24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

A LOUDSPEAKER ARRAY FOR 2 PEOPLE TRANSAURAL REPRODUCTION

Marcos Simón
Institute of Sound and Vibration Research, University of Southampton, SO17 1BJ, Southampton, United Kingdom
email: M.F.Simon-Galvez@soton.ac.uk

Filippo Fazi
Institute of Sound and Vibration Research, University of Southampton, SO17 1BJ, Southampton, United Kingdom

Transaural reproduction allows rendering 3D binaural audio through loudspeakers. Although this is a field that has been extensively studied for single-listener reproduction, reproduction of Transaural audio for more than one listener is still an open research question. This paper introduces a signal processing approach for performing 2-people Transaural reproduction using a combination of 2 single-listener cross-talk cancellation (CTC) beamformers, so that the CTC is maximised at one listener position and the beamformer side-lobes radiate little energy not to affect the other listening position. Off-line response predictions using measured transfer functions show a performance similar to that which would be obtained using more complex approaches, where the sound-field is controlled at the ears of both listeners simultaneously.

Keywords: Binaural, Transaural, Multilistener, Loudspeaker Arrays.

1. Introduction

Binaural audio allows for the reproduction of convincing 3D audio, relying on the fact that two independent signals are delivered to the two ears of one or multiple. This is of particular advantage with respect to other sound-field control methods, as it only requires an accurate reproduction on a small spatial area, whilst giving very convincing and immersive virtual acoustic cues. Nowadays, binaural audio is experiencing an increase in popularity thanks to the boost of virtual reality (VR) and augmented reality (AR) applications.

Although binaural audio is mostly consumed using headphones, binaural audio reproduction with loudspeakers, also termed *Transaural* [1], represents an alternative to headphone reproduction, and allows to perceive a very similar experience if the acoustical conditions are adequate [2]. The concept is not new, it was invented in the 1960 [3], but to this day it is still an area of active research. Transaural reproduction needs to cancel the cross-talk between left and right binaural signals at the opposite ears of a listener, which has lead to the adoption of the term cross-talk cancellation (CTC) as a synonym of systems for binaural reproduction with loudspeakers. Hence, the larger the channel separation between the reproduced pressure at both ears of a listener, the better the performance of a CTC system. For a long time, scientists and engineers have tried to achieve this effect using 2 loudspeakers [4, 5, 6]. However, 2-loudspeaker based control approaches showed to cause a strong colouration on the reproduced signal, which lead to the development of more complex speaker geometries [7, 8].

Whilst new, more complex, geometries offer a much improved sound quality, a drawback that Transaural systems have suffered for a long time from is their very narrow sweet-spot, that is dependent on the listener's head position and orientation [9]. To solve this problem, researchers have proposed to adapt the sweet-spot using head-tracking techniques and continuous updating of the loudspeakers CTC filters [10, 11, 12].

Transaural systems have been largely studied for performing single-listener reproduction, but little research has been carried out on 2-people or multi-listener Transaural systems. As 2-people CTC requires a minimum

of four loudspeakers, initial research was focused on the optimisation of the loudspeaker positions [13, 14] to increase the performance of the sound-field control for both listeners.

This paper introduces a novel real-time implementation for the simultaneous rendering of binaural material for 2 listeners using a linear loudspeaker array. Although the idea of using a head-tracked line array for 2-people reproduction has been attempted beforehand, earlier studies showed difficulties adapting the array response for off-axis listening positions [15]. To overcome this problem, this paper proposes a new approach in which 2 independent single-listener (2-point) CTC-beamformers are combined, minimizing the processing complexity with respect to that required by a 4-point CTC beamformer, and in which the beamformers are created using fully-adaptive dynamic CTC techniques [16]. The structure of the paper is as follows: Section 2 reviews the listener adaptive theory, Section 3 shows a prediction of performance using measured transfer functions from a 28 loudspeaker line array on an anechoic chamber and Section 4 introduces the real-time implementation of the system.

2. Theory

A diagramatic representation of a 2-listener, linear-array-based CTC system is depicted in Fig. 1.Considering a Cartesian coordinates system, $\mathbf{x} = (x_1, x_2, x_3)$, Fig. 1 shows a loudspeaker array of L radiators with coordinates $\mathbf{y}_l = (y_{1l}, y_{2l}, 0)$ and whose radiated acoustic pressure field is controlled at 4 points in the space, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$. The notation employed along the document is that $p(\mathbf{x}_m, j\omega) = p$ is the radiated pressure at a point with coordinates $\mathbf{x}_m = (x_{1m}, x_{2m}, 0)$ with $\omega = 2\pi f$ being the radiating frequency. The acoustic pressures at the field points corresponding to the 2 listener's ears are defined as p_1 and p_2, for the first listener (1 in Fig. 1), and p_3 and p_4, for the second listener (2 in Fig. 1).

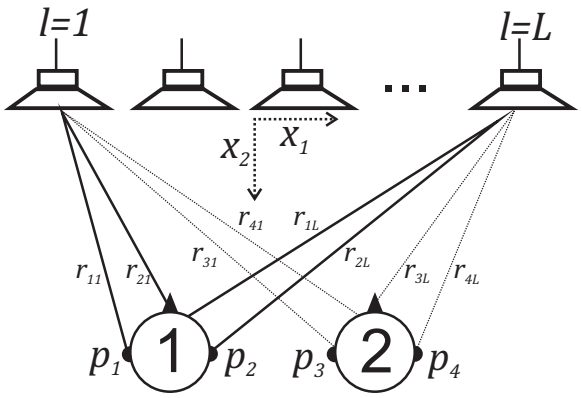


Figure 1: Control geometry assumed for 2-people CTC.

The transfer functions between the array loudspeakers and the four pressure control points are contained in a matrix, $\mathbf{C}$, which is defined as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix}, \tag{1}$$

where

$$\mathbf{C}_1 = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}, \text{ and } \mathbf{C}_2 = \begin{bmatrix} \mathbf{c}_3 \\ \mathbf{c}_4 \end{bmatrix}. \tag{2}$$

The formulation used here assumes that each loudspeaker behaves as an ideal point-monopole source radiator, and hence, for each control point, a vector $\mathbf{c}_m$ is formed by $\mathbf{c}_m = \left[c_{m1}e^{-jkr_{m1}}, \cdots, c_{mL}e^{-jkr_{mL}}\right]$, where an $e^{j\omega t}$ time dependence is assumed with $k = \omega/c_0$ and c_0 being the speed of sound. The quantity $c_{ml} = 1/r_{ml}$ is an attenuation factor, with $r_{ml} = \|\mathbf{x}_m - \mathbf{y}_l\|$.

In the proposed two sets of single-listener CTC-beamforming filters, $\mathbf{H}_1$ and $\mathbf{H}_2$ are created, each of which controls only two of the four pressure points. The CTC-beamforming filters $\mathbf{H}_1$ will maximise the difference of acoustic pressure between p_1 and p_2 and the set of filters $\mathbf{H}_2$ between p_3 and p_4. The set of filterss for the

first and second single-listener CTC-beamformers are obtained by solving the least-squares inverse problems given by

$$\mathbf{H}_1 = \mathbf{C}_1^{\mathrm{H}} \left[\mathbf{C}_1\mathbf{C}_1^{\mathrm{H}} + \psi\mathbf{I}\right]^{-1}, \tag{3}$$

for the first single-listener, and

$$\mathbf{H}_2 = \mathbf{C}_2^{\mathrm{H}} \left[\mathbf{C}_2\mathbf{C}_2^{\mathrm{H}} + \psi\mathbf{I}\right]^{-1}, \tag{4}$$

for the second single-listener, where ψ is a regularisation parameter used to control the amount of electric power used by the array filters [17]. The full explanation of how these CTC-beamformers can be implemented in a, dynamic, listener-position-adaptive manner can be found in [16, 18].

A vector of reproduced signals at the ears of both listeners, $\mathbf{p}$, is obtained by combining the outputs of both CTC-beamformers, so that

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = \mathbf{p}_1 + \mathbf{p}_2, \tag{5}$$

where

$$\mathbf{p}_1 = \mathbf{C}\mathbf{H}_1\mathbf{v} \text{ and } \mathbf{p}_2 = \mathbf{C}\mathbf{H}_2\mathbf{v}, \tag{6}$$

where $\mathbf{v}$ is the matrix of left and right binaural signals to be delivered to the listener's ears, written as

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \tag{7}$$

In this case, Eq. 5 allows to monitor the pressures created by the two single-listener CTC-beamfomers.

Alternatively, filters can be created by controlling the reproduce sound-field at the 4 pressure control points. This h is obtained by incorporating the matrix of transfer functions $\mathbf{C}$ when calculating the filter set, so that

$$\mathbf{H} = \mathbf{C}^{\mathrm{H}} \left[\mathbf{C}\mathbf{C}^{\mathrm{H}} + \beta\mathbf{I}\right]^{-1}. \tag{8}$$

An analysis of the performance that is achieved by using filter sets created according to Eqs. 3 and 4 or to Eq. 8 is shown in the next section.

3. Predictions of free-field performance

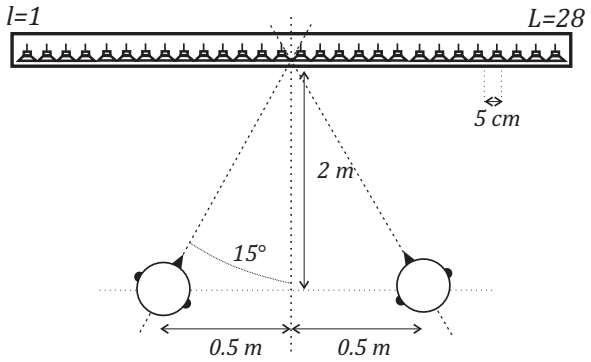
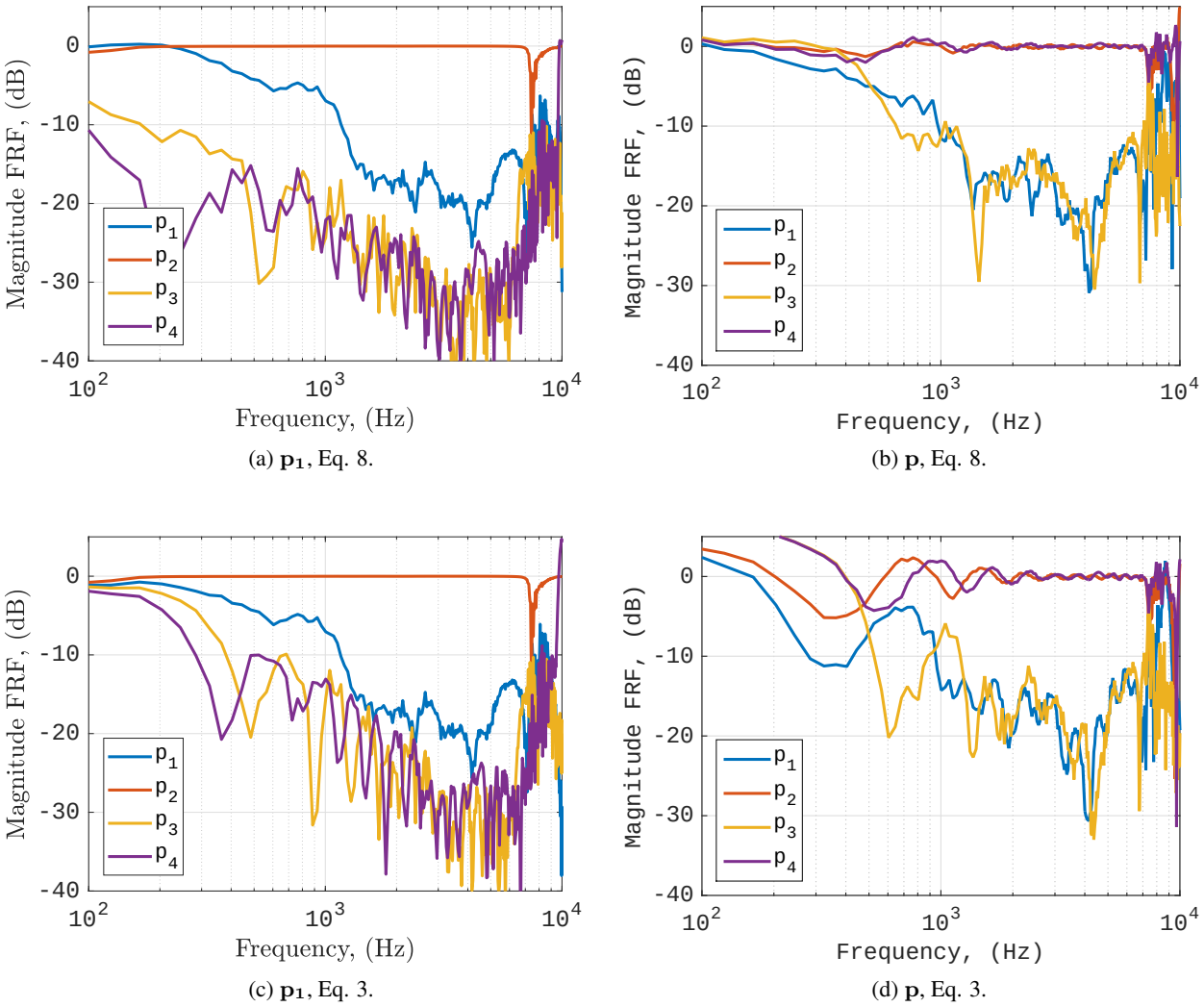


Figure 2: Control geometry used for the performance off-line simulations.

The performance of the formulation presented in Section 2 was simulated off-line using the transfer functions from a 28 loudspeaker array with an inter element distance $d = 5cm$, as that shown in Section 4. The transfer functions were measured using a Neumann KU-100 binaural microphone. The loudspeaker array was driven with an Innosonix MA32/LP amplifier and a Ferrofish AC-16 MKII was used as input soundcard. The loudspeaker transfer functions were measured using sine sweeps and deconvolved using the commercial

software Matlab®. The loudspeaker array and binaural microphone placement is sketched in Fig. 2. The loudspeaker array was placed 2 m away from two listeners separated by 1 m, in a symmetrical configuration with respect to the central plane that divides the loudspeaker array, with each listener forming and angle of 15° from such plane.



(a) $\mathbf{p_1}$, Eq. 8.

(b) $\mathbf{p}$, Eq. 8.

(c) $\mathbf{p_1}$, Eq. 3.

(d) $\mathbf{p}$, Eq. 3.

Figure 3: Predicted responses at the 4 pressure control points representing the ears of the listeners from Fig. 2 geometry, for the cases when sound is only beamed to the control point 2 ((a) and (c)) and when acoustic pressure is beamed at control points 2 and 4 ((b) and (d)).

Fig. 3 shows the predictions of array response at the control points of Fig. 2, to illustrate the channel separation that can be obtained at the listeners' ears. The results are shown using $v_1 = 0$ and $v_2 = 1$ along the whole frequency range, so that the pressure is maximised at the listeners' right ears. The responses, predicted using the formulation of Section 2 are compared with simulations using a single CTC-beamformer in which the 4 pressure control points are controlled using filters created according to Eq. 8.

The left hand side results of Fig. 3 show the effect of beaming acoustic energy to only the right ear of listener 1. The top results (Fig. 3a) were calculated according to Eq. 8. These results show that the reproduced pressure at p_2 presents an almost flat spectrum, with the response being lower at the contralateral ear, p_1 and largely reduced at p_3 and p_4. The bottom plot (Fig. 3c) was calculated according to Eq. 3. For this control configuration, the responses of p_3 and p_4 show a similar level to that of p_2, as in this case p_3 and p_4 are not taken in account by Eq. 3.

The right hand side of Fig. 3 shows the results when both the output of $\mathbf{p}_1$ and $\mathbf{p}_2$ are combined, as in Eq. 5, again with $v_2 = 1$ and $v_2 = 0$, so that the acoustic pressure is maximised at control points 2 and 4. These plots represent a common listening situation, in which acoustic energy is beamed to both listeners right ears. These results show that the channel separation is mostly determined by that obtained at the different ears of each listener, p_1 *vs.* p_2 and p_3 *vs.* p_4. The particularity of the outcome is that the channel separation between using the set of filters $\mathbf{H}$ (Eq. 8) or $\mathbf{H}_1$ and $\mathbf{H}_2$ (Eqs. 3 and 4) is practically similar above 1 kHz. Below 1 kHz, the responses at p_2 and p_4 have about 5 dB of ripple. Such ripple is caused by the phase variations of the side-lobes from the other listener's CTC-beamformer, causing a large boost below 500 Hz. The responses shown in Fig. 3b, however, show an almost flat level along frequency, as the side-lobes are also controlled at the other listener's position along the whole frequency range. Nevertheless, the results presented in Fig. 3d encourage the use the proposed formulation to perform 2 people listening reproduction, provided special attention is put to equalise the response below 1kHz.

4. Real-time 2 people system

The formulation shown in Section 2 was implemented in MAX MSP 6 on a computer running Windows 8.1, and a Microsoft Kinect 2 was employed to perform the head-tracking of the two listeners. A block diagram of the implementation can be observed in Fig. 4. The computer receives information from the Kinect on the listeners' position and modifies the output of both single-listener CTC-beamformers, so that the CTC is maximised between the ears of each listener. A binaural audio source is connected to the input of the computer, with the array output being sourced by an Innosonix MA 32/LP amplifier and all the audio connections between the different devices carried out via MADI.

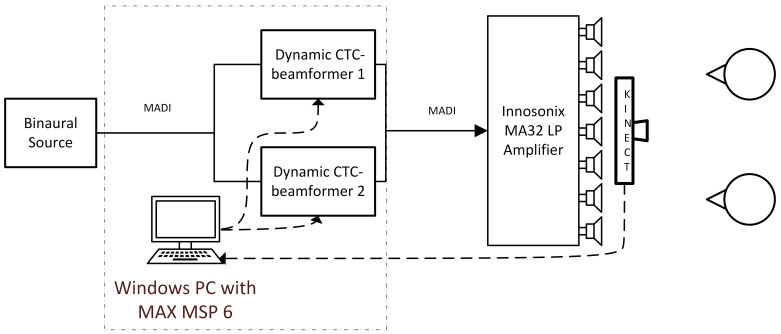


Figure 4: Schematic of the real-time implementation of the 2-people transaural reproduction array.

The 28 loudspeaker array used for the real-time implementation can be observed in Fig. 5, where the image shows two listeners observing a clip with binaural audio and both are immersed in a 3D binaural field.

Figure 5: Technology demonstration with two simultaneous listeners.

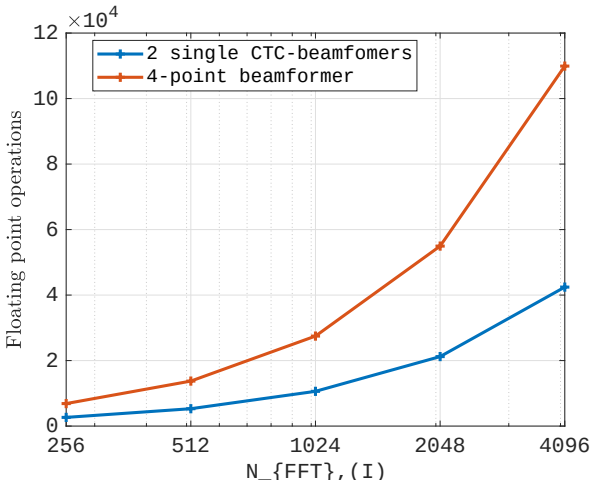


Figure 6: Technology demonstration with two simultaneous listeners.

A complexity analysis is presented that assesses the number of floating point operations required per listener-position update, comparing the proposed approach using a set of two single-listener CTC-beamformers, which require 2 inversions of 2×2 sized matrices, with the use single 2-listener (4-control point) CTC-beamformer, which requires the inversion of a 4×4 matrix. Every time one of the two listeners moves, the control filters have to be calculated an updated. This analysis is estimated in the basis of how many arithmetical operations are needed for a matrix inverse of size $N \times N$, which requires a total of $N^(2.373)$ [19]. These results, shown in Fig. 6, have been calculated for various N_{FFT} sizes. The results, plotted on a logarithmic x axis show a large complexity reduction of the proposed approach, which increases proportional to the N_{FFT} size.

5. Conclusion

This manuscript has introduced a signal processing scheme for 2-people transaural audio reproduction with a loudspeaker array of an arbitrary number of loudspeakers. The presented formulation is based on the linear superposition of two sub-systems, each of which controls only 2 of the 4 pressure control points corresponding to both ears of a single listener. This reduces the processing complexity with respect to other approaches in which the 4 pressure control points are controlled simultaneously.

Loudspeaker array response predictions using measured transfer functions have shown a similar behaviour to that using a 4-point control approach, especially in the upper part of the frequency range, where the sound beams created by the array are very narrow. This encourages the use of the proposed algorithm for performing 2 people transaural reproduction, although this shows, however, a non completely uniform frequency response at low frequencies. Further work will investigate techniques to improve the uniformity of the reproduced response along the whole frequency range.

6. Acknowledgement

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. No new data was generated in this work.

References

1. W. G. Gardner, “3-D Audio Using Loudspeakers,” Ph.D. dissertation, Massachusetts Institute of Technology, 1997.

2. D. Kosmidis, Y. Lacouture-Parodi, and E. A. P. Habets, “The influence of low order reflections on the interaural time differences in crosstalk cancellation systems,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 2873–2877.

3. S. Atal and R. Schroeder, “Apparent sound source translator,” Patent, Feb. 22, 1966, uS Patent 3,236,949. [Online]. Available: http://www.google.co.uk/patents/US3236949

4. D. B. Ward and G. Elko, “Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation,” *Signal Processing Letters, IEEE*, vol. 6, no. 5, pp. 106–108, May 1999.

5. 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. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12148

6. O. Kirkeby, P. A. Nelson, F. Orduña Bustamante, and H. Hamada, “Local sound field reproduction using digital signal processing,” *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1584–1593, 1996.

7. J. Bauck, “A simple loudspeaker array and associated crosstalk canceler for improved 3d audio,” *J. Audio Eng. Soc*, vol. 49, no. 1/2, pp. 3–13, 2001.

8. T. Takeuchi and P. A. Nelson, “Optimal source distribution for binaural synthesis over loudspeakers,” *The Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 2786–2797, 2002.

9. Y. L. Parodi and P. Rubak, “Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakersa),” *The Journal of the Acoustical Society of America*, vol. 128, no. 3, pp. 1045–1055, 2010. [Online]. Available: http://scitation.aip.org/content/asa/journal/jasa/128/3/10.1121/1.3467763

10. M. A. Casey, W. G. Gardner, and S. Basu, “Vision steered beam-forming and transaural rendering for the artificial life interactive video environment (alive),” in *Audio Engineering Society Convention 99*, Oct 1995. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7714

11. T. Lentz and O. Schmitz, “Realisation of an adaptive cross-talk cancellation system for a moving listener,” in *Audio Engineering Society Conference: 21st International Conference: Architectural Acoustics and Sound Reinforcement*, Jun 2002. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=11175

12. Marcos F. Simón Gálvez, Takashi Takeuchi and Filippo Maria Fazi, “A listener adaptive optimal source distribution system for virtual sound imaging,” in *Audio Engineering Society Convention 140, Paris, France*, 2016.

13. Y. Kim, O. Deille, and P. Nelson, “Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners,” *Journal of Sound and Vibration*, vol. 297, no. 1, pp. 251 – 266, 2006.

14. B. Masiero, “Source positioning in a two listener crosstalk cancellation system,” in *Proceedings of NAG-DAGA Rotterdam*, 2009.

15. M. H. Hiroaki Kurabayashi, Makoto Otani and M. Kayama, “Development of dynamic crosstalk cancellation system for multiple-listener binaural reproduction.”

16. Marcos F. Simón Gálvez and Filippo M. Fazi, “Sweet-spot-independent binaural reproduction with a listener-adaptive loudspeaker array,” in *Proceedings of the International Congress on Acoustics, Buenos Aires, Argentina*, 2016.

17. Marcos F. Simón Gálvez, Stephen J. Elliott and Jordan Cheer, “Personal audio loudspeaker array as a complementary tv sound system for the hard of hearing,” *IEICE Trans. Fundamentals.*, vol. E97(9), 2014.

18. Marcos F. Simón Gálvez and F. M. Fazi, "Listener adaptive filtering strategies for personal audio reproduction over loudspeaker arrays," in *Audio Engineering Society Sound Field Control conference, University of Surrey, Guildford, United Kingdom*, 2016.

19. T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms, 3rd Ed.* MIT press, Cambrdige, MA, 2009.