Binaural rendering using near-field loudspeakers

Adrien Vidal, Philippe Herzog, Christophe Lambourg, Patrick Boussard, Libor Husník

Binaural rendering using near-field loudspeakers

Adrien Vidal[1,2], Philippe Herzog[2], Christophe Lambourg[1], Patrick Boussard[1], Libor Husnik[3]

[1] *Genesis, Domaine du petit Arbois, BP 69, 13545 Aix-en-Provence, France, Email: adrien.vidal@genesis.fr*

[2] *Laboratoire de Mécanique et d'Acoustique, CNRS UPR 7051, Marseille, France*

[3] *Czech Technical University in Prague, Faculty of Electrical Engineering, Prague, Czech Republic*

Abstract

Binaural rendering through loudspeakers has been studied for decades and is still subject to improvements. To optimize these systems, two major research trends are usually being combined. The first one consists in varying the number of loudspeakers and their positions, and the second one rather uses signal-processing approaches to improve crosstalk-cancellation techniques. Most of the time, transaural systems are implemented in a non-anechoic room whose influence has to be accounted for to compute the crosstalk filters. Treating the room influence by signal processing is not straightforward and leads to high numbers of filter coefficients. Such filters require large computational time and resources and can lead to unwanted audible effects. The approach proposed here consists in using near-field loudspeakers in order to maximize the energy ratio between direct and diffuse fields. In this way the room equalization is simplified, and makes it possible to optimize sound quality even in small rooms with high influence. However, the main problem with using near-field loudspeakers is the robustness of the sound rendering to the misalignment of the listener. To find the best compromise, a large number of loudspeaker configurations are simulated and objective indicators are calculated to assess an expected sound quality. A few configurations are chosen so as to be representative of the range of values of the indicators and are implemented in a medium-size room. Listening to the resulting sound makes it possible to evaluate the predictions of sound rendering quality from the objective indicators.

Introduction

The positioning of electro-acoustic sources is an important factor for the rendering quality in Crosstalk Cancellation Systems (CCS). The initial configurations were based on the stereo standard, using two loudspeakers at an angle of 60° respective to the listener. In the late 90s, Kirkeby *et al* proposed the "stereo-dipole" solution [1], which involves two closely spaced loudspeakers. Using free-field simulations, they showed that this configuration leads to a larger sweet spot size than the traditional stereo configuration. Takeuchi and Nelson [2] then dealt with this method in depth and generalized it to the Optimal Source Distribution solution which involves several frequency bands, with pairs of loudspeakers at various angles.

The stereo-dipole configuration is largely used, even though it is still subject to discussion [3]. The optimal placement of loudspeakers in CCS is still an issue, *eg* in a recent work [4], in which the author proposes to use elevated stereo-dipoles.

A major problem in implementing CCS is the room influence, which leads to the use of filters with a high number of coefficients. In this study, we suggest placement of the listener in the near field of loudspeakers in order to maximize the energy ratio between direct and diffuse fields. It also improves the condition number of the underlying inverse problem. Using numerical simulations, the CCS rendering is computed for a large number of configurations for near-field and far-field loudspeakers. A few objective quality indicators are then computed to rate the configurations. According to the simulations outcome, three configurations are chosen and implemented in a listening room. Measurements and informal listening tests of these systems allow us to relate anechoic simulation results and real rendering.

Simulation of CCS reproduction

Diffraction on a spherical head model

Diffraction by a spherical head model is used to simulate the CCS rendering. 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 [5]–[8] but with various parameters (sphere radius and ears locations). According to previous studies, the radius vary from 8.5 to 9 cm, so we chose a radius of 8.75 cm that has been commonly used since the work of Hartley and Fry [5]. In the same studies, ears locations vary from 90° to 100° in azimuth and from -20° to 0° in elevation. In accordance with Blauert [9], we assume that ears are placed slightly at the back of the head, at 100° from the frontal axis and on the horizontal plane. This configuration implies symmetry with regard to the horizontal plane, so the simulations are only done for positive elevations. Thereafter, the term "interaural axis" refers to the axis connecting the two ears. First, the transfer functions between sources and ears are calculated, and then the crosstalk cancelation filters are computed.

The computational operations follow the same protocol as presented in [7]. Diffraction is computed for 256 frequencies linearly distributed from 86 Hz to 22.05 kHz. The scattered sound pressure is divided by the free-field sound pressure to obtain the Head-Related Transfer Functions (HRTF), as described in [10]. However, the reference pressure is taken at a distance r_0= 1 m from the centre of the head to keep the level differences due to the different source distances.

$$HRTF(r,\theta,\phi,f) = \frac{P_S(r,\theta,\phi,f)}{\frac{j\rho f Q_0}{2r_0} e^{\frac{-j2\pi f r_0}{c}}} \qquad (1)$$

P_S is the sound pressure at the ear; r, θ, ϕ are respectively the distance, the azimuth and the elevation between the source

and the centre of the head; f is the frequency, ρ is the volumic mass of air and Q_0 is the unitary volume velocity.

The computation is performed for distances ranging from 30 to 50 cm by 5-cm steps. The computation is also performed at 1 m which is here considered as a far-field solution. The azimuth value range is from 5° to 175° by 5° steps and the elevation range is from 0° to 80° by 10° steps. A total of 1890 configurations have thus been tested.

Equivalent filters

The previously calculated HRTF are converted into temporal filters, thereafter called "direct filters". First, inverse Fourier transforms of the HRTF lead to Head Related Impulse Responses (HRIR). These FIR direct filters are called "initial phase form". To facilitate the inversion process the HRIR are then converted into minimum-phase filters using the Hilbert transform [11]. Interaural Time Difference (ITD) between the left and right responses is preserved using an estimation based on peak detection. These resulting FIR filters are called "minimum phase form". The Fourier transform of these HRIRs are used to compute the crosstalk cancellation filters. Each filter form is presented in Figure 1.

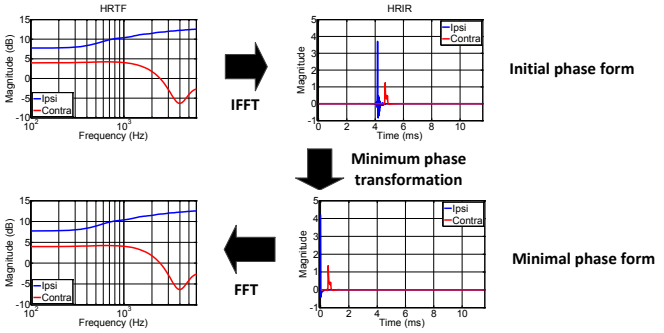

Figure 1: Minimum phase filter calculation

Crosstalk Cancellation filters calculation

The system is described by this equation:

$$\begin{bmatrix} Out(1) \\ Out(2) \end{bmatrix} = \begin{bmatrix} Cd & Ci \\ Ci & Cd \end{bmatrix} . \begin{bmatrix} Hd & Hi \\ Hi & Hd \end{bmatrix} . \begin{bmatrix} In(1) \\ In(2) \end{bmatrix} \quad (2)$$

Where Hd and Hi are the crosstalk cancellation filters, Cd and Ci the direct filters, $Out(1)$ and $Out(2)$ the outputs of the system, $In(1)$ and $In(2)$ the inputs of the system. In corresponds to a binaural recording and Out corresponds to the signal recorded at the listener ears.

$C(f) = \begin{bmatrix} Cd & Ci \\ Ci & Cd \end{bmatrix}$ is the HRTF matrix, where Cd is the ipsilateral HRTF and Ci the contralateral HRTF.

The matrix $H(f) = \begin{bmatrix} Hd & Hi \\ Hi & Hd \end{bmatrix}$ of crosstalk cancellation filters is then calculated using Tikhonov regularization [12]:

$$H = (C^*.C + \beta.Id)^{-1}.C^*.A \quad (3)$$

Where * denotes the Hermitian operator, β the regularization parameter, Id the identity matrix and A a target response.

Here the target response A is chosen as the frequency response of the 4th-order Butterworth filter, with a passband between 150 Hz and 6 kHz. Regularization parameter β is kept constant over the frequency scale. Its value is designed to be 80 dB lower than the maximum norm of the HRTF matrix H.

To ensure the causality of the results, a delay is added to the crosstalk cancellation filters Hd and Hi. This delay is equal to the half-length of these filters.

The simulations are performed in the temporal domain, and the analysis of these results is made in the frequency domain. For the simulation process, $In(1)$ is a Dirac and $In(2)$ is a vector of zeros. These signals are convolved with the matrix of inverse filters, and then with the matrix of direct filters in the initial phase form. We use the initial phase form to avoid compensating a possible bias introduced by the minimal phase representation. The diagram of the simulation process is presented in Figure 2, where $x(1)$ and $x(2)$ are the input signals for the loudspeakers.

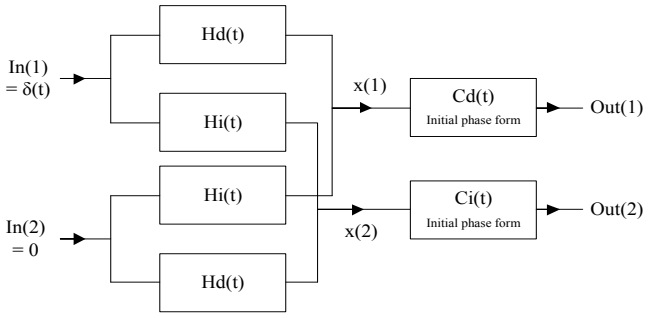

Figure 2: Diagram of the simulation process

Objective indicators calculation

We then compute indicators on the simulation result, in a similar way as in [4] but their number is higher. The indicators are computed in the frequency domain, in the frequency band [150 Hz – 5 kHz] for each spatial configuration. First, each frequency response is converted into 1/48-octave band gains. Indicators are calculated for each 1/48-octave band, and averaging over the frequency band. These averages are converted into a score between 0 and 1, where 1 is the best score. The laws of score attribution follow Gaussian functions defined by their standard deviation σ and their centroid K. Each score attribution law is shown in Figure 4.

Two indicators are commonly used to evaluate the CCS rendering, the Performance Error (PE) and the Channel Separation (CHSP) [4]. PE is related to the fidelity of the restitution for the ipsilateral ear, and CHSP corresponds to the level difference between the two ears.

Performance Error (PE)

This indicator is related to the restitution accuracy for the ipsilateral ear. It assesses the ability to reliably reproduce the signal $In(1)$ into the signal $Out(1)$.

$$PE = \left| 20.\log\left(\frac{|Out(1)|}{|In(1)|} \right) \right| \quad \text{[dB]} \qquad (4)$$

The average PE over the frequency range is PE_{mean}. A mean difference of 1.5 dB is considered as the maximal acceptable difference, and the attributed score S_{PE} is:

$$S_{PE} = e^{\frac{-PE_{mean}^2}{2\sigma^2}} \quad , \quad with \quad \sigma = 1.5 \qquad (5)$$

Channel separation (CHSP)

This indicator describes the level difference between the two ears. This indicator shows the ability to reproduce a signal *Out(2)* with a lower amplitude than *Out(1)*. We assume that an infinite channel separation is not necessary to get a high quality CCS, we thus define a target channel separation for which the score is considered as maximal. This target channel separation is the maximal channel separation computed on the spherical head for a source placed 20 cm away from the head centre. This corresponds to the closest sound source location that can be reproduced (*eg* noise from an opened window in a car). A representation of the target CHSP is shown in Figure 3.

$$CHSP = \left| 20.\log\left(\frac{|Out(1)|}{|Out(2)|} \right) \right| - CHSP_{target} \quad \text{[dB]} \qquad (6)$$

For CHSP values greater than 0 dB, values are set to 0 and the average CHSP over the frequency range is $CHSP_{mean}$. The score S_{CHSP} is attributed using a Gaussian function with σ=7.5:

$$S_{CHSP} = e^{\frac{-CHSP_{mean}^2}{2\sigma^2}} \qquad (7)$$

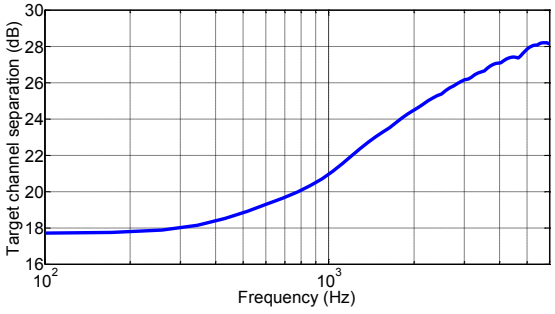

Figure 3: Target channel separation

Free-field simulations generally give very good results for PE and CHSP, although real implementation faces difficulties not accounted for in such simulations. Therefore loudspeaker configurations cannot be ranked according to these usual indicators. Further indicators are needed, that take into account the room influence and sources limitations to predict the rendering in real implementations.

Room has a complex influence, and is not straightforward to model. Room effect is modelled here by a diffuse field that is homogeneous in the whole listening room. The related pressure at the ears is derived from the power radiated outside the listening area. A similar approach is used in [13].

Radiance (Rad-PE and Rad-CHSP)

The Radiance indicators correspond to PE and CHSP calculations accounting for the diffuse field. To estimate the related pressure of the diffuse field, the acoustic power of a source equivalent to the loudspeaker pair is computed. First the pressure radiated outside the listening area is computed over a sphere with radius r = 1.5 m centred on the head. This distance is farther than all the loudspeaker positions and coherent with the distances from the walls of the listening room. The sphere is discretized as N points almost equally spaced, and the quadratic mean pressure over the sphere P_{ext} is computed according to the following equation:

$$P_{ext} = \sqrt{\frac{1}{N}\sum_{\theta,\phi}^{N}\left(P_1(r,\theta,\phi) + P_2(r,\theta,\phi)\right)^2} \quad \text{[Pa]} \qquad (8)$$

Here the number of point N is 1000. P_1 and P_2 are respectively the pressure emitted by the ipsilateral and contralateral loudspeakers on the sphere at the coordinates (r, θ, ϕ):

$$P_i(r,\theta,\phi,f) = \frac{x(i)}{d_i(r,\theta,\phi)} e^{\frac{-2j\pi f d_i(r,\theta,\phi)}{c}} \quad \text{[Pa]} \qquad (9)$$

Where $d_i(r,\theta,\phi)$ is the distance between the loudspeaker i and the point at coordinates (r,θ,ϕ). $x(i)$ is the input signal of the loudspeaker i (*cf* Figure 2).

The equivalent acoustic power W of the loudspeaker pair is derived from this mean pressure:

$$W = \frac{4\pi r^2 P_{ext}^2}{\rho c} \quad \text{[W]} \qquad (10)$$

Energy density of diffuse field is considered as constant in the entire listening room, and the related pressure is computed as [14]:

$$P_{dif} = \sqrt{4\rho c \frac{W}{A}} \quad \text{[Pa]} \qquad (11)$$

Where A is the Sabine room absorption, $A = \sum_i \alpha_i S_i$. For each surface i of a room, α_i is the absorption coefficient and S_i the area of the corresponding surface. For the simulations, a total surface S_{tot} of 80 m^2 is chosen and a mean coefficient absorption α_{moy} is chosen frequency dependant. Its value logarithmically ranges 0.3 to 0.8 from 150 Hz to 5 kHz. These correspond to a suitable, largely damped, medium listening room [15]. As the chosen absorption coefficient is superior to 0.2, the Eyring relation is preferred to the Sabine

relation [14]: $A(f) = -\ln(1-\alpha_{moy}(f))S_{tot}$. The mean room absorption over the frequency range is so A=64 m^2. This point differs from [13].

For the assessment of small impairments in audio systems, ITU [15] recommends using listening room with specific parameters. ITU recommends a floor area between 20 m^2 and 60 m^2 and an average reverberation time T over the frequency range 200 Hz to 4 kHz equal to $T = 0.25(V/V_0)^{1/3}$ where V is the room volume and V_0 is a reference volume of 100 m^3. Reverberation time is related to room volume and Sabine room absorption according to the following equation [14]: $T = 0.16V/A$. Considering a ceiling height of 2.5 m, the recommendation leads to a room absorption area between 40 m^2 and 84 m^2. The mean of the chosen value A=64 m^2 thus corresponds to the middle range of recommended values.

Two radiance indicators are computed, according to the following equations:

$$Rad_{PE} = \left| 20\log\left(\frac{\sqrt{|P_{dif}|^2 + |Out(1)|^2}}{|In(1)|} \right) \right| \quad \text{[dB]} \quad (12)$$

$$Rad_{CHSP} = \left| 20\log\left(\frac{|Out(1)|}{\sqrt{|P_{dif}|^2 + |Out(2)|^2}} \right) - CHSP_{target} \right| \quad \text{[dB]} \quad (13)$$

These definitions are analogue to the PE and CHSP definitions, including a diffuse-field term.
Scores $S_{radiance}$ are given by the same law as PE and CHSP, see Eq (5) and Eq (7) replacing PE and CHSP by Radiance$_{PE}$ and Radiance$_{CHSP}$ respectively.

Robustness to head misplacement (PE-R and CHSP-R)

3D sound systems and especially CCS are sensitive to the listener's position, and misplacement or head movement corrupt the rendering. We assume that the listener's position is controlled, and that the listener pays attention not to move the head. But incertitude of the position is possible, and we assume the amplitude of this error to be a 2-cm translation of the head in each direction or a 5° rotation of the head. To quantify this error on the rendering, the simulation is computed using direct filters for which the head is displaced.

Four kinds of movement are tested: 2-cm lateral displacement according to the *x, y, z* axes in the Cartesian representation and 5° head rotation in the azimuthal plane. Each movement is computed for the two axis directions or direction of rotation, which leads to height tested movements. The reconstruction is computed in the same way as presented before, then PE and CHSP indicators are computed for each kind of movement. The Performance Error Robustness (PE-R$_2$) is then the minimal score among all versions of PE, and the Channel Separation Robustness (CHSP-R$_2$) is the minimal score among all versions of CHSP. Using these definitions, only the "worst" displacement is taken into account.

A larger displacement is also computed, which should correspond to head movements (not misplacement). The same calculation as before is performed considering a 5-cm lateral displacement and a 10° rotation. These indicators are named PE-R$_5$ and CHSP-R$_5$.

Efficiency (Eff)

This indicator quantifies the increase of the loudspeaker volume velocity induced by crosstalk cancellation filters. This indicator can be related to the final dynamic range of the system.

$$Eff = \frac{Q_{CCS}}{Q_{monopolar}} \quad (14)$$

Q_{CCS} is the total volume velocity of the loudspeaker pair to get 1 Pa on the ipsilateral ear using crosstalk cancellation filters, and $Q_{monopolar}$ is the total volume velocity of the loudspeaker pair to get 1 Pa on the ipsilateral ear without using crosstalk cancellation filters.

Doubling the volume velocity is considered unacceptable, and the attributed score S_{eff} is then:

$$S_{Eff} = e^{-\frac{(-Eff-K)^2}{2\sigma^2}} \quad with \quad \sigma = 0.5 \quad and \quad K = 1 \quad (15)$$

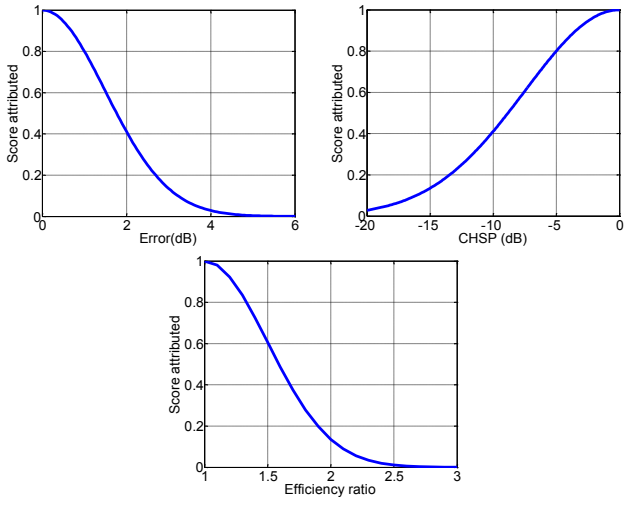

Figure 4: Score attribution laws

Simulation results

Results for the horizontal plane are plotted in Figure 5. Each configuration corresponds to a loudspeaker pair, but only one loudspeaker position of the pair is represented. Results for various elevations at 40 cm distance are represented in Figure 6.

The two usual indicators (PE and CHSP) are not plotted for these representations, because the values are almost maximal for all configurations and no difference is visible.

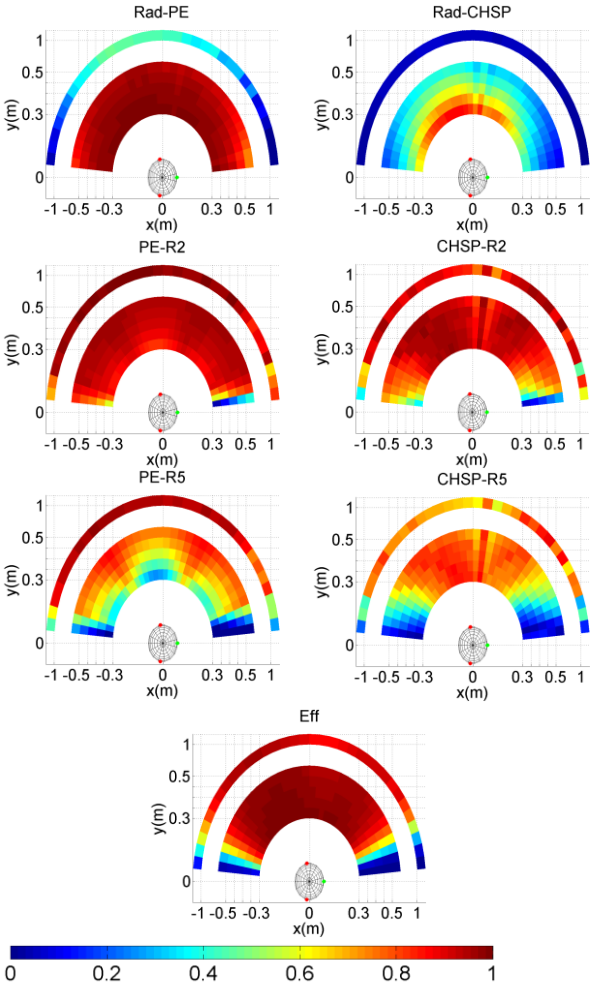

Figure 5: Simulation results in the horizontal plane. For clarity, the scale is not the same between far-field and near-field configurations: 1-m configurations are represented closer to the listener than they should be.

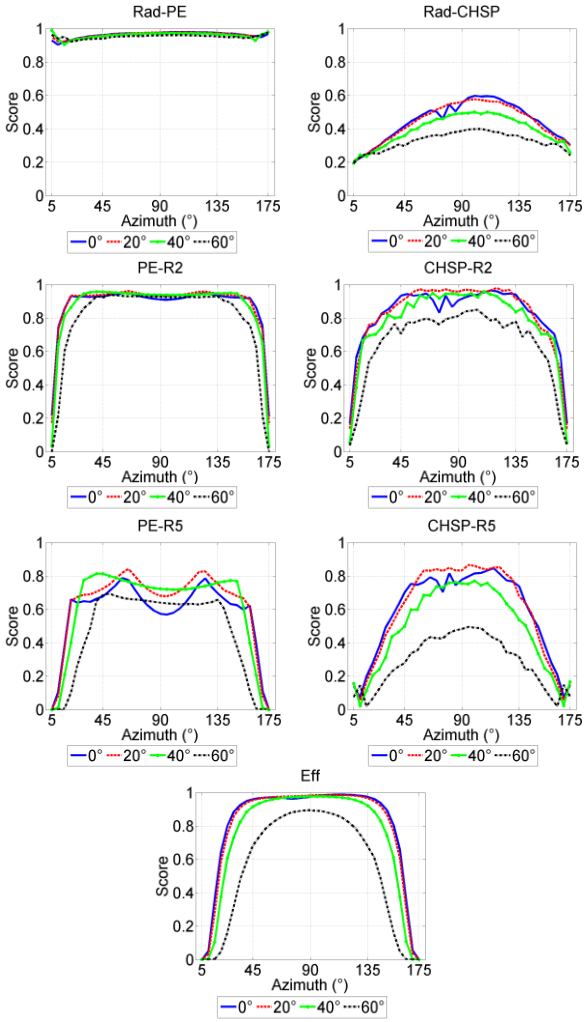

Figure 6: Simulation results for sources at 40 cm from the centre of the head for various elevations.

Rad-PE and Rad-CHSP scores are similar, but Rad-CHSP scores are lower than Rad-PE scores. Radiance scores show the largest variations over the distance range: far-field configurations get the worst scores. Especially for Rad-CHSP scores, an angular dependence is also visible and the best configurations are near the interaural axis, slightly in the rear in the horizontal plane. It seems that the best scores are obtained for the configurations close to the ear. Scores are slightly better in the rear, probably because a model with ears positioned at the rear is chosen. Elevation seems having a slight effect, high elevated configurations get the lower scores.

Robustness scores show differences according to the position. For small displacements (PE-R_2 and CHSP-R_2), scores are excellent for almost all the configurations, except for those close to the median plane. A slight range dependence is visible for PE-R_2, and far-field configurations get the best scores. For larger displacements (PE-R_5 and CHSP-R_5), the trends are similar. For PE-R_5, bad scores are also observed near the median plane and the same range dependence is visible: far-field configurations get the best scores. The range dependence is more visible for PE-R_5 than PE-R_2. The angular dependence is amplified for CHSP-R_5: in the azimuthal plane only configurations between 45° and 140° give good scores. According to Figure 6, slightly elevated loudspeakers give better scores.

Efficiency scores show variations similar to CHSP-R. The closer to the median plane a configuration is, the lower its score is, and high elevated configurations get the lower scores.

According to these proposed indicators, room influence affects PE and CHSP in a similar way: the configurations which are close to the ear get the best scores. Almost all configurations are robust to head misalignment (2-cm displacements), except configurations close to the median plane. Head movements (5-cm displacement) induce higher differences according to positions: PE for configurations in near-field area and configurations near the median plane is not robust to head-movement, and neither is CHSP for configurations far from the interaural axis. Slightly elevated configurations get slightly better scores than configurations in the median plane. We did not find any physical explanation of the fact that slightly elevated configurations are more robust. The area around the near-field position (θ = 110°, ϕ = 20°) seems to be the better one.

Implementation for sample configurations

According to these results, 3 configurations are chosen and are implemented in a small-sized, acoustically treated room. The configurations are:

A) ($\theta = \pm 110°$, $\phi = 30°$, r = 40 cm). This is an optimal configuration according to previously computed indicators.

B) ($\theta = \pm 45°$, $\phi = 0°$, r = 40 cm). This configuration gets good scores, but unlike the configuration A) loudspeakers are in front of the listener and in the horizontal plane. Preliminary listening has shown that the perceived source locations may be correlated with the loudspeaker locations. This configuration makes it possible to test whether rear sources can still be perceived even when using frontal loudspeakers.

C) ($\theta = \pm 7°$, $\phi = 0°$, r = 40 cm). This configuration is the near-field version of the stereo-dipole. A spanning angle of 7° is used instead of 5°, because the loudspeaker size does not allow the use of spanning angles below 7°.

All these configurations are placed at the same distance from the listener, in the middle of the room. The ceiling height of the room is 2.5 m, and the floor surface, which is not square, is approximately 18 m^2. It is slightly lower than ITU recommendation [15]. The loudspeakers used in the experiment are Cabasse Alcyone satellites 2 and the system stands 110 cm above the floor. The diagram of the implementation process is presented in Figure 7. *H* matrix is computed using a simulated anechoic transfer function in the same way as for the simulation, and an equalization filter *F* is introduced. As a first approximation, only the loudspeakers are equalized using 5 biquad filters computed from anechoic measurements following a method presented in [16].

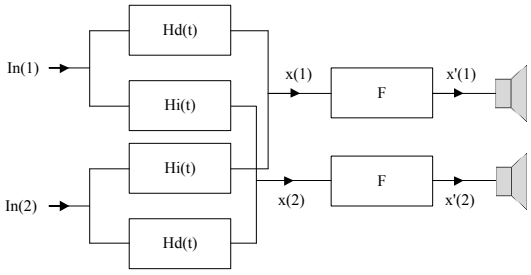

Figure 7: Diagram of the implemented configurations

The scores attributed from the simulation are shown in Figure 8. The PE and CHSP scores are unity for all configurations and are not represented here. Rad-PE scores are very high and close to unity for all the configurations, whereas Rad-CHSP scores are lower and a clear ranking is visible: the best configuration is A) and the worst is C). The robustness and efficiency scores are very good for configurations A) and B), and very low or null for configuration C).

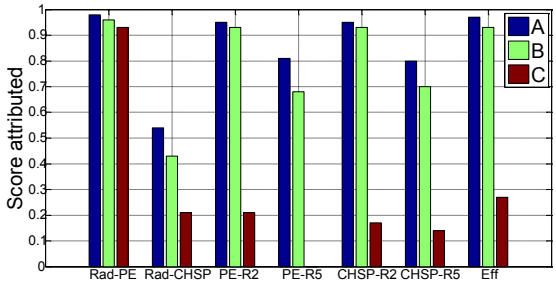

Figure 8: Attributed simulation scores for the tested configurations

Objective measurements

A rigid sphere made of ABS plastic is used to measure the configuration responses. The sphere parameters are the same as those used in the simulation (8.75-cm radius and 100° ears location) and the microphones are GRAS 40PR. For selected incidence angles, the transfer functions of the sphere are measured in an anechoic chamber. The measurements are close to the simulation and differ by less than 1 dB over the frequency band [100 Hz – 6 kHz].

Each configuration is measured using a 30-second pink noise signal on one track and silence on the other. A separate measurement is performed for each microphone. The transfer functions between input and output are computed using the H1 estimate. The average of the deviations between the restitution on ipsilateral ears and the target response is computed from these two measurements. This average is used to compute a room equalization filter *F* using 10 biquad filters [16]. The configurations are then measured again, using both loudspeaker and room equalization. The resulting frequency responses are plotted in Figure 9. For each case, PE and CHSP indicators are computed using these measurements smoothed in 1/6th-octave bands. The mean scores for the left and right measurements are reported in Figure 10.

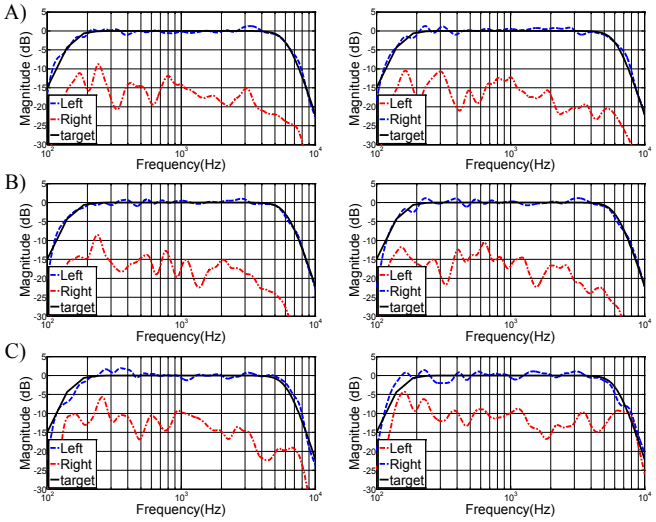

Figure 9: Frequency responses of implemented systems A), B) and C). The measurements on the left column are made with pink noise on the left input signal, and those on the right column are made with pink noise on the right input signal. The frequency responses are smoothed in 1/6th-octave bands.

To evaluate the distance influence, the configurations are also implemented in a far-field version, at a 1-m distance. Configuration C) is then implemented with a spanning angle of 5°.

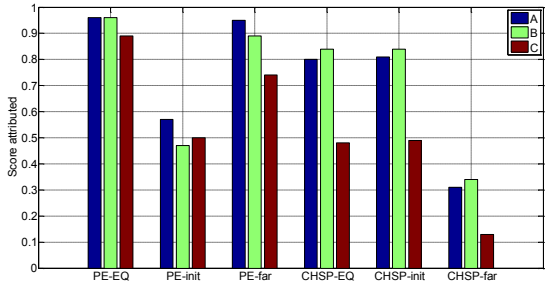

Figure 10: Attributed scores processed on the measurements. PE-EQ and CHSP-EQ are attributed scores with room equalization; PE-init and CHSP-init are obtained without room equalization; PE-far is the far-field version with room equalization.

The results show that configurations A) and B) are quite similar. The reconstruction on the ipsilateral side is correct within a range of +/- 1 dB and CHSP is always superior to 10 dB and even 15 dB over most of the frequency band. Configuration C) differs from the others: the reconstruction on the ipsilateral side is correct within a range of +/- 2 dB and CHSP is around 10 dB for the whole frequency band. The scores reported in Figure 10 highlight significant differences between this configuration and the two others, especially for CHSP.

Figure 10 shows that even with room equalization PE is very close to unity but is not maximal for all configurations. For configurations A) and B), CHSP is quite good and for configuration C) it is mediocre. We can notice that the CHSP of configuration B) is slightly better than that of configuration A), and the equalization has no effect on CHSP. For all the measured configurations, PE is slightly lower than in the simulation and CHSP is significantly lower. Indeed CHSP is more sensitive than PE to room influence. This difference in sensitivity could be seen in Figure 8: Rad-PE scores are higher than Rad-CHSP scores.

For configuration C), the measured PE is slightly lower and CHSP is significantly lower than other configurations. As can be seen in Figure 8, this configuration gets lower Radiance scores than the two others. Radiance indicators predict quite well differences between this configuration and the two others.

The far-field configurations are characterized by a significantly lower CHSP. When increasing the distance, PE is slightly lower for configurations B) and C).

Misplacement is not considered a major aspect here, because the system is set up carefully, and the arrival time of each loudspeaker is controlled using a microphone placed at the centre of the system. It should be possible to measure misaligned configurations, but these measurements would cumulate room influence and misalignment. As the simulation showed, the room has a higher influence on the rendering than misplacement, and this seems to be confirmed by measurements.

Informal listening tests

In addition to these measurements, informal listening tests were performed using near-field CCS. Three monophonic stimuli were recorded in an anechoic chamber at five different locations: azimuths 0°, 30°, 60°, 90° and 120° in the horizontal plane at a 80-cm distance. These stimuli were recorded through a B&K 4100-D dummy head and then diffuse-field-equalized according to the data provided by the manufacturer. The first stimulus is a French phonetically balanced sentence pronounced by a man, the second one is a pulse train, and the last one is a noise burst. The seat height was adjustable to adapt to the subject's size. Five people participated to this informal experiment. All stimuli were played through each near-field CCS configuration and compared with playback through Beyerdynamic DT-990 Pro headphones. This reference listening is called "binaural listening" thereafter.

The sources are quite well localized using CCS, apart from the rear sources with configuration A), which are sometimes perceived as frontal sources. An encouraging point is that using rear loudspeakers (configuration C), frontal sources are quite correctly perceived. With binaural listening, the sources are not all well localized, especially the front sources which are perceived inside the head.
For the configuration with elevated loudspeakers some subjects perceive elevated sources whereas virtual sources are localized in the horizontal plane. We assume this is due to the use of a head model without torso to compute inverse filters. Indeed, the importance of torso for elevated sources has already been shown [17].
The timbre of the stimuli is perceived as very close to that of binaural playback. Moreover, a constant restitution of timbre according to the source location is perceived for all configurations for the speech signal. For the other stimuli a variation of the spectral content is perceived, especially for the lateral and rear sources. However, this variation is also perceived in binaural listening. The spectral variation according to the source position seems to be an inherent flaw of binaural recordings.

Conclusions

In this paper, a comparison between CCS rendering with different loudspeaker positions is attempted. First of all, CCS rendering was simulated for high numbers of positions, especially in the near-field area. The simulation of the rendering was computed using a spherical head model. Three groups of indicators were computed, which were expected to predict the CCS behaviour. For each indicator, scores were attributed according to physical considerations.

Performance Error (PE) and Channel Separation (CHSP) are commonly used to evaluate CCS rendering, but in the free-field simulation these indicators did not differentiate configurations according to the position. They were thus modified to suit a more realistic situation. The room influence is assessed through Radiance indicators. These indicators reveal a high-range dependence, the near-field configurations being the best. A slight angular dependence is also visible, the configurations close to the interaural axis gave the best scores.

The second considered phenomenon is head misalignment, assessed through the Robustness indicators. According to these criteria, all configurations should not be equal, those which are close to the interaural axis and far enough from the head are the most robust for PE. The configurations near the median plane are the least robust for CHSP.

A loss of dynamic induced by crosstalk cancellation filters is assessed through the Efficiency indicator. The configurations near the median plane gave bad scores and the distance has no effect. In conclusion, excepted for PE robustness, the best configurations are those which are the closest to the ear.

Three near-field sample configurations were implemented in a typical listening environment, in order to compare anechoic simulations and real implementation. Measurements were made to objectively evaluate CCS rendering, using PE and CHSP indicators. Radiance indicators predicted quite well the PE and CHSP loss. Rad-PE indicator was however slightly lower than the measured PE without equalization. On the contrary Rad-CHSP was higher than the measured CHSP. The equalization improved significantly the measured PE and had no effect on CHSP. However, the equalization did give the maximal PE score, especially for the configuration which had the worst initial PE score.

Moreover, informal listening with 5 listeners was conducted to complement objective measurements. Localization was better perceived for the configurations that gave good simulated and measured scores. The timbre perception is not significantly different between the configurations, and close to binaural listening. These are however very preliminary tests, and emphasize the need for more rigorous listening tests, which are planned in the near future.

The configurations were also implemented at 1-m distance, which is considered as a far-field solution. The results for far-field configurations were significantly worse than for the near field, especially for CHSP. The differences according to the span angle were similar in near-field and far-field implementations.

An unexpected conclusion is that the standard stereo-dipole is almost the worst configuration in the simulation. The measurements showed that the results of the stereo-dipole are also worse than the other configurations, at any distance. This conclusion is in contradiction with previous work [1], [4], but these studies did not take into account the room influence. However, the near-field version of the stereo-dipole is better than all far-field configurations: the distance thus seems to have a greater importance than the span angle.

Indicators are defined in a rather arbitrary way, and should be improved using perceptive criteria. Preliminary computations using different indicator parameters were made. It led to different indicator values, but leading to very similar ranking of the configurations.

Ongoing work therefore deals with improving the proposed indicators, based on results from pending rigorous listening tests. Further work should address the implementation of configurations using more realistic HRTF, instead of using a rigid sphere model. The perception of elevated configurations could then be different. Another perspective should be the study of configurations with more than two loudspeakers, which may improve the robustness to room influence.

References

[1] O. Kirkeby, P. 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.

[2] T. Takeuchi and P. Nelson, “Optimal source distribution for binaural synthesis over loudspeakers,” *J. Am. Soc. Am.*, vol. 112, no. 6, pp. 2786–2797, 2002.

[3] M. Bai and C.-C. Lee, “Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction,” *J. Am. Soc. Am.*, vol. 120, no. 4, pp. 1976–1989, 2006.

[4] Y. L. Parodi, “A systematic study of binaural reproduction systems through loudspeakers,” Phd thesis, Aalborg university, 2010.

[5] R. Hartley and T. Fry, “The binaural localization of pure tones,” *Phys. Rev.*, vol. 13, no. 6, pp. 431–442, 1921.

[6] R. Duda and W. Martens, “Range dependence of the response of a spherical head model,” *J. Am. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, 1998.

[7] D. Brungart and W. Rabinowitz, “Auditory localization of nearby sources. Head-related transfer functions,” *J. Am. Soc. Am.*, vol. 106, no. 3, pp. 1465–1479, 1999.

[8] R. Algazi, C. Avendano, and R. Duda, “Estimation of a spherical-head model from anthropometry,” *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 472–479, 2001.

[9] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization, Revised Edition*. 1997.

[10] H. Moller, “Fundamentals of Binaural Technology,” *Appl. Acoust.*, vol. 36, pp. 171–218, 1992.

[11] A. Oppenheim and R. Schafer, *Digital signal processing*, Prentice-Hall international Editions. 1975.

[12] O. Kirkeby and P. Nelson, “Digital filter design for inversion problems in sound reproduction,” *J. Audio Eng. Soc.*, vol. 47, no. 7/8, pp. 583–595, 1999.

[13] M. Galvez and F. Fazi, “Loudspeaker arrays for transaural reproduction,” presented at the International Congress on Sound and Vibrations, Firenze, 2015.

[14] A. Gade, “Acoustics in Hall for Speech and Music (pp 301-350),” in *Springer Handbook of Acoustics*, Springer New-York., Thomas Rossing, 2007.

[15] “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,” *Recommendation ITU-R BS.1116-1*, 1997.

[16] G. Ramos and J. Lopez, “Filter design method for loudspeaker equalization based on IIR parametric filters,” *J. Audio Eng. Soc.*, vol. 54, no. 12, pp. 1162–1178, 2006.

[17] R. Algazi, C. Avendano, and R. Duda, “Elevation localization and head-related transfer function analysis at low frequencies,” *J. Am. Soc. Am.*, vol. 109, no. 3, pp. 1110–1122, 2001.