Room Compensation for Binaural Reproduction with Loudspeaker Arrays

Marcos Felipe Simón Gálvez [1] and Filippo Maria Fazi [1]

[1] The authors are within the Institute of Sound and Vibration Research, University of Southampton, Southampton, Hampshire, SO17 1BJ, United Kingdom.
(M.F.Simon-Galvez@soton.ac.uk)

Abstract

This paper introduces a method for compensation of room reflections for binaural reproduction with loudspeaker arrays. The paper considers a line array placed in a room with parallel walls. By knowing the position of the array with respect to the room walls, it is possible to control the radiation pattern so that the lateral reflections are not excited, hence leading to a more uniform reproduced frequency response. Simulations of performance are presented on the document considering first order reflections together with a perturbation analysis to assess the robustness of the method. Practical measurements are also included to assess the feasibility of the method in a real-time implementation of a binaural system.

Keywords: Binaural reproduction, Transaural, Loudspeaker Array, room acoustics

PACS no. 60.Fg, 60.Gk

1 Introduction

Reflections from the room surfaces modify the perceived localization of binaural virtual audio cues rendered through loudspeakers. Previous research has calculated binaural responses inside reverberant environments [1] and it has been shown that room reflections modify the observed interaural time difference (ITD) [2]. This effect has also been studied subjectively, finding that reflections arriving 5 ms and 10 ms after the direct sound clustered the perceived localisation towards the front half plane and increased the probability for front back reversals [3]. Other studies suggested that the reproduction environment had to be relatively absorbent for an accurate virtual source localisation [4], which is something difficult to obtain in many listening situations.

In order to compensate for room reflections, some applications for soundfield control have considered incorporating the room in the loudspeaker transfer functions [5]. Based on this same idea, other researchers have proposed to reduce some of the room reflections but excite others that are beneficial to panning, whilst at the same time smoothing the reproduced response [6]. Alternative research streams have used adaptive approaches with microphones placed in the room [7]. Nevertheless, these methods have been studied numerically, and it is not clear how they perform in a real environment wherein loudspeaker and microphones have small positioning errors.

The method presented herein uses a novel approach for the reduction of lateral reflections. Using the method of mirror source images, a system is introduced that places notches in the loudspeaker array radiation pattern at the positions wherein the walls are crossed by the imaginary line that connects the listener's head with the mirror loudspeaker arrays, hence reducing the excitation of first order reflections. This system needs just a few parameters to operate, which are the position of the listener in the room with respect to the array, and the position of the lateral walls with respect to the array. These parameters can be obtained by using an optical acquisition system as this used by listener adaptive loudspeaker arrays [8] to track the listener's position.

The structure of this document is as follows: Chapter 2 explains the concept of cross-talk cancellation, Chapter 3 explains the implementation of the system for reducing the first order reflections, chapter 4 introduces a robustness analysis and Chapter 5 presents a set of real-time performance measurements.

2 Cross-Talk Cancellation

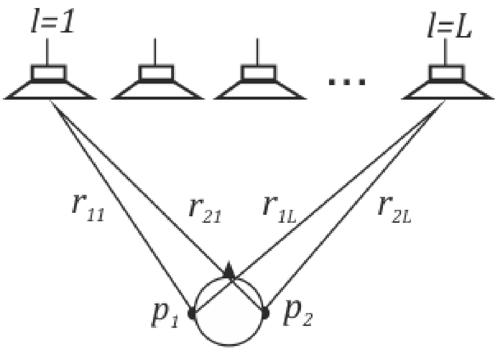


Figure 1: Geometry of the loudspeaker array used to control the acoustic pressure at the listener's ears.

Binaural reproduction over loudspeakers works by controlling the sound-field at the listener's ears. An example is introduced in Fig. 1, where it is shown a system comprising a set of L control loudspeakers with coordinates $\mathbf{y}_l, l = 1,\cdots,L$. The loudspeakers are driven by signals $\mathbf{v} = [v_1, \quad v_2]^{\mathrm{T}}$, where the subscripts 1 and 2 refer to left and right ears and it is assumed a single radiating frequency so that v=$v(j\omega)$, where ω=$2\pi f$ is the radiating frequency. The listener's ears can be represented by M=2 control microphones of coordinates $\mathbf{x}_m, m = 1,\cdots,M$. The acoustic pressures at the listener's ears are defined as $\mathbf{p} = [p_1, \quad p_2]^{\mathrm{T}}$, which may also be written as

$$\mathbf{p}=\mathbf{Cv}, \tag{1}$$

where **C** is the matrix of transfer functions between the loudspeaker and the control points simulating the listener's ears, which taking as reference the geometry of Fig. 1 and assuming that each loudspeaker behaves as a point source can be written as

$$\mathbf{C} = \frac{\rho_0}{e\pi}\begin{bmatrix} \dfrac{e^{-jkr_{11}}}{r_{11}} & \cdots & \dfrac{e^{-jkr_{1L}}}{r_{1L}} \\ \\ \dfrac{e^{-jkr_{21}}}{r_{21}} & \cdots & \dfrac{e^{-jkr_{2L}}}{r_{2L}} \end{bmatrix}, \tag{2}$$

where $\rho_0 = 1.28\ \mathrm{kgm}^{-3}$ is the air density, k=ω/c_0 is the wavenumber and c_0 is the speed of sound in the air. A positive time convenience e^{jwt} is used. The binaural signals that are to be synthesised at the ears of the listener are defined by the elements of the complex vector $\mathbf{d} = [d_1, \quad d_2]^{\mathrm{T}}$. Ideally, the matrix of cross-talk cancellation filters will reproduce the left signal channel, d_1, only at the left ear and the right signal channel, d_2, only at the right ear. This is obtained by introducing a filter matrix of cross-talk cancellation filters, **H**, so that **v**=**Hd** and hence the input signals are modified to assure the cross-talk cancellation. The acoustic pressures at the listener's ears can then be expressed as

$$\mathbf{p}=\mathbf{CHd}. \tag{3}$$

The matrix **H** can be obtained using inverse filtering techniques, as for example shown in [9, 10]. For the formulation presented here, a weighted approach is used, which allows to perform a frequency dependent selection of the acoustic control [11, 12]. This is written as

$$\mathbf{H} = \left[\mathbf{C}^{\mathrm{H}}\mathbf{\Psi}\mathbf{C} + \beta\mathbf{I}\right]^{-1}\mathbf{C}^{\mathrm{H}}\mathbf{\Psi}\mathbf{p}_{\mathrm{T}} \tag{4}$$

where β is a regularisation parameter that control the amount of energy used by the loudspeaker array filters [9] and the superscript H stands for the Hermitian conjugate transpose. The filters are multiplied by a vector of target pressures, $\mathbf{p}_{\mathrm{T}}$, which takes different form for the control filters of each ear, so that $\mathbf{p}_{\mathrm{T}} = [1,0]^{\mathrm{T}}$ for the left ear and $\mathbf{p}_{\mathrm{T}} = [0,1]^{\mathrm{T}}$ for the right ear, giving place to $\mathbf{H}_{\mathrm{L}}$ and $\mathbf{H}_{\mathrm{R}}$. The matrix **Ψ** weights the importance of each control point in the optimisation. This is used in the next sections to control the amount to which first order reflections are not excited at the low frequencies. The matrix **Ψ** is a M×M diagonal matrix in which each diagonal element corresponds to a control point and can vary between 0 (as if discarded) and 1 (fully taken in account) to account for the weight of such control point in the optimisation.

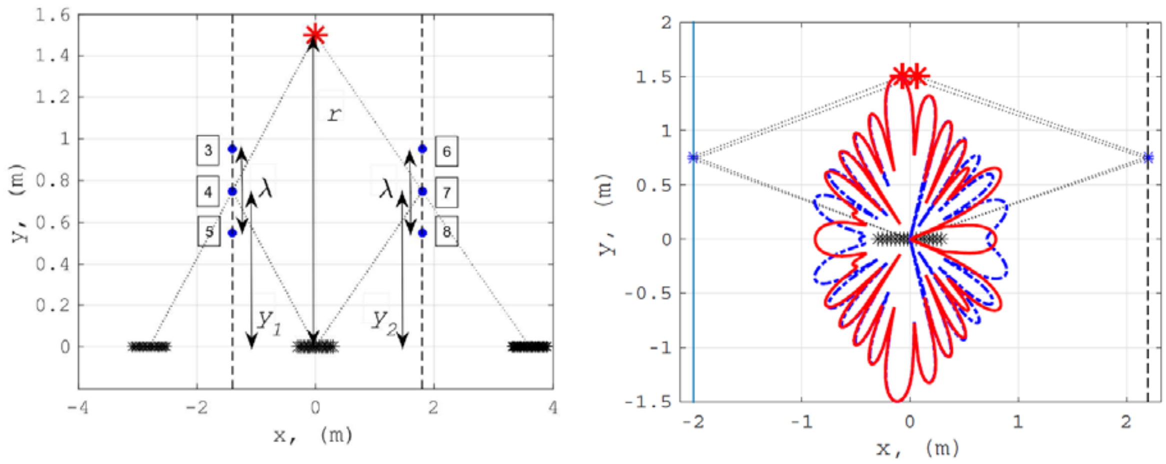
A performance metric employed along the paper is the cross-talk cancellation spectrum, defined as the channel separation at the listener's ears

$$CTC_{\mathrm{L}} = \frac{|p_1|^2}{|p_2|^2}, \text{ and } CTC_{\mathrm{R}} = \frac{|p_2|^2}{|p_1|^2}. \tag{5}$$

In a symmetrical listening configuration $CTC_{\mathrm{L}} = CTC_{\mathrm{R}}$. Therefore, in the next sections of this document this is just referred as *CTC*.

3 Avoiding Reflections

In order to reduce the effect of the lateral reflections, it is first needed to understand how much these affect the direct signal from the array. To this end, the incoming reflections from the lateral walls are modelled using the image source method [13, 14]. A sketch of the geometry used for the control of the lateral reflections can be observed in Fig. 2a, where it is shown a 16 loudspeaker line array and the mirror images produced by the two lateral walls.



(a) Control points definition

(b) Radiation pattern

Figure 2: The left hand side plot shows the schematic used for controlling the first order reflections of the loudspeaker array with respect to the lateral walls. The thin stars represent the loudspeaker array and its mirror images, whilst that the thick stars are the control points simulating the ears of a listener. The closed circles are the extra control points introduced to notch the radiation pattern of the array at these positions. The right hand side plot compares the radiation pattern of the array when the lateral reflections are compensated (solid line) with this obtained by traditional filter creation techniques (dashed line) [15].

The minimisation of the energy sent towards the room reflectors is obtained by using a modified matrix of transfer functions, $\mathbf{C}_{\mathrm{C}}$, which introduce extra control points on the radiation directions corresponding to the first order reflections, as shown in Fig. 2a. The array filters are obtained by using the following vectors of target pressures, defined as

$$\mathbf{p}_{\mathrm{TC,L}} = [1,0,0,0,0,0,0,0], \tag{6}$$

and

$$\mathbf{p}_{\mathrm{TC,R}} = [0,1,0,0,0,0,0,0], \tag{7}$$

The first two elements of the vectors correspond to the control points for the listener ears, whilst that the other 6 control points are placed at the lateral walls to control the first order reflections. The control points 4 and 7 correspond to those reflections which will arrive to the listener's head centre, whilst that the control points 3,5,6 and 8 are introduced to increase the robustness of the method. The position along the x axis of all the control points is fixed, whilst that the position of elements 4 and 7 is fixed along frequency, the position of elements 3,5,6 and 8 along the y axis is frequency dependent, and is given by

$$\mathbf{x}_m(y) = [\sim, y_m, \sim] = \left[r, r, y_1 - \lambda/2, y_1, y_1 + \lambda/2, y_2 - \lambda/2, y_2, y_2 + \lambda/2\right], \tag{8}$$

where r is the distance from the loudspeaker array to the listener ears and y_1 and y_2 are the y axis coordinates of the walls control points. The reader may refer to Fig. 2a for a better understanding of these quantities. In this case, a $\lambda/2$ spacing between the control points has been employed, as previously used in personal audio applications to control the loudspeaker array according to the radiation frequency [12]. An example of the loudspeaker array radiation pattern that is obtained by using this formulation is shown in Fig. 2b, wherein it can be seen that by introducing extra control points the array radiates less energy towards the walls in the direction of the first order reflections.

One of the drawbacks of small size loudspeaker arrays is that these are only effective when their aperture is comparable to the radiation wavelength [16]. For the application considered here, this means that, at low frequency, a small array will not be very effective at in reducing the energy radiated towards the walls without this affecting the cross-talk cancellation. To avoid this effect, a frequency dependent weighting of the control points can be used .

An example is brought below. In this case, the matrix $\mathbf{\Psi}$ was arranged so that elements 1 and 2 of its diagonal are equal to 1, whilst that the rest of the diagonal element follow a first order high-pass filter distribution given by

$$w_l = \frac{\dfrac{j\omega}{\omega_c}}{1+\dfrac{j\omega}{\omega_c}}. \tag{9}$$

Using the formulation described in the sections above, a simulation was performed considering a loudspeaker array of L=16 drivers spaced d=0.0386 m, with a total aperture of 58 cm. In this case, $\omega_c = 2\pi 800$ rad/s, assuring good crossover between cross-talk cancellation at the low frequencies and reduction of first order reflections at higher frequencies.

In order to include in the simulations the lateral walls first order reflections, the performance was assessed by calculating the response according to

$$\mathbf{p}_\mathrm{R} = \mathbf{C}_\mathrm{R}\mathbf{H}_\mathrm{R}\mathbf{d}, \tag{10}$$

where in this case $\mathbf{C}_\mathrm{R}$ is an M×3L matrix of transfer functions taking in account the image loudspeaker arrays and $\mathbf{H}_\mathrm{R}$ is an L×3 matrix of cross-talk cancellation filters of identical columns to account for the modified size of $\mathbf{C}_\mathrm{R}$.

The cross-talk cancellation and frequency response results is shown in Fig. 3: for a case in which the response is calculated in the free-field, (“Free field"), a case in which the response is calculated with free-field filters but assessed in a environment with first order reflections by using Eq. 10, (“Uncompensated"), and a case with extra control points included to reduce the effect of first order reflections, (“Compensated"), in which the reproduced pressures are also calculated with Eq. 10. The results show an increase of the cross-talk cancellation for the “Compensated" response, leading to large, over-predicted, cross-talk cancellation figures of about 50 dB at some frequencies. The frequency responses also predict a much smoother response for the “Compensated" case.

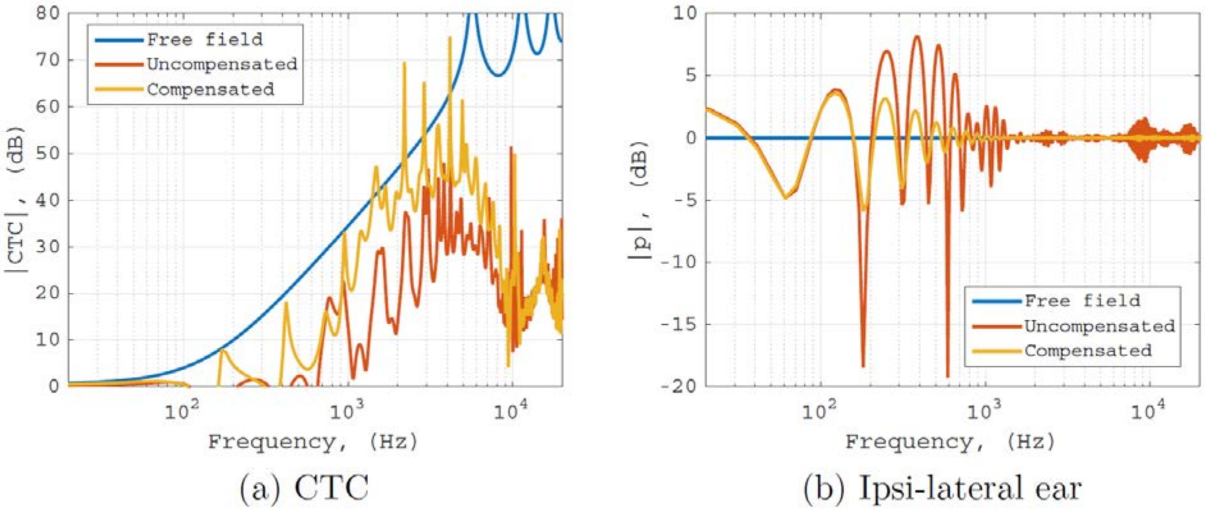


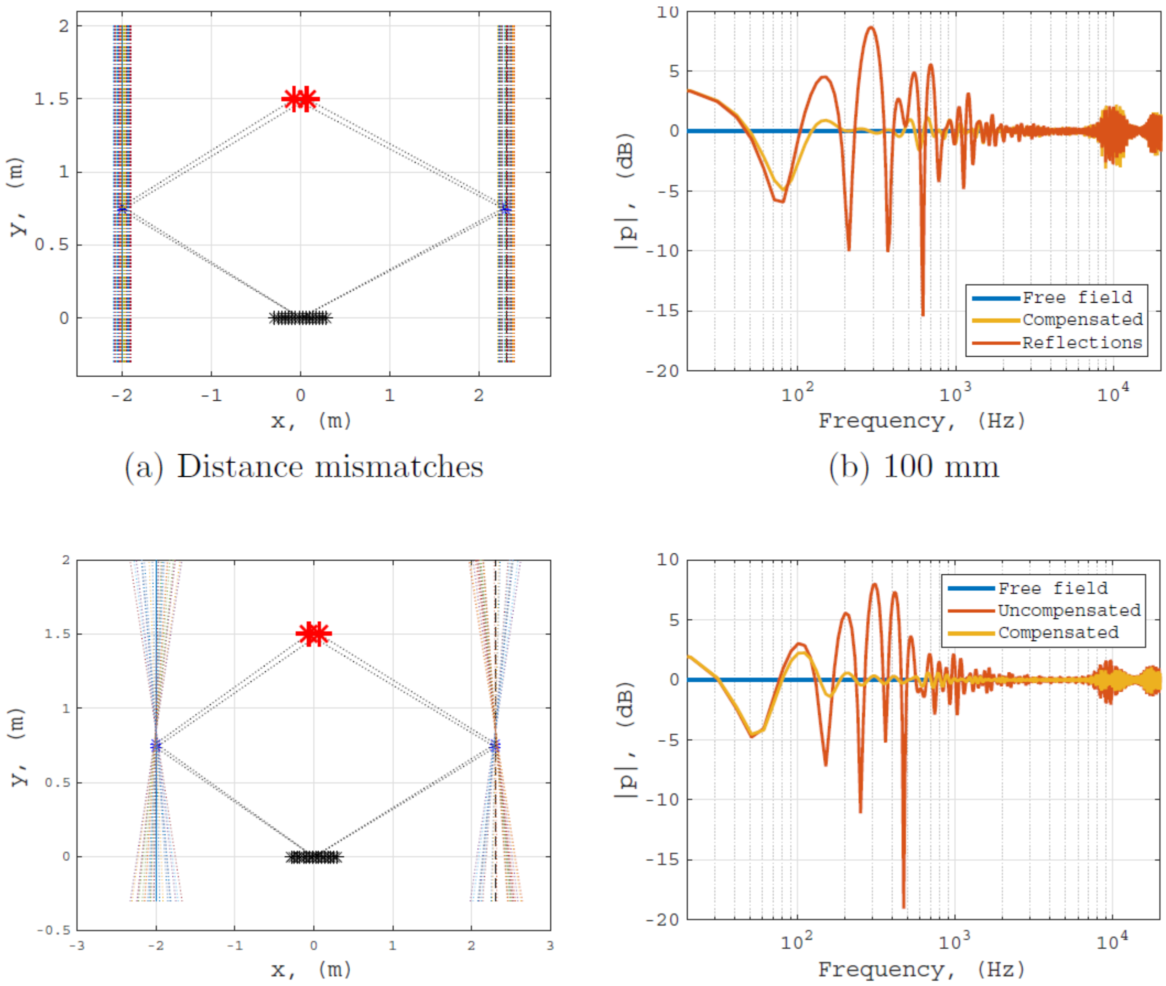
Figure 3: Cross talk cancellation and frequency responses under the effect of lateral reflections (“Uncompensated") and when these are compensated (“Compensated") compared with the free-field performance.

Note that in this case a room with parallel walls perpendicular to the loudspeaker array has been considered. However, the method proposed herein can be used with more complex geometries with lateral walls that are not perpendicular to the loudspeaker array.

4 Method Robustness

The simulations of previous section predicted a larger cross-talk cancellation and a smoother frequency response when using compensation for lateral reflections, given that the exact geometry of the room is introduced as an input parameter in Eq. 4 when calculating the control filters using the updated matrix of transfer impedances $\mathbf{C}_C$. In a practical set up, however, there will be small mismatches and errors in the exact walls position. To assess the robustness of the proposed formulation, a perturbation analysis is included here, in which the average performance of the array is calculated introducing errors in the position and orientation angle of the walls. A script was run 50 times and the results were averaged to obtain the on average performance of the formulation with respect to wall location errors.

The first type of error considered was that when an array is perpendicular to the lateral walls of a room and the distance with respect to both lateral walls was different to that used to calculate the compensated control filters in Eq. 4. The compensation was calculated for a given room geometry, and the reproduced acoustic pressure was calculated using a modified matrix $\mathbf{C}_\mathrm{N}$ in Eq. 10, in which errors with a maximum standard deviation of 100 mm in the walls lateral positions were introduced when calculating the reproduced pressure with Eq. 10. The results of this simulation are shown in Fig. 4c, wherein can be observed that when compensation is employed the reproduced response is more uniform. Fig. 4d shows the results for the case in which the orientation of both lateral walls with respect to the array was varied, considering errors with a maximum standard deviation of 10°. It can be observed that the result for the “Compensated" filters is also robust with respect to errors in the wall orientation and the reproduced response is more uniform than if using free-field filters.



(a) Distance mismatches (b) 100 mm

(c) Angle mismatches (d) 10°

Figure 4: Average performance of the array for errors with a standard deviation of 10 cm (top plots) and 10 degrees (left plots).

5 Practical Measurements

This section presents a set of practical measurements to assess the proposed formulation in a real-time application. To this end, a set up as this presented in Fig. 5 was disposed in an anechoic chamber, in which a Kemar binaural microphone was used to measure the reproduced response. To model the lateral walls, two reflectors were placed at the sides of the array, forming a set up as this used in the simulations. The reflectors were spaced 1.2 m to the left and 1.3 m to the right of the loudspeaker array.

Figure 5: Setup used for the practical measurements.

The response of two sets of filters was measured in the scenario of Fig. 5; a set of free-field filters, and a set of reflection compensation filters. The measured frequency and cross-talk cancellation responses are shown in Fig. 6. The cross-talk cancellation response shows an increase in performance of up to 5 dB below 1.6 kHz when compensation filters are used. The measured frequency response below 1 kHz is also smoother when compensation filters are used with respect to when free-field filters are used.

In order to further assess the efficiency of the formulation with a practical array, a perturbation analysis was carried out. Filters were calculated for a given wall geometry and then the walls distances from the array modified to introduce errors. These results are shown for the measured frequency and cross-talk cancellation responses under the label “Perturbations" in Fig. 6, wherein it is shown how the frequency response at the left ear is much smoother for the “Compensated" filters than for the “Uncompensated" filters. In the case of the cross-talk cancellation, the response is slightly better below 1.6 kHz.

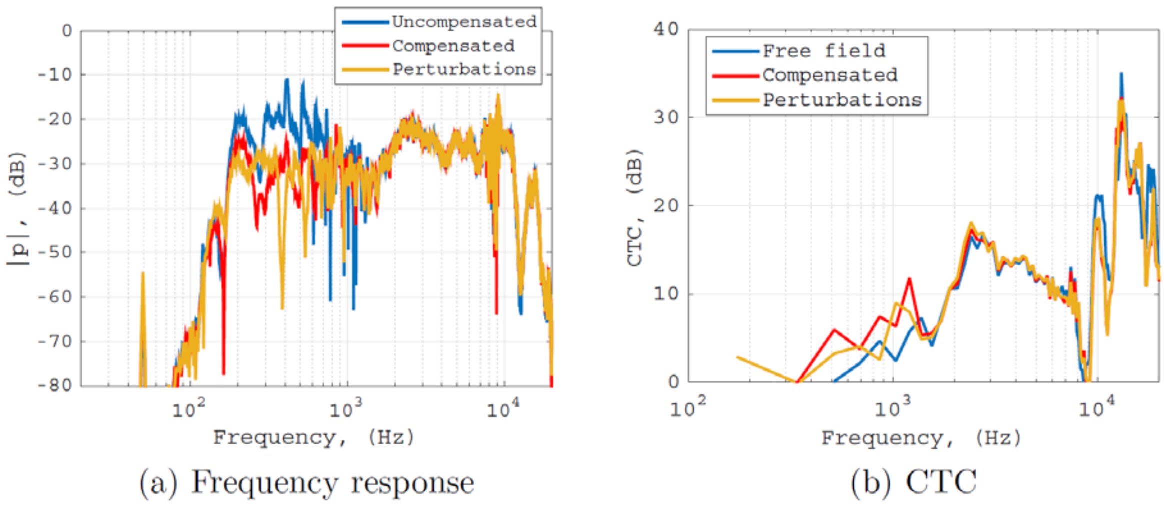
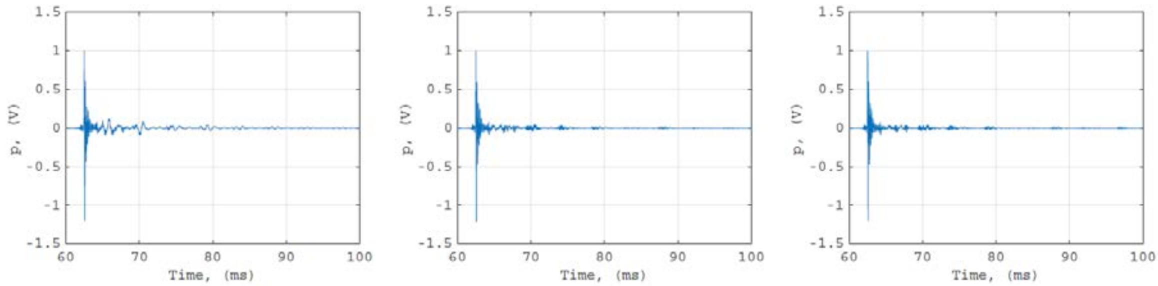


Figure 6: Measured frequency response and cross-talk cancellation performance in the setup of Fig. 5 for free-field filters (Uncompensated), compensation filters (Compensated) and compensated filters with modified wall distances to introduce errors (Perturbations).



(a) Free-field (b) Compensated (c) Perturbations

Figure 7: Measured impulse responses in the setup of Fig. 5 for free-field filters (Uncompensated), compensation filters (Compensated) and compensated filters with modified wall distances to introduce errors (Perturbations).

Fig. 7 shows the impulse responses of the array in the scenario of Fig. 5, calculated by inverse Fourier transformation of the measured frequency responses of Fig. 6. The impulse response corresponding to the free-field, uncompensated, case shows strong peaks which repeat with a certain cadence. When the radiation pattern of the array is compensated to account for first order reflections, these peaks are largely reduced, as shown in Fig. 7b. The impulse response corresponding to the perturbation analysis is shown in Fig. 7c, wherein it can be seen that even though the distance between the array and the walls was modified, the impulse response has a lower amount of reflections compared with the free-field case.

6 Conclusions

This paper has introduced a method for lateral reflections compensation for binaural reproduction over a linear loudspeaker array given the position of the listener and the loudspeaker array in a room. The method was first analysed by means of simulations that predicted an improvement in audio quality. A perturbation simulation has also predicted the method to be robust with respect to errors in the room geometry estimation.

The proposed formulation was implemented in real-time in a 16 loudspeakers line array and tested in an anechoic chamber with two reflectors representing the lateral walls. An informal listening test proved that by using this compensation the audio quality is improved. This formulation can be employed, for example, in small size smart loudspeaker arrays that adapt to the reproduction environment to give a better audio quality and reduce the effect of reflections.

Acknowledgements

The authors of the paper would like to acknowledge the support of the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership. The authors also thank Ferdinando Olivieri for his comments about the frequency dependent selection of the wall control points.

References

[1]D. B. Ward, “On the performance of acoustic crosstalk cancellation in a reverberant environment,” *The Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 1195–1198, 2001.
[2]D. Kosmidis, Y. Lacouture-Parodi, and E. Habets, “The influence of low order reflections on the interaural time differences in crosstalk cancellation systems,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 2873–2877.
[3]A. Sæbø, “Effect of early reflections in binaural systems with loudspeaker reproduction,” 1999.
[4] Takashi Takeuchi, Philip A Nelson and H. Hamada, “The effect of reflections on the performance of virtual acoustic imaging systems,” in *Proceedings of the Active 97 conference, Budapest, Hungary, 1997*, 1997.
[5]T. Betlehem and T. D. Abhayapala, “Theory and design of sound field reproduction in reverberant rooms,” *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2005.
[6]A. Canclini, D. Marković, F. Antonacci, A. Sarti and S.Tubaro, “A room-compensated virtual surround system exploiting early reflections in a reverberant room,” in *20th European Signal Processing Conference (EUPSICO 2012)*, Bucharest, Romania, 2012.
[7]S. Spors, H. Buchner, and R. Rabenstien, “A novel approach to active listening room compensation for wave field synthesis using wave-domain adaptive filtering,” in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP ’04). IEEE International Conference on*, vol. 4, May 2004, pp. iv–29–iv–32 vol.4.
[8]Marcos F. Simón Gálvez, Takashi Takeuchi and Filippo M. Fazi, “A listener adaptive optimal source distribution system for virtual sound imaging,” in *Audio Engineering Society Convention 140, Paris, France*, 2016.
[9]Marcos F. Simón Gálvez, S. J. Elliott, and J. Cheer, “A superdirective array of phase shift sources,” *The Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. 746–756, 2012.
[10] Marcos F. Simón Gálvez, S. J. Elliott, and J. Cheer “Time domain optimization of filters used in a loudspeaker array for personal audio,” *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 23, no. 11, pp. 1869–1878, Nov 2015.
[11] M. Shin, F. M. Fazi, P. A. Nelson, and F. C. Hirono, “Controlled sound field with a dual layer loudspeaker array,” *Journal of Sound and Vibration*, vol. 333, no. 16, pp. 3794 – 3817, 2014.
[12] F. Olivieri, F. M. Fazi, M. Shin, and P. Nelson, “Pressure-matching beamforming method for loudspeaker arrays with frequency dependent selection of control points,” in *Audio Engineering Society Convention 138*, May 2015.
[13] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small room acoustics,” *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
[14] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, Eds. plus 0.5em minus 0.4emJohn Wiley & Sons, 2000.
[15]Marcos F. Simón Gálvez and Filippo M. Fazi, “Loudspeaker arrays for transaural reproduction,” in *Proceedings of the 22nd International Congress on Sound and Vibration, Florence, Italy*, 2015.
[16] F. Fahy and J. Walker, *Advanced Applications in Acoustics, Noise & Vibrations*, F. Fahy and J. Walker, Eds. plus 0.5em minus 0.4emLondon: Spon Press, 2004.