Low-complexity, listener's position-adaptive binaural reproduction over a loudspeaker array

Marcos F. Simón Gálvez¹⁾, Takashi Takeuchi¹⁾ and Filippo Maria Fazi¹⁾

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

Summary

This work presents a method for binaural reproduction over a loudspeaker array that adapts to the listener's position by using a cross-talk cancellation approach that is updated in real-time. This is obtained by combining the audio signal processing system together with a computer vision mechanism that estimates listener's position. A novel approach to adapt the cross-talk cancellation filters is introduced here, which employs filters created for a central listening position using a free-field propagation model and far field beamforming techniques, hence requiring little memory and processing. The paper introduces simulations to show the effectiveness and robustness of the formulation together with free-field measurements of performance using a 16 loudspeaker compact array.

1 Introduction

Binaural audio reproduction allows for the creation of stable virtual audio images by reproducing at the ears of the listener ideally the same signals as those created by the real sound source. Binaural reproduction of audio through loudspeakers, hereafter referred to also as transaural reproduction, allows for the delivery of these signals without the need of wearing headphones. This method can be employed to design high-realism spatial audio home systems and is based on the principle of cross-talk cancellation, which reduces the acoustical cross-talk that occurs when the signal intended for the *ipsi-lateral* (same side) reaches also the *contra-lateral* (opposite side) ear. This method was first proposed by Atal and Schroeder [1]. After this invention, binaural reproduction over loudspeakers has been largely studied by many researchers as for example by Cooper and Bauck [2] and later by Kirkeby *et al.* [3, 4], whose work relied on the use of a stereo pair of loudspeakers. After that, transaural reproduction has evolved towards more robust reproduction by including purposefully-designed regularisation strategies [5]. The robustness of these systems with respect to the angular span of the two loudspeakers was also studied by Ward and Elko [6], who realised that in order to provide a good cross-talk at low frequencies,

a larger loudspeaker span is required than at high frequencies. This issue was later addressed by using special loudspeaker arrangements that allow for an uncoloured audio reproduction of the whole audio spectrum, as for examples those presented by Bauck [7] or by Takeuchi and Nelson [8].

Despite efforts in making cross-talk cancellation systems more robust, one of their key limitations is the fact that listener movements exceeding 75-100 mm away from the so-called sweet-spot may completely destroy the desired spatial audio effect [9, 10]. In order to obtain a cross-talk cancellation system which allows for the listener to move freely, some researchers have used systems including a larger number of loudspeakers, commonly known as loudspeaker arrays [11]. It has been shown that the use of a loudspeaker array of more than two sources can be more robust than a system with only two loudspeakers, as it allows for a greater cross-talk cancellation, is less sensitive to errors, and provides better performance in reverberant environments [12]. Example of loudspeaker array systems that have been used for transaural reproduction include a circular array surrounding the head of the listener [13] and linear arrays with a source spacing that maximises the cross-talk performance over a wide frequency range [14]. Linear arrays of large number of loudspeakers surrounding a TV have been also used to reproduce 22.2 multichannel sound [15]. Even though these systems allow for a larger sweet-spot area they still do not enable the listener to move freely over a wide area.

With the aim to provide an improved sweet-spot-independent reproduction, other work has incorporated listener tracking into audio reproduction devices. This has been employed for example in stereo reproduction systems [16, 17, 18, 19] making localisation of panned virtual sources robust to listener movements. Similarly, listener tracking has also been employed in cross-talk cancellation audio reproduction to increase the system robustness for listening positions outside the sweet-spot. This method was applied to transaural systems with two loudspeakers by Gardner [20], López and González

[21], Behler and Lentz [22], Mannerheim and Nelson [23], Song et al. [24], and Łopatka and Czyżewski [25]. The use of listener-adaptive systems with a larger number of loudspeakers for a single listener was also considered by Kurabayashi et al. [26], who used a 4 loudspeaker array to achieve single-listener reproduction. Thanks to the advancement of technology, modern computer vision systems such as those provided by the Microsoft Kinect SDK can be easily incorporated in audio reproduction systems. The complexity of the problem resides in the adaptation of the filters used to generate the signals feeding the loudspeakers, hereafter referred to as control filters, to the listener's position, whilst assuring at the same time a good audio quality, free of audible sound artefacts that may be caused by the filter adaptation process. Whilst the works presented in the bibliographic references above rely on a set of pre-computed filters for different listener's positions that are loaded into the system when the listener moves, other researchers considered time-domain adaptive filters [27]. Control architectures implemented in the time domain have also been studied [28] providing a large improvement in terms of processing complexity for a 2 loudspeaker cross-talk-canceller requiring, however, further equalisation.

This paper presents a novel time-domain approach for the adaptation of the cross-talk cancellation filters to the listener's position. This technique is especially suited for application to linear arrays employing a large number of loudspeakers. Differently from previous approaches, this formulation requires a single set of fixed cross-talk cancellation filters and a digital delay network, whose parameters are adapted in real-time to the listener's position. Note that the term "listener-position adaptive filters" states for array control filters that adapt to the listener's position and it must not be confused with adaptive filtering design theories as for example those used in active control of sound [29]. The structure of the paper is organised as follows: the standard cross-talk cancellation formulation is reviewed in Section 2. Section 3 introduces the formulation used to adapt the response to the listener position, together with a performance analysis in terms of robustness to listener's head misalignment. The real-time implementation of the technique is described in Section 4 for a 16-channel linear loudspeaker array. Experimental measurements are presented that show the performance of the system at different listening positions.

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

2 Cross-Talk Cancellation

A diagrammatic representation of binaural reproduction over a loudspeaker array using cross-talk cancellation is introduced in Fig. 1. The system comprises M control loudspeakers driven by signals $\mathbf{v}(j\omega) = [v_1(j\omega), \dots, v_M(j\omega)]^T$, where $\omega = 2\pi f$ is the angular frequency of the signal to be reproduced and the subscript 1 to M represents the loudspeaker index. This produces a sound-field $\mathbf{p}(j\omega) = [p_L(j\omega), p_R(j\omega)]^T$ at left and right ears of the listener, which can be written as

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

where $\mathbf{C}$ is the matrix of plant transfer functions. Taking the geometry of Fig. 1 as reference for an array of M loudspeakers, assuming that each loudspeaker behaves as an ideal point source with omnidirectional radiation pattern and neglecting the acoustic diffraction caused by the listener's head (free-field model), the matrix of transfer functions between each loudspeaker and the two control points which represent the ears of a listener is written as

$$\mathbf{C}(j\omega) = \frac{\rho_0}{4\pi} \begin{bmatrix} \frac{e^{-jk r_{L1}}}{r_{L1}} & \dots & \frac{e^{-jk r_{LM}}}{r_{LM}} \\ \frac{e^{-jk r_{R1}}}{r_{R1}} & \dots & \frac{e^{-jk r_{RM}}}{r_{RM}} \end{bmatrix}, \quad (2)$$

where ρ_0 is the air density, $k = \omega/c_0$ is the wavenumber and c_0 is the speed of sound in the air. A positive time convenience $e^{j\omega t}$ is used throughout this work. As observed in the diagram in Fig. 1, r_{L1} and r_{LM} are the distances between the left ear pressure control point and loudspeaker 1 and M , respectively, and r_{R1} and r_{RM} are the distances between the right ear pressure control point and loudspeaker 1 and M , respectively.

The binaural signals that are to be synthesised at the ears of the listener are defined by the elements of a complex vector $\mathbf{d}(j\omega) = [d_L(j\omega), d_R(j\omega)]^T$. In order to reproduce those signals at each ear, the $M \times 2$ filter matrix $\mathbf{H}$ containing the cross-talk cancellation

filters is introduced so that $\mathbf{v}(j\omega) = \mathbf{H}(j\omega)\mathbf{d}(j\omega)$. The pressures at the ears are thus given by

$$\mathbf{p}(j\omega) = \mathbf{C}(j\omega)\mathbf{H}(j\omega)\mathbf{d}(j\omega). \quad (3)$$

For the case of a loudspeaker array, in which the number of loudspeakers is larger than the number of pressure control points, the matrix $\mathbf{H}$ can be obtained by

$$\mathbf{H}(j\omega) = [\mathbf{C}(j\omega)^H \mathbf{C}(j\omega) + \beta \mathbf{I}]^{-1} \mathbf{C}(j\omega)^H, \quad (4)$$

where β is a regularisation parameter that limits the amount of energy of the filters, that is proportional to the norm of the columns of matrix $\mathbf{H}$ [30]. The superscript H stands for the Hermitian conjugate transpose operator. The above filters can be calculated in the frequency domain, as shown here, or in the time domain as described for example in [31]. Eq. 4 leads to a set of optimal filters that minimise the least square error between the binaural signals to be reproduced, $\mathbf{d}$ and the actual reproduced pressure $\mathbf{p}$ [32].

In order to analyse the cross-talk cancellation performance of the loudspeaker array a metric known as the cross-talk cancellation spectrum is introduced. This is obtained by including the crosstalk matrix, $\mathbf{R}(j\omega)$, which is defined as

$$\mathbf{R}(j\omega) = \mathbf{C}(j\omega)\mathbf{H}(j\omega) = \begin{bmatrix} R_{LL}(j\omega) & R_{LR}(j\omega) \\ R_{RL}(j\omega) & R_{RR}(j\omega) \end{bmatrix}, \quad (5)$$

The ratios between the elements of this matrix define the cross-talk cancellation spectrum for each ear, which are defined as

$$CTC_L(j\omega) = \frac{|R_{LL}(j\omega)|^2}{|R_{RL}(j\omega)|^2}, \quad CTC_R(j\omega) = \frac{|R_{RR}(j\omega)|^2}{|R_{LR}(j\omega)|^2}. \quad (6)$$

The cross-talk cancellation spectrum is a dimensionless quantity, whose level is typically plotted in decibels.

3 Adaptive System

The formulation described above allows to minimise the acoustic cross-talk for a given listener position which the filters were created for, i.e., the sweet-spot. If, however, the listener moves away from this position the cross-talk increases significantly and hence the ability to produce virtual acoustic images is largely reduced. In order to overcome this issue the filters have to be updated.

A method is proposed hereafter to perform such adaptation. The array driven by the filters $\mathbf{H}(j\omega)$ generates two sound beams that deliver the binaural signals to the listener's ears at a given, on-axis reference position. When the listener moves away from

this position, these beams are electronically *steered* towards the new listener's position by applying a set of digital delays to the loudspeaker signals. These digital delays are represented by the steering matrix $\mathbf{D}(j\omega)$, which is a $M \times M$ diagonal matrix. This process yields a new matrix of *steered* control filters, $\mathbf{H}_S(j\omega)$, which is given by

$$\mathbf{H}_S(j\omega) = \mathbf{D}(j\omega)\mathbf{H}(j\omega). \quad (7)$$

The elements of matrix $\mathbf{D}(j\omega)$ corresponds to the complex weights of a far-field delay and sum beamformer [33]. Note that the beamformer is only used to steer the cross-talk cancelling beams, and it is not employed to perform any radiation pattern control. The matrix $\mathbf{D}(j\omega)$ is built as

$$\mathbf{D}(j\omega) = \begin{bmatrix} e^{-jkx_1 \cos(\theta)} & 0 & \dots & 0 \\ 0 & e^{-jkx_2 \cos(\theta)} & \dots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & e^{-jkx_M \cos(\theta)} \end{bmatrix}, \quad (8)$$

where x_l corresponds to the position of each loudspeaker assuming the array lies along the x axis and θ is the angular location of the listener with respect to the array axis, as shown in Fig. 2.

Figure 2: Geometry used to adapt the filters of the loudspeaker array.

As it is shown below, this formulation enables the listener to move in space along θ , following a circular trajectory, but the filters do not compensate for head rotation (the listener's head is always supposed to be oriented towards the centre of the array) and for variation of the distance between listener and array, even though small listener's displacements along the radial direction do not degrade the cross-talk cancellation performance.

3.1 Simulations of System Performance

A set of simulations was carried out to study the performance of the method introduced above at different listening positions, considering an array of

Figure 3: Geometry of the loudspeaker array (square boxes) used to control the pressure at 2 control points (stars) representing the ears of the listener for various listening positions corresponding to 0° , 15° and 30° , respectively.

16 point monopole sources with $d = 0.039$ m spacing. Two sets of control filters are compared, those created according to Eq. 4, here called “Inverse filters”, and those created according to Eq. 7, called “Steered”. In this case the “Inverse filters” were created for each new listener position, whilst the “Steered” filters were obtained by combining a set of control filters created for a reference on-axis listener position ($x = 0$ m, $y = 1.5$ m, $\theta = 0^\circ$) termed “Centered”, in conjunction with the steering matrix defined in Eq. 8. Frequency independent regularisation was applied using $\beta = 0.03$. This value was selected empirically and shown to provide a good balance between filter norm and cross-talk cancellation. For more information of how to select the regularisation parameter please refer to these references [30, 36].

The simulations were performed using the free-field transfer impedances of point monopole sources, as these represent a good model of the loudspeaker radiation at low frequencies and have been previously used to estimate the performance of loudspeaker arrays [30, 34]. Furthermore, the main study of the paper concerns the control filter updating and hence room effects are to be avoided. If the reader is interested in how rooms affect the response of a line array please refer to [35, 12]. A control geometry as shown in Fig. 3 was employed to simulate the performance for the different listener positions. The cross-talk cancellation spectrum and the frequency response at the listener’s ears are employed as objective metrics of performance. In this case, a large cross-talk cancellation and a flat reproduced response are considered ideal, as it would allow to reproduce any desired binaural signal.

A first comparison of performance is shown in

Fig. 4, calculated for three positions 1.5 m away from the loudspeaker array for observation angles at 0° , 15° , and 30° . The left-hand side graphs in the figure show the cross-talk cancellation performance whilst the right-hand side graphs show the frequency response at the listener’s ears. For comparison, the performance obtained with an array with filters optimised for a central listening position, calculated for $\theta = 0$, (“Centred” filters) is also presented. The results show that the three set of filters, i.e., “Inverse filters”, “Steered” and “Centred”, give the same performance when the listener is at the reference position, as in this case the delay network is not acting and the control filters are identical. In this case, a large cross-talk cancellation, up to 80 dB at some frequencies, was obtained in the numerical simulations, with some peaks at those frequencies at which the array is best conditioned [8]. The figures corresponding to 15° and 30° show that the cross-talk cancellation performance provided by the “Inverse Filters” is still large, whilst that the other two approaches obtain a much lower performance. The cross-talk cancellation performance obtained for the “Centred” approach is practically 0 dB, showing some peaks at certain frequencies, whilst the “Steered” method shows peaks of about 30 dB of cancellation at 15° and 30° . Although this represents a much lower figure than that obtained with the “Inverse filters” it is shown below that the filters created with both approaches, “Inverse filters” and “Steered”, behave very similarly in a practical realisation of the system. The comparison between left and right cross-talk of the “Steered” configuration (L and R in the figure) shows a difference in performance when the array is steered, with maximum deviations of about 15 dB at certain frequencies.

The frequency responses are reduced below 3 kHz due to the use of regularisation, which can be compensated for by applying an equalisation filter to flatten the response. As the listener is moved away from the reference position, the difference between the frequency responses of the “Inverse filters” and the “Steered” approach increases, showing a discrepancy proportional to the steering angle θ . This variation is, however, small, with a maximum difference of 5 dB at 20 kHz when the listener is at the 30° position, as shown in Fig. 4.

The radiation patterns obtained with the loudspeaker array at different frequencies are shown in Fig. 5. These show that the radiation patterns of the “Inverse filters” and the “Steered” control approach are very similar. The “Inverse filters” can achieve greater cross-talk cancellation at some frequencies, which can be observed in the directivity plots for 1 kHz, 2 kHz, and 10 kHz. At other frequencies, however, the radiation pattern of the “Steered”

Figure 4: Cross-talk cancellation (CTC) (left hand side plots) and frequency (right hand side plots) responses at various listener positions, i.e., 0° (top plots), 15° (middle plots) and 30° (lower plots). “Steered (L)” and “Steered (R)” correspond to the CTC magnitudes for each beam.

approach is almost identical to that of the “Inverse filters”.

3.2 System Robustness

So far it has been shown that the “Inverse filters” have better cross-talk cancellation performance than the “Steered” approach. A robustness simulation is presented here in order to further analyse this effect, in which the cross-talk cancellation performance is assessed for listener positions that do not correspond exactly to those for which the filters were calculated. In this case, cross-talk cancellation filters have been calculated for a listener position at 15° , with the average cross-talk cancellation

performance calculated when the position of the listener presents a mismatch of ± 5 cm. This method has been previously used to assess the robustness of a super-directional loudspeaker array [36, 35] and has shown to provide a good estimate of the performance in case of loudspeaker and control point position mismatches. An average over 100 random position was calculated, as shown in Fig. 6.

In order to perform this analysis each transfer function Z between a source of the loudspeaker array and one of the listener’s ears was formed by using the Green function of a point monopole source with a distance component modified to account for the mis-

Figure 5: Radiation patterns obtained at different frequencies with the “Inverse filters” (dashed line) and with the “Steered” method (solid line) for the right channel beam. The two octagons represent the listener’s ears.

Figure 6: Listener position variation (thin stars) compared to the nominal listening position for which the control filters were created for (thick stars).

match in the listener’s position. This is written as

$$Z(j\omega) = \frac{e^{jk r_M}}{r_M}, \quad (9)$$

with $r_M = r(1 + a\gamma)$, r being the nominal distance value from a given source to the given control point simulating an ear of the listener, and γ is a random number with Gaussian distribution. The coefficient a was selected so that the standard deviation of the

Figure 7: Study of the robustness for head misalignment with respect to the two listener-adaptive control methods. Results calculated for a listener 15° away from a central listening position.

mismatch is 10 cm. The transfer impedances were used to form a matrix of *noisy* transfer impedances, $\mathbf{C}_N$, which is employed to form a new matrix of pressures at the listener ears, $\mathbf{R}_N$, allowing to assess the cross-talk cancellation performance.

The cross-talk cancellation results when the listener

is slightly away from the nominal position at which the filters were designed for is shown in Fig. 7. The results show the performance of the “Inverse filters” and of the “Steered” approach with no mismatches in the listener position and the average performance when random position errors are present, shown in the legend as “Noisy” results. As previously shown in Fig. 4, the performance of the “Inverse filters” is better than that of the “Steered” filters for the reference position. The performance of the two methods are, however, very similar when a small error in the position of the listener is present, as it is expected to happen in a practical realisation of the system, thus encouraging the use of the “Steered” formulation for real-time applications.

4 Real-Time System Implementation

4.1 Signal Processing

Figure 8: Block diagram of the signal processing scheme for adapting the array output to the listener position.

The array steering method described above can be used to implement a real-time audio reproduction system which adapts to the listener position, hence allowing to reproduce binaural material accurately when the listener is outside of the sweet-spot. A block diagram of the Digital Signal Processing required for the implementation of the system is shown in Fig. 8. The left and right binaural signals are first convolved with the filters obtained from the matrix $\mathbf{H}(j\omega)$ for a central listening position, as described in Eq. 4. Each signal is convolved with M filters for each channel, thus giving a total of $2M$ filters. In this case, filters

with a length of 512 coefficients are used. The left and right filtered signal for each loudspeaker are then combined and fed to a bank of variable digital delays (one for each loudspeaker), which is used for steering the binaural sound beams generated by the array towards the listener when the latter moves.

The M steering delays are updated in real-time depending on listener, which is provided by a Microsoft Kinect sensor used to track the listener’s movement. After the steering delays, the signals are fed to the M loudspeakers. The left and right input channels can also be filtered by equalisation filters to compensate for a given set of HRTFs or for the frequency response of the loudspeakers. This system has been implemented in MAX MSP 6 in a Windows computer using the loudspeaker array shown in Fig. 9.

4.2 Performance Measurements

(a)

(b)

Figure 9: Detail of the 16 source loudspeaker array (a) and measurement set up in the anechoic chamber (b).

The performance of the small 16-channel loudspeaker array in Fig. 9 with listener-adaptive filters created with the method described in the previous

Figure 10: Impulse response for the left input channel measured at the sensors of the binaural microphone located at 0° (a) and cross-talk cancellation (CTC) performance of the loudspeaker array when a listener is at 0° (b), 15° (c) and 30° (d).

sections of this paper was measured in the large anechoic chamber of the Institute of Sound and Vibration Research. The loudspeaker array is the same as that used, for example, in [37] and consists of 16 electrodynamic loudspeakers, with a 0.039 m spacing between the centres of neighbouring drivers. The cross-talk cancellation performance was measured with a Neumann KU-100 binaural microphone. The experimental set up is shown in the lower plot of Fig. 9. The measurements were performed at the same positions as those used for the simulations described in Section 3.1. The radiation patterns of the array were also measured, at a distance of 1.5 m from the centre of the array.

The measured left channel impulse response at the reference listening position, at 0° with respect to the array, is shown in Fig. 10a, where a large response difference between left and right ears can be observed. The cross-talk cancellation response is shown in Figures 10b, 10c and 10d. The graphs, corresponding to

the three different orientations, show a very similar frequency curve. The variation between the responses of the left and right channel (light and dark curves in each plot) is also small. The measured performance is much lower than the free field simulated responses for the unperturbed case, especially at the high frequencies, but is not very different from the simulated response with listener position errors shown in Fig. 7.

The measured radiation patterns generated by the array at different frequencies for three steering angles corresponding to 0°, 15° and 30°, are shown in Fig. 11. These present a similar pattern to those in the simulations discussed in Section 3.1. At the low frequencies, the cross-talk cancellation is achieved by placing a notch in the direction of the contralateral ear of the listener. At the high frequencies, where the array is very directional, the array beams towards the ipsi-lateral ear. Clearly, the “Steered” formulation is able to adapt the radiation pattern of the array as the listener moves, maintaining a good

Figure 11: Measured free-field radiation patterns for the right ear beam at different frequencies. The array steering angles are 0° (left hand side plots), 15° (centre plots) and 30° (right hand side plots).

cross-talk cancellation performance independently of the listener position.

5 Conclusion

A method to achieve listener-adaptive binaural reproduction with a loudspeaker array was presented. The method combines filters created for a central reference listening configuration with a beamforming approach used to steer the binaural sound beams generated by the loudspeaker array towards the listener, thus achieving a large cross-talk cancellation for a listener position in front of the array. Although this only allows for listener's position compensation over a circular trajectory a certain distance apart from the loudspeaker array, small-scale listener's displacements along the radial direction do not degrade the cross-talk cancellation performance.

The performance of the new formulation was compared by means of simulations to that obtained in previous implementations, in which “inverse” cross-talk cancellation filters are directly pre-calculated for a grid of listening positions and are updated as the listener moves. It also has been shown by computing

numerical simulations that the performance of the new method is practically identical to that obtained with the “inverse” filters when small errors in the listener position are considered. This introduces a significant improvement with respect to systems based on a grid of pre-measured responses, as the memory and computational requirements of the proposed method are much lower.

The above formulation was implemented in a fully-operational real-time system, using a loudspeaker array of 16 sources and a Microsoft Kinect sensor to allow a sweet-spot independent reproduction of virtual acoustic images. The performance of the system was measured for different listening positions, showing a large cross-talk cancellation figure at each of the measurement points.

6 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.

Thanks to Dylan Menzies, Andreas Franck, and Ferdinando Olivieri for the long conversations about the system operation. The authors are also very obliged to Peter Otto and the University of California, San Diego, who have kindly provided the array used for the experiments.

References

- [1] S. Atal and R. Schroeder, "Apparent sound source translator," Feb. 22 1966, US Patent 3,236,949.
- [2] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *Journal of the Audio Engineering Society*, vol. 37, no. 1/2, pp. 3–19, 1989.
- [3] 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.
- [4] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *Journal of the Audio Engineering Society*, vol. 47, no. 7/8, pp. 583–595, 1999.
- [5] E. Choueiri, "Spectrally uncolored optimal crosstalk cancellation for audio through loudspeakers," Patent, Mar. 22, 2012, WO Patent App. PCT/US2011/050,181.
- [6] 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.
- [7] J. Bauck, "A simple loudspeaker array and associated crosstalk canceler for improved 3d audio," *Journal of the Audio Engineering Society*, 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] C. Kyriakakis, "Fundamental and technological limitations of immersive audio systems," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 941–951, May 1998.
- [10] 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.
- [11] 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.
- [12] Marcos F. Simón Gálvez and F. M. Fazi, "Loudspeaker arrays for transaural reproduction," in *Proceedings of the 22nd International Congress on Sound and Vibration, Florence, Italy*, 2015.
- [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] S. Merchel, "Sweet Spotter-Listener-adaptive stereo compensation system." [Online]. Available: <http://www.sebastianmerchel.de/sweetspotter.html>
- [17] Marcos F. Simón Gálvez, D. Menzies, F. M. Fazi, T. de Campos, and A. Hilton, "A listener position adaptive stereo system for object-based reproduction," in *Audio Engineering Society 138th Convention, Warsaw, Poland*, May 2015.
- [18] T. Momose, M. Otani, M. Hashimoto, M. Kayama, and K. Itoh, "Adaptive amplitude and delay control for stereophonic reproduction that is robust against listener position variations," *J. Audio Eng. Soc.*, vol. 63, no. 1/2, pp. 90–98, 2015.
- [19] Marcos F. Simón Gálvez, D. Menzies, F. M. Fazi, T. de Campos, and A. Hilton, "Listener tracking stereo for object based audio reproduction," in *Tecniacustica 2016 (Valencia) - European Symposium in Virtual Acoustics and Ambisonics*, 2016.
- [20] W. G. Gardner, "3-D Audio Using Loudspeakers," Ph.D. dissertation, Massachusetts Institute of Technology, 1997.
- [21] J. J. López and Alberto González, "3-D audio with dynamic tracking for multimedia environments," in *Proceedings of the 2nd COST-G6 Workshop Digital Audio Effects*, 1999.

- [22] G. Behler and T. Lentz, “Dynamic cross-talk cancellation for binaural synthesis in virtual reality environments,” in *Audio Engineering Society Convention 117*, Oct 2004.
- [23] P. Mannerheim and P. A. Nelson, “Virtual sound imaging using visually adaptive loudspeakers,” *Acta Acustica united with Acustica*, vol. 94, pp. 1024–1039, 2008.
- [24] M.-S. Song, C. Zhang, D. Florencio, and H.-G. Kang, “An interactive 3-d audio system with loudspeakers,” *Multimedia, IEEE Transactions on*, vol. 13, no. 5, pp. 844–855, Oct 2011.
- [25] K. Lopatka and A. Czyżewski, “Adaptive acoustic crosstalk cancellation in mobile computer device,” *Elektronika : konstrukcje, technologie, zastosowania*, vol. 55 (4), pp. 32–36, 2014.
- [26] 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.
- [27] L. Wang, F. Yin, and Z. Chen, “A stereo crosstalk cancellation system based on the common-acoustical pole/zero model,” *EURASIP J. Adv. Signal Process*, vol. 2010, pp. 3:1–3:11, Feb. 2010.
- [28] S. Cecchi, A. Primavera, M. Virgulti, F. Bettarelli, J. Li, and F. Piazza, “An efficient implementation of acoustic crosstalk cancellation for 3d audio rendering,” in *Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit International Conference on*, July 2014, pp. 212–216.
- [29] P. A. Nelson and S. J. Elliott, *Active Control of Sound*. London: Academic Press, 1992.
- [30] 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.
- [31] —, “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.
- [32] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduña Bustamante, “Fast deconvolution of multichannel systems using regularization,” *IEEE Transactions on Audio Speech and Language Processing*, vol. 6, no. 2, 1998.
- [33] 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. John Wiley & Sons, 2000.
- [34] 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.
- [35] —, “The effect of reverberation on personal audio devices,” *The Journal of the Acoustical Society of America*, vol. 135, no. 5, pp. 2654–2663, 2014.
- [36] 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.
- [37] F. Olivieri, M. Shin, F. M. Fazi, P. A. Nelson, and P. Otto, “Loudspeaker array processing for multi-zone audio reproduction based on analytical and measured electroacoustical transfer functions,” in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, Sep 2013.