on the right loudspeaker is built in the same way.

A.2 EQUALIZATION OF SELECTED CONFIGURATIONS

For the seven selected configurations tested in Sec. 3, transaural filters were computed in three steps:

1. Impulse responses of all systems were measured ina semi-anechoic room with a mannequin head, B&K 4100-D.
2. Transaural filters were computed using Eq. (2) but with Dyn reduced to 30 dB and with a target response A corresponding to a second-order band-pass Butterworth filter between 150 Hz and 6 kHz.
3. For each listener, a post-equalization was determined individually: frequency response function of the transaural systems was measured in the listening room using binaural microphones (DPA 4060). It was smoothed in one-sixth–octave bands and averaged over the two ears. A minimal phase FIR filter was then computed using the window method [39], with a duration of 85 ms.

THE AUTHORS


Adrien Vidal


Philippe Herzog


Christophe Lambourg


Jacques Chatron

Adrien Vidal graduated in 2013 from the engineering school PHELMA (Grenoble-INP) in physics and signal processing. As a research engineer at Genesis and LMA, he received his Ph.D. from Aix-Marseille University in 2017. Then he studied sonification in sport at Movement Science Institute as a post-doctoral researcher. Since 2019, he has been a research engineer at the PRISM laboratory working on 3D sound systems and sound perception.

•

Philippe Herzog graduated in electronic engineering from ENSEA in 1982. He then worked in the aeronautics division of the Crouzet (now Sextant) group. In 1987, he received a Ph.D. degree in acoustics from Le Mans University. He then worked at the University of Sherbrooke until he joined the French CNRS in 1988. He also stayed 2 years at CTTM and 3 years at Swiss Federal Institute of Technology. In 2000, he joined the LMA laboratory in Marseille. Philippe is the author of about 150 scientific papers. He has been president of the French section of the AES and president of the French Acoustical Society. In 2019, he co-founded the ARTEAC-LAB engineering company, focused on electroacoustics and sound field control.

•

Christophe Lambourg graduated in electronics from Paris XI University in 1991. He received a Ph.D. degree in acoustics from Le Mans University in 1997. He was involved as a consultant for 2 years in a research project on the improvement of acoustic comfort and sound systems of railway stations (SNCF). From 2000 to 2008, within the Signal Development engineering company, he was the head of the Paris branch and conducted numerous research and development studies for industrial key accounts in different fields of acoustics and signal processing. In 2008, he joined the GENESIS company, in which he was responsible for acoustic studies. Since 2019, he has been president of the ARTEAC-LAB company, which he co-founded with Philippe Herzog. During his professional career, he participated in numerous collaborative research projects, involving public research laboratories and industrialists. His main areas of expertise are audio signal processing, sound quality, room acoustics, and electroacoustics.

•

Jacques Chatron is a CNRS engineer in the Sounds Team of the Mechanics and Acoustics laboratory in Marseille since 2009. He works on auditory perception (more precisely loudness, localization, low frequencies, etc.). He develops experimental devices (hard and soft sides) and is responsible for the “Hearing & Perception” resource center of the laboratory. Before that, he started two companies (the first in 1985 and second in 1990) and was a sound engineer for 24 years in his two companies.