MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

A study on the effect of reflections and reverberation for low-channel-count Transaural systems

Simón Gálvez, Marcos Felipe[1]
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

Blanco Galindo, Miguel
Centre for Vision, Speech and Signal Processing, University of Surrey, GU2 7XH, UK

Fazi, Filippo Maria
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

ABSTRACT

Cross-talk cancellation allows for the reproduction of binaural audio through loudspeakers. This uses a digital signal processing network that controls the acoustic pressure at the listener's ears. Although this can be achieved by using only two drivers, loudspeaker arrays extend the operating frequency range and are more robust against mismatches between nominal and actual transducer transfer functions. This document presents a numerical study on the trade-off between cross-talk cancellation performance and the number of channels of a loudspeaker array in two practical room environments. A third order image source model is used to predict the performance of loudspeaker arrays using 2, 3, 4, 5 and 7 speakers. The simulated results show how low-order reflections and the reverberant pressure created by the loudspeaker arrays affect the cross-talk cancellation and the reproduced acoustic pressures. The results show that arrays using 5 sources or more can minimise colouration on the reproduced response and maximise the effectiveness of the cross-talk cancellation over the entire audio frequency range.

Keywords: Loudspeaker Arrays, Transaural, Cross-talk cancellation,
I-INCE Classification of Subject Number: 25, 75, 76.

[1]M.F.Simon-Galvez@soton.ac.uk

1. INTRODUCTION

Transaural audio [1] allows for the reproduction of binaural signals through loudspeakers. This is typically achieved by using a set of inverse filters that are created for a given listening position. The inverse filters facilitate the delivery of two signals at the listener's ears that are meant to match a desired binaural signal. This is obtained by cancelling the reproduced signal at the opposite ear, and it is therefore that such devices are also typically termed as cross-talk cancellation systems. This method represents an alternative to binaural reproduction over headphones and it can be beneficial for some situations in which the user also requires interaction with the real world.

The first examples of cross-talk cancellation were developed using stereo set-ups, with the control network implemented using analogue electronics [2]. The development of digital signal processing saw an outbreak in the amount of research on stereo-based cross-talk cancellation systems [3, 4], with emphasis put around finding optimum spans to maximise sound field control along the whole frequency range [5, 6]. Studies showed that a wider span was required at low frequencies than at high frequencies, which motivated the use of frequency dependent arrangements of stereo systems to obtain uncoloured audio reproduction along the whole frequency range [7]. Later technology development used three-way systems to further improve the robustness of the solution at critical frequency bands [8].

Cross-talk cancellation systems are often created assuming a free-field operation mode - their filters are created using analytical models or measurements of transfer functions in anechoic chambers. This assumption, however, does not hold in real-life applications, and the performance of inverse systems does degrade with the influence of room reflections and the listening environment [9]. This effect was first studied from the point of view of a stereo-dipole [10], with results showing that reflections were considerably modifying the interaural level difference of a system (ILD) but that these had little different in the interaural time difference (ITD). This result was encouraging and showed that reflections did not affect much the Transaural performance, however, it was performed with a single side reflection, something far from a practical listening scenario. Later research found that reflections with a delay of 5 ms and 10 ms could affect localisation [11]. Later studies have also been performed, but have not given enough evidence of how detrimental reflections are for localisation [12] or have not been able to show significant evidence of how detrimental reflections are for Transaural localisation [13]. These studies have shown that while reflections degrade the performance of Transaural systems, in practice it is hard to draw overall conclusions as many factors affect it: distance from the source, room characteristics (reflection, distance from the wall) and the *radiation performance* of the Transaural system itself. This latter aspect is what this paper focuses on.

Loudspeaker arrays have been used in the past over two-channel systems due to: 1) their higher robustness against inaccuracies between assumed transducer transfer functions used to design the filters and those given by real loudspeakers [14, 15] and 2) by their limited interaction with the room acoustics due to their additional directivity [15, 16]. This document builds onto previous work that has investigated the free-field performance of cross-talk cancellation systems based on loudspeaker arrays of different numbers of speakers (i.e.; 2, 3, 4, 5 and 7 drivers [17]), but in this case focusing on the

in-room response of the different driver array systems.

The paper structure is presented as follows: Section 2 of the paper introduces the signal processing required for the cross-talk cancellation sound-field control, and Section 3 focuses on simulation and the analysis on the in-room cross-talk cancellation performance. Section 4 summarises the paper outcomes and introduces some future work.

2. ARRAY SIGNAL PROCESSING

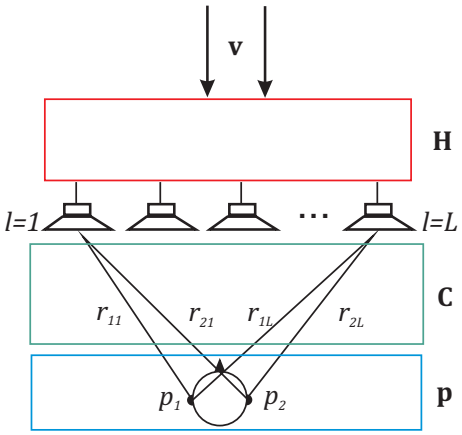


Figure 1: Geometry of a loudspeaker array used to control the pressure at two field points corresponding to the ears of a listener. The various blocks of the signal processing scheme used for the acoustic control are depicted.

Considering a Cartesian coordinate system, $\mathbf{x} = (x_1, x_2, x_3)$, Fig. 1 shows a loudspeaker array of L sources with coordinates $\mathbf{y}_l = (y_1, y_2, 0)$ whose radiated acoustic pressure field is controlled at two points in space, $\mathbf{x}_1$ and $\mathbf{x}_2$, which simulate the ears of a listener. The notation $p_m \equiv p(\mathbf{x}_m, j\omega)$ is used to denote the complex radiated pressure as a function of radian frequency, $\omega = 2\pi f$, at each control point.

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

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

It is assumed that each loudspeaker behaves as an ideal monopole source, such that $\mathbf{c}_m = \left[c_{m1}e^{-jkr_{m1}}, \cdots, c_{mL}e^{-jkr_{mL}}\right]$, where an $e^{j\omega t}$ time dependence is assumed, $k = \omega/c_0$ and c_0 is the speed of sound. The quantity $c_{ml} = 1/r_{ml}$ is a distance attenuation factor where $r_{ml} = \|\mathbf{x}_m - \mathbf{y}_l\|_2$.

The vector of reproduced signals at the listener's ears, $\mathbf{p}$, is defined as

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \tag{2}$$

and is found from the relation

$$\mathbf{p} = \mathbf{CHv}, \tag{3}$$

where $\mathbf{H}$ is a set of cross-talk cancellation filters given by

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

and $\mathbf{v}$ is the vector of desired binaural signals

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \tag{5}$$

The above acoustic control problem can be stated as a minimisation problem of the form,

$$\min\{||(\mathbf{CH} - \mathbf{I}_2)\,\mathbf{v}||_2 + \beta ||\mathbf{Hv}||_2\}, \tag{6}$$

where $\mathbf{I}_2$ is a 2×2 identity matrix. The objective of the above problem is to minimise the cross-talk due to the contralateral binaural signal at the ipsilateral ear. The parameter β is the Tikhonov regularisation factor that is used to limit the energy used by the array by reducing $||\mathbf{Hv}||_2$. The addition of regularisation in turn improves system conditioning and increases system robustness with respect to inaccuracies in the speaker response [16]. However, regularisation does also introduce a bias in the minimisation of $\mathbf{CH} - \mathbf{I}_2$ in Eq. 6, and it must be chosen carefully as otherwise it can lead to compression in the perceived virtual stage [18].

For the results presented in this paper, the channel separation is used as an initial performance metric. The channel separation of the reproduced signal is expressed in terms of the cross-talk cancellation (CTC) spectrum, which states the ratio between the absolute acoustic pressure at the two control points p_1 and p_2. This is written as

$$\mathrm{CTC}_{i,q} = 20 \log_{10} \left(\frac{|p_i|}{|p_q|} \right), \tag{7}$$

with the subscripts i and q alternating between 1 and 2 depending on the binaural channel for which the CTC spectrum is calculated. As in this paper we are assuming a symmetric configuration, it is assumed that $\mathrm{CTC}_{1,2} = \mathrm{CTC}_{2,1}$. The results presented below use the terminology of *ipsilateral* ear, where the acoustic pressure is to be maximised, and *contralateral* ear, where the acoustic pressure is to be minimised.

Another metric used in this study is the array effort. The array effort characterises how much energy the cross-talk cancellation filters require, as often they need large boosts of acoustical energy to control the sound-field at frequencies at which the system is ill-conditioned [7]. The array effort is defined as the norm of the control filters, divided by the norm of the input signal, h_S, that a single loudspeaker requires to obtain the same pressure as that produced by the cross-talk cancellation system in a given ear. The normalised array effort (AE) is thus defined as

$$\mathrm{AE} = 10 \log_{10} \left(\frac{\sum_{l=1}^{L} \left(|H_1|^2 + |H_2|^2 \right)}{|h_S|^2} \right). \tag{8}$$

This quantity is proportional to the amount of electric power employed by the array filters, 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, the negative effect of ill-conditioning of the propagation matrix is mitigated and thus the array is made more robust to changes in the actual reproduction system (gain, phase, positions and self-noise) [14], yet at the expense of reducing the cross-talk cancellation. The array effort is used below to compare different array geometries under the same energy constraints.

3. ROOM EFFECT ON ARRAY RESPONSE

To model the performance of the different loudspeaker arrays under study, a close listening distance has been assumed, resembling a scenario of a user interacting with a laptop mounting a Transaural system that renders 3D audio, as per the left hand side of Fig. 2. For all the simulations presented below, a listener in the axis of symmetry of the loudspeaker array at 0.6 m has been assumed. Different arrays have been considered with a maximum aperture of 35 cm and different inter-element configurations, as shown in the right hand side of Fig. 2.

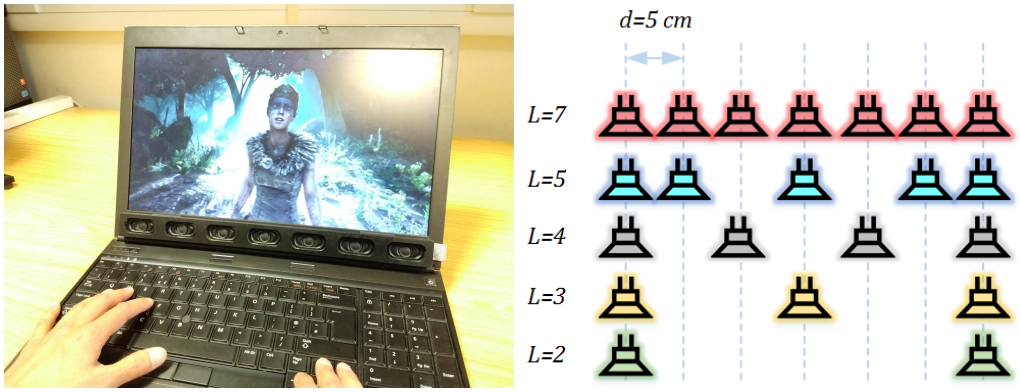


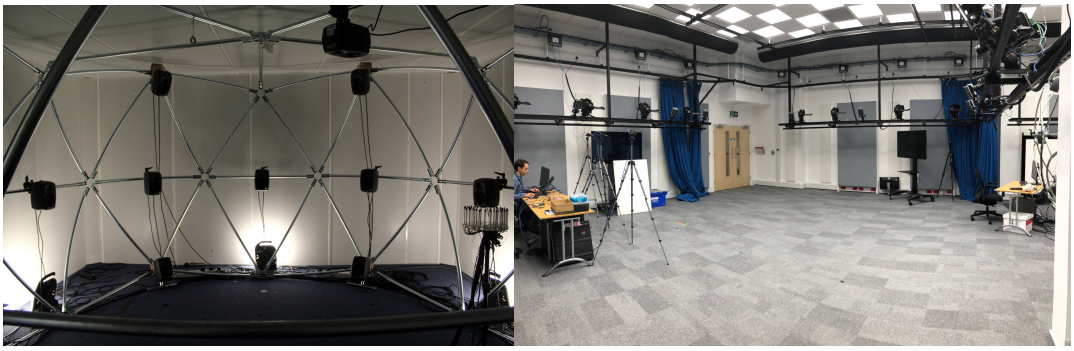
Figure 2: Close up of a loudspeaker array system for the reproduction of Transaural audio and the different array geometries that have been modelled.

3.1. Modelling the room

The arrays' reverberant performance was simulated in two real world practical environments: the Vision Lab (Vislab) and the Audio Booth at University of Surrey, which can be observed in Fig. 3. The Audio Booth has dimensions 4.12 x 4.98 x 2.1 and a mid-frequency reverberation time (T_{MF60}) below 0.2 s. Vislab is a larger audio-visual lab with dimensions 7.79 x 11.84 x 4.02 m and a reported T_{MF60} <0.4 s, although this was measured with a large curtain covering the back half of the room, reducing its effective volume. Octave band absorption coefficients were obtained by selecting appropriate materials from [19] to match that of the study environments. The simulations presented below assume a loudspeaker array placed in the centre of both rooms at 1.2 m of height.

The effect of the room is modelled by generating "reverberant" transfer functions using the image source model. The image source model method is a widely used method

to simulate room transfer functions, which operates by mirroring the sound source through each room boundary, creating equivalent "free field" sources. For the simulations presented in this paper, reflections up to third order are considered. Using this method, a set of reverberant transfer functions $\mathbf{C}_{REV}$ were used to simulate a set of binaural pressures under the influence of the room, $\mathbf{p}_{REV}$, with the CTC spectrum obtained with the new set of reverberant pressures according to Eq. 7. For easiness of observation, the reverberant results presented below have been filtered using second octave bandwidth Gaussian smoothing [20].



(a) Audio Booth (b) Vislab

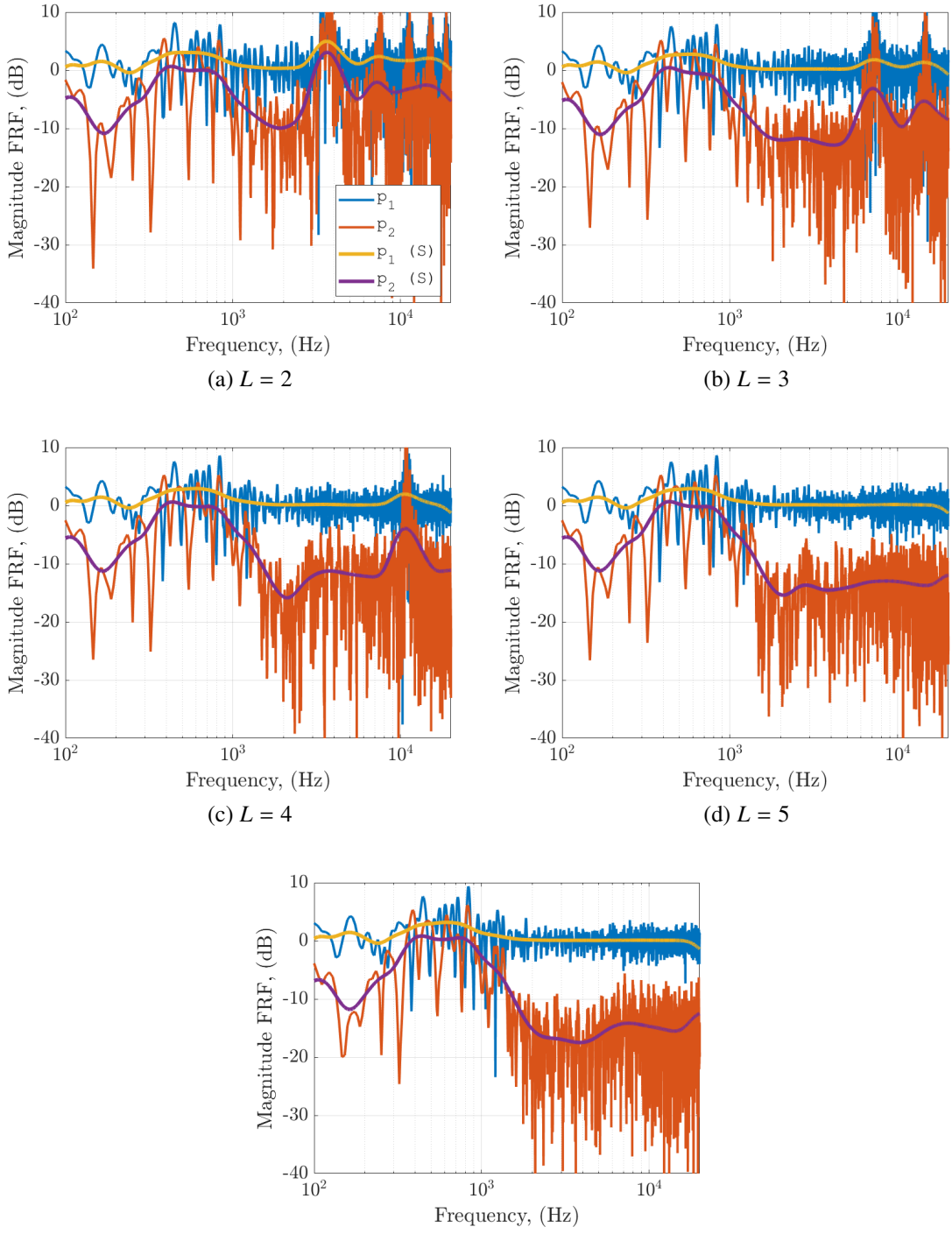
Figure 3: Rooms used to model the performance of the loudspeaker array: the Audio Booth and the Vision Laboratory (Vislab) at the Centre for Vision Speech and Signal Processing of University of Surrey.

3.2. Cross-Talk Cancellation performance

The different loudspeaker geometries using different number of speakers (2 to 7 drivers) are shown in Fig. 2b. Using the formulation and the free-field propagation model described in Section 2, array control filters were created considering a listener placed on the axis of symmetry of the array at a distance of 0.6 m. The array filters were created so that the array effort, defined as in Eq. 8, was kept below 10 dB for all the array geometries. A set of reverberant pressures were calculated with this set of filters according to Eq. 3, using the set of simulated $\mathbf{C}_{REV}$ as described in previous section.

The simulated reverberant pressures are shown in Fig. 4 for one of the channels of the binaural signal; in this case, a left binaural channel is to be reproduced, so that the acoustic pressure is maximised at p_1 but minimised in p_2. The plots show how the different arrays present colouration in the reproduced frequency responses at certain frequencies. Whilst all the arrays seem to present the same low frequency response, a pattern can be identified in which the frequency response at the ipsilateral ear, p_1, is made more uniform as the number of loudspeakers is increased. This is clearly visible for the loudspeaker arrays using 2, 3 and 4 speaker systems, which present colouration of up to 5 dB at certain frequencies, while the geometries using 5 and 7 loudspeakers are free of colouration above 1.5 kHz.

The simulated CTC spectrum for different array geometries is shown in Fig. 5 for the two rooms under study, together with the free-field performance included for comparison.



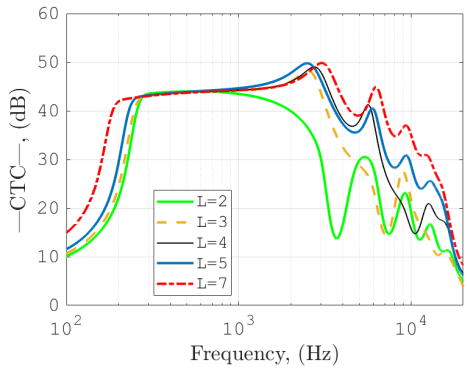
(a) $L = 2$ (b) $L = 3$

(c) $L = 4$ (d) $L = 5$

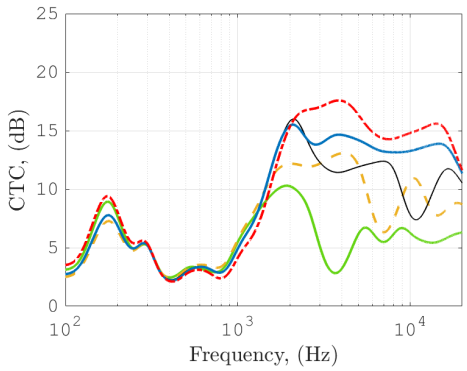
(e) $L = 7$

Figure 4: Raw (thin lines) and smoothed (thick lines) reproduced reverberant pressures in the Vislab environment for the different geometries of study.

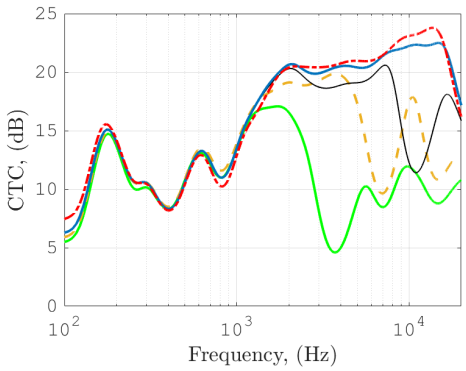
For each of the rooms included in the study, the simulated reverberant results present a similar CTC performance at low frequency between all the array geometries, i.e., between 100 Hz and 1.5 kHz. Above 1.5 kHz, a clear difference in performance can be seen by the systems using more than 2 loudspeakers. While the predicted maximum performance in free field is similar for all the different array geometries, the predicted



(a) Free-field



(b) Vislab



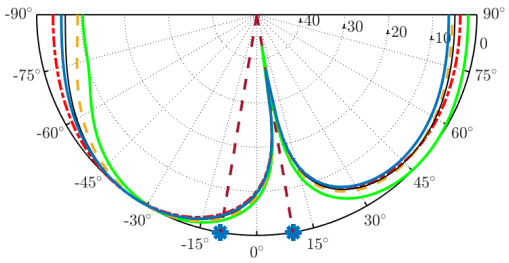
(c) Audio Booth

Figure 5: Array geometries (2 to 7 speakers) and the corresponding cross-talk cancellation spectrum modelled in the Vision-laboratory (Fig. 3b) and in the Audio Booth (Fig. 3a)

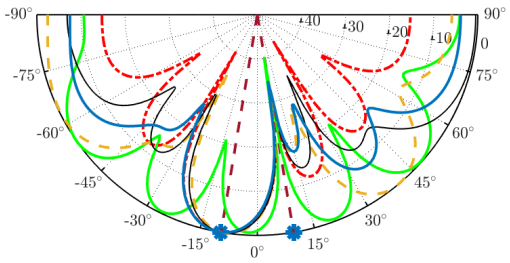
.

reverberant performance increases with the number of sources, same as the uniformity of the free-field response above 2 kHz. As it is shown below, the free-field response and the reverberant response are related. The authors suggest the reader to refer to [17] for a better understanding on how the different arrays perform in the free-field. The CTC systems using 2 loudspeakers peaks in performance between 1 and 3 kHz, and obtains a maximum of 10 dB (Vislab) and 15 dB (Audio Booth). Adding a third speaker increases the frequency range of *effective* CTC to about 6 kHz, with has also been demonstrated in other theoretical studies [21, 17] and practical systems [8]. For the system using three loudspeakers, the CTC spectrum presents maximums of 13 dB (Vislab) and of 20 dB (Audio Booth). A fourth loudspeaker increases the range of effective CTC up to 8 kHz, with the systems peaking at 16 dB (Vislab) and 20 dB (Audio Booth). Adding a fifth loudspeaker increases the range of operation to the whole audio frequency range, with maximums of 16 dB (Vislab) and obtaining 22 dB at 10 kHz. Using 7 loudspeakers increases the performance above 2 kHz between 2 and 4 dB in the Vislab, however, the performance is very similar to that of the 5 speakers geometry in the Audio Booth.

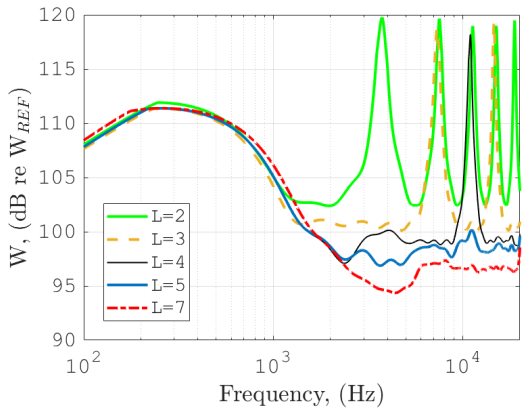
The differences in performance between the 5 speakers and 7 speaker arrays in the two rooms are due to the unequal absorption characteristics of the two rooms. The effect of reflections and the level of reverberant pressure created in a room are directly proportional to the radiated acoustic power of a source and inversely proportional to the



(a) f =1 kHz



(b) f =3 kHz



(c) Radiated power

Figure 6: Radiation patterns for the geometries of study of Fig. 2b and acoustic radiated power. The blue dots in the radiation patterns symbolise the position of the listener's ears.

average absorption of the room walls [22]. Due to this effect, arrays that radiate more energy in all directions will contribute to a larger excitation of reflections on a given room, and depending on the absorption of the room these will contribute more or less into the reverberant pressure. Fig. 6 shows the radiation patterns and the acoustic radiated power for the arrays of study. The radiation patterns show how at 1 kHz, all the array geometries behave in a similar manner, but that at 3 kHz the arrays with larger number of sources beam towards the ipsilateral ear whilst reducing the level of the sidelobes. Similar to the low frequency radiation patterns, the low frequency acoustical power, shown in Fig. 6c, is very similar for all the array geometries, which is due to the fact that all the array geometries present the same aperture. At higher frequencies, the arrays with larger number of sources minimise the radiated acoustical power. It can be seen how the radiated power for the arrays using 3, 4 and 5 sources present peaks in the spectrum. Those peaks are caused by the inefficiency of the low channel array systems to cancel the cross-talk and at the same time produce a uniform sound pressure at the ipsilateral ear. As shown in the simulated reverberant pressures of Fig. 4, these peaks increase the reverberant pressure level at the peak frequencies and can be detrimental for the sound quality.

4. CONCLUSIONS

This document has looked into the effect of the room reflections on Transaural systems using different numbers of loudspeakers (2, 3, 4, 5 and 7). An image source model has been used to simulate the effect of two rooms representing a multi-purpose laboratory and a conditioned listening room.

The simulations manifest how arrays of 2, 3, and 4 drivers give a coloured in-room frequency response, due to the acoustic radiated power also peaking at specific frequencies, corresponding to multiples of half the speaker spacing. The simulated results also show how for loudspeaker arrays using 5 of 7 speakers, a flat frequency response is obtained, which is expected to provide a better sounding experience. The simulated results for the CTC spectrum predict a similar low frequency performance for all the array geometries, however, these exhibit how the array systems using 5 or 7 loudspeakers can achieve a higher CTC in the mid and high frequency range. These findings predict that all the array geometries would be able to place virtual acoustic images well outside of the speaker span, however, the systems using 2, 3, and 4 loudspeakers would struggle to produce very lateralised images and near-to-the-head images, as these require to create large level difference cues that, as predicted, those systems could not achieve in practical environments.

The experiments presented in this document have assumed a steady-state analysis, and cannot model how critic time effects that occur in our hearing system as the precedence effect [23] will affect the rendered virtual acoustic images. Further work is required to understand how the delay between the original reproduced binaural signal and the room reflections affect the 3D performance. Further studies are also required to understand how far into the reverberant field a Transaural system with a given radiation performance can be to render virtual acoustic images.

5. ACKNOWLEDGEMENTS

The authors would like to thank 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 want to thank the Spanish Acoustical Society for their help to the authors on travelling to the Internoise 2019 conference.

6. REFERENCES

[1] Duane H. Cooper and Jerald L. Bauck. Prospects for transaural recording. *J. Audio Eng. Soc*, 37(1/2):3–19, 1989.

[2] S. Atal and R. Schroeder. Apparent sound source translator, February 22 1966. US Patent 3,236,949.

[3] Ole Kirkeby, Philip A. Nelson, and Hareo Hamada. The "stereo dipole": A virtual source imaging system using two closely spaced loudspeakers. *J. Audio Eng. Soc*, 46(5):387–395, 1998.

[4] Ole Kirkeby and Philip A. Nelson. Digital filter design for inversion problems in sound reproduction. *Journal of Audio Engineering Society*, 47(7/8):583–595, 1999.

[5] Darren B. Ward and G.W. Elko. Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation. *Signal Processing Letters, IEEE*, 6(5):106–108, May 1999.

[6] Yesenia Lacouture Parodi and Per Rubak. Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakersa). *The Journal of the Acoustical Society of America*, 128(3):1045–1055, 2010.

[7] Takashi Takeuchi and Philip A. Nelson. Optimal source distribution for binaural synthesis over loudspeakers. *The Journal of the Acoustical Society of America*, 112(6):2786–2797, 2002.

[8] Takashi Takeuchi and Philip A. Nelson. Extension of the optimal source distribution for binaural sound reproduction. In *Proceedings of the Institute of Acoustics Reproduced Sound Conference*, 2008.

[9] Marek Olik, Philip J. B. Jackson, Philip Coleman, and Jan Abildgaard Pedersen. Optimal source placement for sound zone reproduction with first order reflections. *The Journal of the Acoustical Society of America*, 136(6):3085–3096, 2014.

[10] T Takeuchi, P.A. Nelson, O. Kirkeby, and H. Hamaha. The effects of reflections on the performance of virtual acoustic imaging systems. In *Proceedings of the Active 97, The International Symposium on Active Control of Sound and Vibration*, 1997.

[11] Asbørn Saebø. *Influence of Reflections on Crosstalk Cancelled Playback of Binaural Sound*. PhD thesis, 2001.

[12] 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)*, pages 2873–2877, May 2014.

[13] Gregory Vincent White. An objective and subjective investigation into the effects of early reflections on transaural audio reproduction, 2016.

[14] Stephen J. Elliott, Jordan Cheer, Jung-Woo Choi, and Youngtae Kim. Robustness and regularization of personal audio systems. *IEEE Transactions on Audio Speech and Language Processing*, 20(7):2123–2133, 2012.

[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] Marcos F. Simón Gálvez, Stephen J. Elliott, and Jordan Cheer. The effect of reverberation on personal audio devices. *The Journal of the Acoustical Society of America*, 135(5):2654–2663, 2014.

[17] Marcos F. Simón Gálvez, Charles Berkeley, Eric Hamdan, and Filippo M. Fazi. A robustness study for low-channel-count cross-talk cancellation systems. In *Audio Engineering Society Conference: 2019 AES Immersive and Interactive Audio Conference*, Jul 2019.

[18] Filippo Maria Fazi and Eric Hamdan. Stage compression in transaural audio. In *Audio Engineering Society Convention 144*, May 2018.

[19] Michael Vorlander. *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. RWTHedition. Springer, Dordrecht, 2007.

[20] R. A. Haddad and A. N. Akansu. A class of fast gaussian binomial filters for speech and image processing. *IEEE Transactions on Signal Processing*, 39(3):723–727, March 1991.

[21] E. C. Hamdan and F. M. Fazi. Three-channel crosstalk cancellation mode efficiency for sources in the far-field. In *Audio Engineering Society Conference: Conference on Immersive and Interactive Audio*, March 2019.

[22] Leo. L. Beranek. *Acoustics*. American Institute of Physics, New York, 1987.

[23] Ruth Y. Litovsky, H. Steven Colburn, William A. Yost, and Sandra J. Guzman. The precedence effect. *The Journal of the Acoustical Society of America*, 106(4):1633–1654, 1999.