



LOUDSPEAKER ARRAYS FOR TRANSAURAL REPRODUCTION

Marcos F. Simón Gálvez and Filippo Maria Fazi

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

email: M.F.Simon-Galvez@soton.ac.uk

Transaural rendering allows for the reproduction of binaural audio material without the need of wearing headphones. This is achieved by the use of cross talk cancellation systems, which are generally implemented with stereo loudspeaker pairs. The robustness and quality of the crosstalk cancellation depends on the frequency of the reproduced sound and on the source span of the stereo set up. As this limits the frequency bandwidth where effective crosstalk cancellation can be achieved, nested stereo systems with different source spans have been proposed as a method to enhance the bandwidth of the crosstalk cancellation. This technique is compared in this paper to a linear loudspeaker array of 16 uniformly sources used as a crosstalk cancellation system. The performance of the uniform array is first compared to that of a nested stereo array by means of off-line simulations. The robustness of the crosstalk cancellation is analysed with respect to room reflections and reverberation. This effect is also studied by investigating the total acoustic energy produced by the linear array, in comparison with that produced by the nested stereo array.

1. Introduction

Binaural audio reproduction allows for the creation of stable virtual audio images. A binaural signal contains spatial information such as interaural level difference (ILD) and interaural time difference (ITD) [1]. The binaural signals can be obtained by recording an audio source with a dummy head, or can be also synthesised by combining an audio track of a mono source with the head-related impulse responses containing the ILD and ITD of a given incoming direction, so that the mono source can be placed, ideally, everywhere in the space. Ideally, binaural reproduction can produce the same signals at the ears as those experience if the listener were at the live event.

Transaural reproduction allows for the reproduction of binaural signals through loudspeakers, hence without the need of wearing headphones. Transaural reproduction is based on the principle of cross talk cancellation, which reduces the cross talk of the signal intended to reproduce in the *ipsilateral* (same side) ear in the *contralateral* (opposite side) ear. This method was first proposed by Atal and Schroeder [2]. After this invention, transaural reproduction has been largely studied by many researchers as for example by Cooper and Bauck [3] and later by Kirkeby *et al.* [4, 5], with both sets of authors using two loudspeakers. After that, transaural reproduction has evolved towards more robust reproduction by including regularisation [6]. The robustness with respect of the required loudspeaker

span was also studied by Ward and Elko[7], who realised that in order to provide a good cross talk at lower frequencies, a larger span was needed. These was later addressed by using special loudspeakers arrangements which allow for an uncoloured audio reproduction of the whole audio spectrum, as for examples those presented by Bauck [8] or by Takeuchi [9].

One key limitation of conventional cross talk cancellation systems arise from the fact that listener movements exceeding 75-100 mm may completely destroy the desired spatial effect[10, 11]. The use of a loudspeaker array of more than two sources can be beneficial with respect to a set up using just two loudspeakers, as it can allow for a greater cross talk cancellation and be less sensitive to errors [12]. This has been implemented previously in the way of a circular array surrounding the head of a listener [13], in a linear array with the sources spaced to maximise the cross talk response along frequency [14]. Line arrays of large number of loudspeakers surrounding a TV have been also used to reproduce 22.2 multichannel sound [15]. Loudspeaker arrays have also been used to provide binaural material to more than a single listener simultaneously [16].

This paper analyses the use of a loudspeaker array of 16 sources for single listener transaural reproduction, compared with a three-way optimal source distribution (OSD) array [9, 17]. The main intention of the paper is to investigate whether the use of an array can provide a better cross talk cancellation performance for a single listener in normal environments. To this end, simulations modelling the array sources as point monopoles are performed, first in the free field in Section 2, and then using a model of the reverberant field in Section 3.

1.1 Optimal Source Distribution Nested Arrays

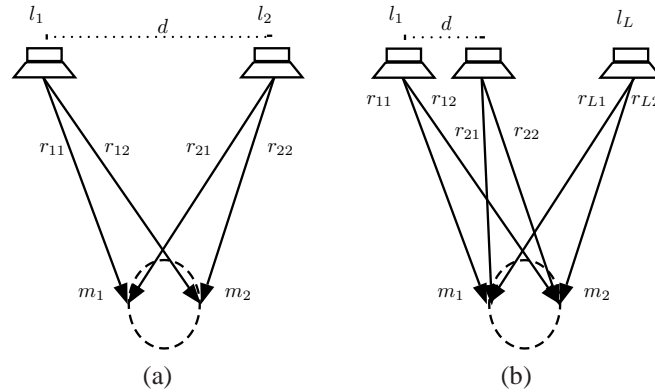


Figure 1: Geometry of a stereo dipole (a) and of a loudspeaker array (b), where the $M=2$ microphones simulate the ears of a listener.

An example of transaural reproduction using cross talk cancellation is introduced in Fig. 1a. The system comprises $L = 2$ control sources, l_1 and l_2 , driven by signals $\mathbf{v} = [v_L(j\omega), v_R(j\omega)]^T$ respectively. This produces a soundfield in the $M = 2$ microphones m_1 and m_2 which respectively captures pressure signals $\mathbf{p} = [p_L(j\omega), p_R(j\omega)]^T$ at left and right ears of the listener. The relation between the control loudspeakers and the microphones simulating the ears of the listener can be written as

$$(1) \quad \mathbf{p} = \mathbf{C}\mathbf{v},$$

where \mathbf{C} is the matrix of plant transfer functions, which based on the geometry of Fig. 1a and assuming that each source behaves as a point monopole source can be written as

$$(2) \quad \mathbf{C} = \frac{\rho_0}{e\pi} \begin{bmatrix} \frac{e^{-jkr_{11}}}{r_{11}} & \frac{e^{-jkr_{12}}}{r_{12}} \\ \frac{e^{-jkr_{21}}}{r_{21}} & \frac{e^{-jkr_{22}}}{r_{22}} \end{bmatrix},$$

where $k = \omega/c_0$ is the wavenumber, $\omega = 2\pi f$ is the radiating frequency and c_0 is the speed of sound in air. A time convention $e^{j\omega t}$ is used. As observed in the geometry of Fig. 1a r_{11} and r_{12} are the distance from the loudspeaker 1 and 2 to the left microphone (microphone 1) and r_{21} and r_{22} are the distance from loudspeaker 1 and 2 to the right microphone (microphone 2).

The binaural signals that are to be synthesised at the receivers are defined by the elements of a complex vector $\mathbf{d} = [d_L(j\omega), d_R(j\omega)]^T$. In order to reproduce those signals at each receiver, a filter matrix containing the cross talk cancellation filters, \mathbf{H} , is introduced so that $\mathbf{v} = \mathbf{H}\mathbf{d}$. This matrix is defined as follows

$$(3) \quad \mathbf{H} = \begin{bmatrix} H_{LL}(j\omega) & H_{LR}(j\omega) \\ H_{RL}(j\omega) & H_{RR}(j\omega) \end{bmatrix},$$

which allows to express the pressure at the ears as

$$(4) \quad \mathbf{p} = \mathbf{C}\mathbf{H}\mathbf{d}$$

For a *fully determined* control system, as the case of a single listener cross talk cancellation system using an stereo set up, a set of filters can be obtained by direct inversion of the matrix of plant transfer functions, which in the case of a 2×2 matrix can be computed analytically. However, to achieve a more stable system, regularisation can be included. In this case the source filters to obtain the required cross talk cancellation system are given by

$$(5) \quad \mathbf{H} = [\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}] \mathbf{C}^H,$$

where β is a regularisation parameter, which allows to control the energy used by the control filters [18]. Regularisation can also be used to increase the robustness of an array to small mismatches in the loudspeaker transfer functions [19].

1.2 Multichannel Loudspeaker Array

A control geometry containing a loudspeaker array can be observed in Fig. 1b. In this case the array uses L loudspeakers and there are $M = 2$ control microphones, corresponding to the ears of a listener. As the number of control loudspeakers is larger than the number of control sources, $L > M$, the matrix of plant transfer functions is not square, which requires to calculate the pseudo inverse of the matrix instead [20]. If regularisation is used, the loudspeaker feeds are obtained in this case by

$$(6) \quad \mathbf{H} = [\mathbf{C}^H \mathbf{C} + \beta \mathbf{I}] \mathbf{C}^H,$$

wherein this case \mathbf{H} is a $L \times M$ matrix.

1.3 Performance metrics

In order to analyse the cross talk cancellation performance of both arrays two metrics are introduced. The first metric is used to assess the channel separation along frequency, which is performed by including the crosstalk matrix, \mathbf{R} , defined by

$$(7) \quad \mathbf{R} = \mathbf{H}^H \mathbf{C}^H \mathbf{C} \mathbf{H} = \begin{bmatrix} |R_{LL}(j\omega)|^2 & |R_{LR}(j\omega)|^2 \\ |R_{RL}(j\omega)|^2 & |R_{RR}(j\omega)|^2 \end{bmatrix},$$

The ratio between the elements of this matrix define the cross talk cancellation spectrum, $\psi(j\omega)$, which for the case of a symmetrical listening situation is given by

$$(8) \quad \psi(j\omega) = \frac{|R_{LL}(j\omega)|^2}{|R_{RL}(j\omega)|^2} = \frac{|R_{RR}(j\omega)|^2}{|R_{LR}(j\omega)|^2}.$$

Apart from this metric, it is also important to characterise how much energy the cross talk cancellation filters require, as often they need large boosts of acoustical energy to control the soundfield at frequencies at which the system is not well conditioned [9]. To this end, a metric known as array effort (AE) is introduced here [18]. The array effort is defined as the norm of the control filters, divided by the norm input signal, h_S , that a single loudspeaker requires to obtain the same pressure as that produced by the cross talk system in a given ear. The normalised array effort is thus defined as

$$(9) \quad AE = \frac{\sum_{l=1}^L (|H_{l1}(j\omega)|^2 + |H_{l2}(j\omega)|^2)}{|h_S|^2}.$$

This quantity is proportional to the amount of electric power employed to maximise the response in one ear and reduce it on the other ear, assuming the electroacoustic interaction between the transducers of the array is negligible. The magnitude of the array filters can be controlled by constraining the array effort to be lower than a given value at each frequency, which is achieved by varying the regularisation parameter, β . By limiting the array effort, ill-conditioning with respect to the inversion of the propagation matrix is also avoided, and so the array is made more robust to changes in the environment [19]. Array effort and acoustic contrast are dimensionless quantities, whose levels are typically plotted in decibels.

2. Free field performance

The performance of both the loudspeaker array (LA) and the OSD cross talk cancellation systems is assessed here by means of crosstalk cancellation performance system in the same control geometry. The OSD array is divided in three frequency bands, low, medium and high. In a practical implementation of an OSD array the input is first filtered in three different bands (using a low-pass, a band-pass, and a high-pass filter), with the three filtered versions of the input reproduced by the corresponding array. The separation between the loudspeakers of the OSD arrays are 120 cm for the low frequency channel, 30 cm for the medium frequency channel and 8 cm for the high frequency channel. The loudspeakers of the LA are separated by 8 cm. The arrays were placed at a distance of 2 m from the microphones representing the listener ears, with the ears of the listener separated by 18 cm. The cross talk performance of both the LA and the OSD loudspeaker array was calculated by limiting the array effort below 10 dB, as it would be done in a practical situation in order to prevent loudspeaker overdrive [21].

The reproduced pressures in the ipsilateral and contralateral ears, the cross talk spectrum, ψ , and the array effort are shown for the LA and for the OSD array in Fig. 2. It can be observed how the response is limited at low frequency. This is due to the use of regularisation for the creation of the filters defined in Equations 5 and 6. This requires a low frequency boost, which is obtained by applying a gain to the filters so that the frequency response at the ipsilateral ear is flattened. The effect can be observed in Fig. 2a. After this equalisation the directional response is the same, but the audio quality is better as there is less colouration.

Fig. 2c shows the cross talk cancellation performance. The low frequency response for both arrays is reduced, due to the use of regularisation. The LA is able to obtain a higher cross talk cancellation starting at 150 Hz, whilst the OSD starts to be effective at about 250 Hz. The performance of the LA is better than that of the OSD for the whole frequency range, with the exception of certain frequencies at which the performance of the OSD is better. The sound radiation patterns, shown in the right hand side of the figure, show how the LA is able to produce a very narrow radiation pattern. The radiation pattern of the OSD shows how, throughout the whole of the frequency range, the control filters create a null in the contralateral ear. This is also how the LA obtains the cross talk cancellation at low fre-

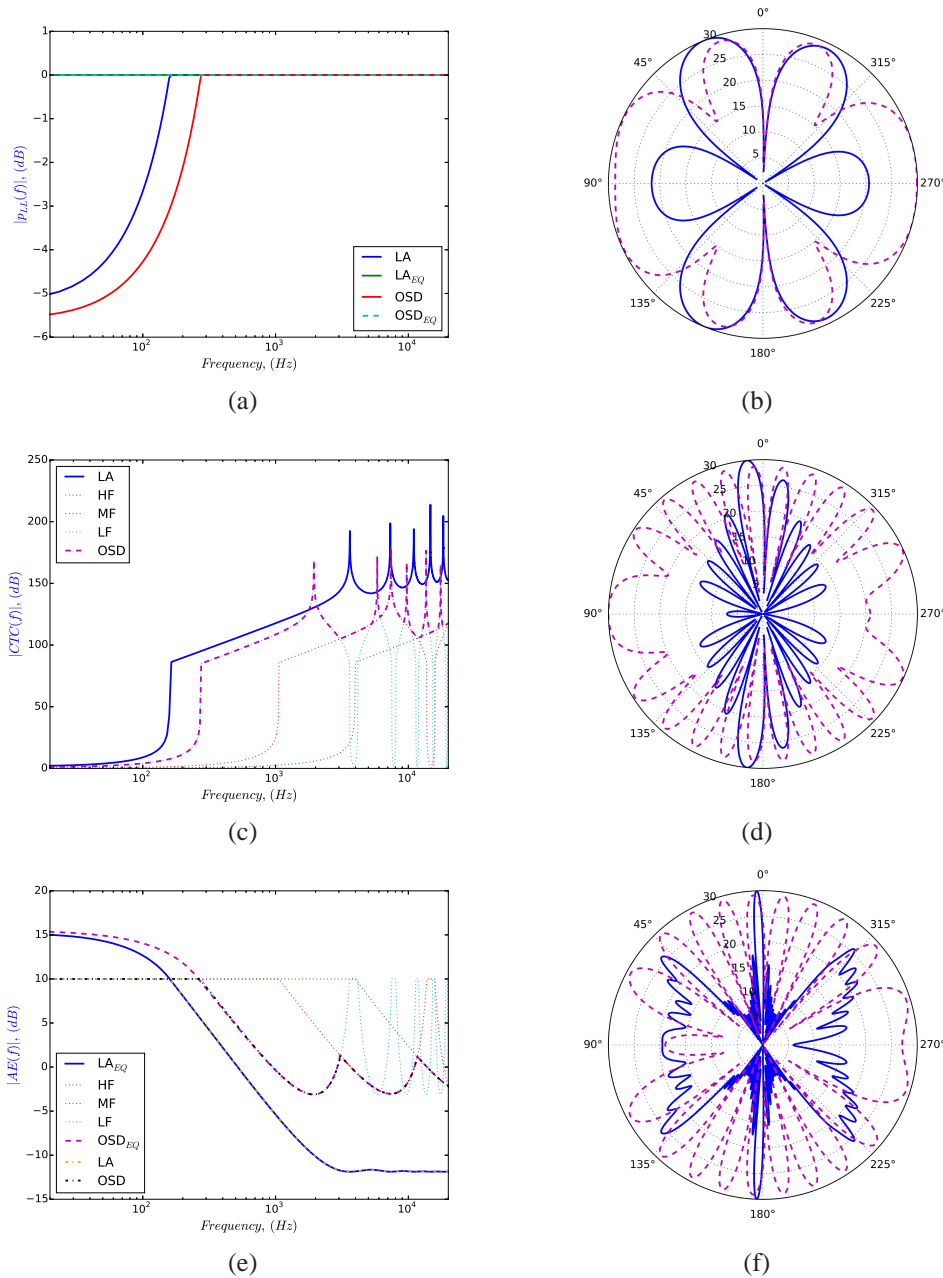


Figure 2: Pressure response at the ipsilateral ear, (a), cross talk cancellation performance, (b), and array effort simulated for the LA and for the OSD array, (c). The right hand side plots show the radiation patterns for the LA (solid line) and the OSD array (dashed line) at 500 Hz, (d), 1.5 kHz, (e), and 6 kHz, (f).

quencies, however, it acts similar as a delay and sum array at higher frequencies, beamforming at the ipsilateral ear.

The array effort obtained with the cross talk cancellation filters of both arrays is shown in Fig. 2e. The array effort for the LA and for the OSD array has been limited by using regularisation to be below 10 dB at every frequency. The filters were then equalised to produce a flat frequency response in the ipsilateral ear. After equalisation is applied, it can be seen how both systems require a larger boost at low frequency, which grows until 15 dB at 100 Hz. The OSD array requires about 1 dB more at low frequency than the LA to obtain the same pressure at the ear. Note that the array effort is a measure of the total energy, hence the average energy of the signal driving each loudspeaker is proportional to AE/L . This suggest that the OSD requires larger loudspeaker signals.

3. Simulated reverberant performance

Using the method presented in [22], it is possible to simulate the reverberant performance of a cross talk cancellation system in a room, based on the free-field radiation pattern of the device and the acoustic characteristics of the room. The acoustic power radiated by each of the two cross talk cancellation systems, $W(j\omega)$, can be estimated by sampling the pressure in a surrounding sphere of radius r [23], which can be written as

$$(10) \quad W(j\omega) = \frac{r^2}{\rho_0 c_0} \sum_{m=1}^{2\pi/\Delta_\theta} \sum_{n=1}^{\pi/\Delta_\phi} |p^2(j\omega, \phi_n, \theta_m, r)| |\sin \phi_n| \Delta_\phi \Delta_\theta,$$

where $\Delta_\theta = 2\pi/N_H$ and $\Delta_\phi = \pi/N_V$ represent the angle in radians between each horizontal and vertical measurement point, where N_H is the number of horizontal measurements and N_V is the number of vertical measurements. As both arrays have a symmetrical radiation pattern with respect to the z axis, the radiated acoustic power has been estimated for both arrays using a semicircular measurement slide of 180 point microphones, hence obtaining an accuracy of 1 degree. The acoustic power has been measured at a distance of 3 m from both radiators. The results of the simulation are shown in Fig. 3a. At low frequencies the LA produces a larger acoustic power which increases until about 150 Hz, where it decreases due to the increase in directivity of the device. The radiated power of both arrays decrease, with the LA radiating about 10 dB less of power than the OSD array thanks to the increased directivity obtained by the larger number of sources used.

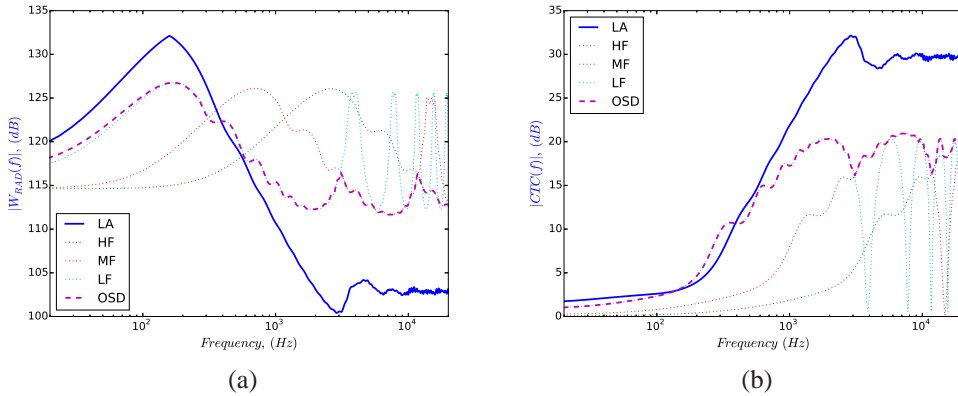


Figure 3: Radiated acoustical power (a) and cross talk cancellation performance (b) with the three way nested OSD array and with the 16 source LA. The reverberant performance is simulated inside a room of 90 m² with an average absorption coefficient $\alpha = 0.6$.

Under steady-state conditions the power input of a source into a diffuse field is balanced by the absorption of the room walls. The space-average squared reverberant pressure is related to the power radiated by the source, W , by [24]

$$(11) \quad \langle |p_R|^2 \rangle = \frac{4\rho_0 c_0}{R} W,$$

where $R = \frac{S\hat{\alpha}}{1-\hat{\alpha}}$, $\langle \rangle$ denotes spatial averaging, S represents the surface of the enclosure walls and $\hat{\alpha}$ is the average absorption coefficient of the walls. This equation allows us for the calculation of the reverberant pressure that any source produces inside a reverberant environment, once the source

radiated acoustic power and the absorptive characteristics of the room are known. The radiated power in a diffuse field is assumed to be the same as that radiated into a free space, as originally shown for a monopole[25]. The reverberant pressure component can then be combined with the direct pressure component radiated by the cross talk cancellation system, leading to the matrix \mathbf{R}_{REV} , defined as

$$(12) \quad \mathbf{R}_{REV} = \mathbf{H}^H \mathbf{C}^H \mathbf{C} \mathbf{H} + \langle |p_R|^2 \rangle \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \langle |R_{LLREV}(j\omega)|^2 \rangle & \langle |R_{LRREV}(j\omega)|^2 \rangle \\ \langle |R_{RLREV}(j\omega)|^2 \rangle & \langle |R_{RRREV}(j\omega)|^2 \rangle \end{bmatrix}.$$

In the case of a symmetrical listening configuration, the space-average cross talk cancellation spectrum is given by

$$(13) \quad \langle \psi(j\omega) \rangle = \frac{\langle |R_{LLREV}(j\omega)|^2 \rangle}{\langle |R_{RLREV}(j\omega)|^2 \rangle} = \frac{\langle |R_{LRREV}(j\omega)|^2 \rangle}{\langle |R_{RRREV}(j\omega)|^2 \rangle}.$$

This formulation has been used to simulate the performance of both the OSD array and the LA in a room with a surface of 90^2 with a frequency independent absorption coefficient along the frequency range $\alpha=0.6$. At low frequencies both cross talk cancellation systems have a similar performance inside the reverberant room. In this frequency range the aperture of both arrays is small compared with the radiated wavelength, and hence they are not efficient to cancel the pressure at the contralateral ear given the power constraint that was imposed to the systems. As the frequency increases both arrays become more directional. Above 500 Hz the cross talk cancellation of the OSD array is between 15 and 20 dB until 20 kHz, thanks to the action of the three separate channels. Above 500 Hz the array becomes more directional thanks to the contribution of the larger number of individual loudspeakers, obtaining about 10 dB more of performance than the OSD array.

4. Conclusion

This paper has presented a performance comparison between a 16 channel loudspeaker array and a three way optimal source distribution array for transaural reproduction. The performance was simulated using free field point Green functions, which allow to model loudspeakers at low frequencies and give a first insight of the device performance. The analysis has been carried out so that both arrays produce the same pressure in the ipsilateral ear, with the control filters created so that they do not exceed a certain level of array effort before these are equalised. The free field operation has shown that the 16 source loudspeaker array is able to obtain a larger performance than the optimal source distribution three-way 2 source loudspeaker array when using a similar level of electrical power.

The performance of the device inside a normal room has been also simulated, based on the power input of the array to the reverberant field, which is given by the total acoustical power the source radiates, the surface of the room walls and the absorption coefficient. This has shown that the 16 source loudspeaker array is able of radiating a much lower acoustical power above 500 Hz. This allows for a much larger cross talk cancellation performance in the reverberant field, which suggests that the use of an array is beneficial for transaural reproduction inside reverberant spaces. This study has shown that the 16-channel LA achieves better cross-talk cancellation than an optimal source distribution array, especially in a reverberant environment. This advantage, however, comes at the price of using a larger number of loudspeakers, which in turn require a larger costs and computational power. It is also likely than an OSD soundbar may be capable of generating higher quality sound since different loudspeakers are used for different frequency bands. Nevertheless, equalisation and the use of sub-woofers may allow for good quality transaural reproduction through a loudspeaker array of contained size.

5. 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.

References

1. J. Blauert, *Spatial hearing : the psychophysics of human sound localization*. Cambridge, Mass. MIT Press, 1997.
2. A. S and S. R, "Apparent sound source translator," Feb. 22 1966, uS Patent 3,236,949.
3. D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.*, vol. 37, no. 1/2, pp. 3–19, 1989.
4. O. Kirkeby, P. A. Nelson, and H. Hamada, "Local sound field reproduction using two closely spaced loudspeakers," *The Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 1973–1981, 1998.
5. O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *Journal of Audio Engineering Society*, vol. 47, no. 7/8, pp. 583–595, 1999.
6. E. Choueiri, "Spectrally uncolored optimal crosstalk cancellation for audio through loudspeakers," Patent, Mar. 22, 2012, wO Patent App. PCT/US2011/050,181. [Online]. Available: <http://www.google.com/patents/WO2012036912A1?cl=en>
7. 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.
8. 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.
9. 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.
10. C. Kyriakakis, "Fundamental and technological limitations of immersive audio systems," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 941–951, May 1998.
11. M. R. Bai and C.-C. Lee, "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 1976–1989, 2006.
12. Y. Huang, J. Benesty, and J. Chen, "On crosstalk cancellation and equalization with multiple loudspeakers for 3-d sound reproduction," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 649–652, Oct 2007.
13. G. T. Daniel Menzel, Helmut Wittek and H. Fastl, "The Binaural Sky: A Virtual Headphone for Binaural Room Synthesis." [Online]. Available: www.mmk.ei.tum.de/publ/pdf/05/05men1.pdf
14. J. Zheng, J. Lu, and X. Qiu, "Linear optimal source distribution mapping for binaural sound reproduction," in *Proceedings of Internoise 2014*, 2014.
15. K. Matsui and A. Ando, "Binaural reproduction of 22.2 multichannel sound with loudspeaker array frame," in *Audio Engineering Society Convention 135*, Oct 2013.
16. H. Kurabayashi, M. Otani, K. Itoh, M. Hashimoto, and M. Kayama, "Development of dynamic transaural reproduction system using non-contact head tracking," in *Consumer Electronics (GCCE), 2013 IEEE 2nd Global Conference on*, Oct 2013, pp. 12–16.
17. Sherwood, "S7-Optimal Sound Distribution Transaural Sounbar." [Online]. Available: <http://www.sherwood-av.com.au/product/s-7-150w-3d-soundbar-with-hdmi/>
18. M. 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.
19. S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio Speech and Language Processing*, vol. 20, no. 7, pp. 2123–2133, 2012.
20. 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.
21. M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "Personal audio loudspeaker array as a complementary tv sound system for the hard of hearing," *IEICE Trans. Fundamentals.*, vol. E97(9), 2014.
22. ———, "The effect of reverberation on personal audio devices," *The Journal of the Acoustical Society of America*, vol. 135, no. 5, pp. 2654–2663, 2014.
23. L. L. Beranek, *Acoustics*. New York: American Institute of Physics, 1987.
24. P. A. Nelson and S. J. Elliott, *Active Control of Sound*. London: Academic Press, 1992.
25. R. H. Lyon, "Statistical analysis of power injection and response in structures and rooms," *The Journal of the Acoustical Society of America*, vol. 45, no. 3, pp. 545–565, 1969.