A low-cost loudspeaker array for personal audio with enhanced vertical directivity

Daniel Wallace[1] & Jordan Cheer

Institute of Sound and Vibration Research, University of Southampton

Building 13, University of Southampton, Southampton, SO17 1BJ, UK

ABSTRACT

Personal audio refers to the production of spatially distinct listening zones within a space that is shared by multiple people. This technology can enable each listener to hear their own desired material without affecting others, facilitate private communication in public spaces, or provide independent volume controls when people of different ages watch TV together. This is accomplished using loudspeaker array processing, and the performance of a given system is subject to practical limitations associated with the array, the zonal geometry and the reproduction environment. The directivity of typical line array designs is only controllable in the horizontal plane, with approximately monopole directivity in the vertical direction. This can reduce real-world performance through excitation of the reverberant sound field. To overcome this limitation a design is presented for an eight-channel line array of loudspeakers, with each loudspeaker designed to be directional in the vertical plane. The design uses inexpensive components and the USB audio protocol to lower the barrier to entry into personal audio research and development. In order to form sound zones, input signals must be processed using transfer responses between the array elements and the desired sound zones, and in keeping with the philosophy of cost reduction, two different methods for analytically approximating these responses are compared against anechoic measurements which are expensive to acquire.

1. INTRODUCTION

In spaces shared by multiple people, sound produced for the attention of one listener may become annoying or distracting to others nearby. Personal Audio systems provide listening zones to minimise this interference, and can be generated using a combination of active noise control techniques at low frequencies, loudspeaker array processing at mid-frequencies, and the individual directivity of loudspeaker drivers at high frequencies [1]. The design of such systems has been discussed in various contexts, such as open-plan offices and museum exhibits [1], television and radio systems [2]–[4] entertainment and communication systems in the transport sector [5], [6], mobile devices [7], and the control of speech privacy in public spaces [8], [9].

Several algorithms and approaches have been developed to generate single zones of sound by driving each of the elements of a loudspeaker array with a filtered copy of a target signal. The sound field control problem is therefore reduced to one of filter design. The requirements of this design problem can be summarised as maximising the level of a signal in a given location, the *bright zone*, and minimising it either elsewhere, or in a specific *dark zone.* The acoustic contrast metric [10] provides an intuitive way to judge system performance in this regard – the metric measures the ratio of mean-squared pressure between the bright and dark zones, and

[1] Corresponding author: D.Wallace@soton.ac.uk

can be used as the cost function to optimally design the necessary filters for the L loudspeakers in the array. Formally, acoustic contrast is defined at each frequency as

$$C = \frac{N_d}{N_b} \frac{\boldsymbol{u}^H \boldsymbol{G}_b^H \boldsymbol{G}_b \boldsymbol{u}}{\boldsymbol{u}^H \boldsymbol{G}_d^H \boldsymbol{G}_d \boldsymbol{u}} \tag{1}$$

where N_b and N_d are the number of bright and dark zone optimisation microphones respectively, $\boldsymbol{u}$ is an $L \times 1$ vector of control filters, $\boldsymbol{G}_b$ and $\boldsymbol{G}_d$ are $N_b \times L$ and $N_d \times L$ matrices of transfer responses, and $(\cdot)^H$ denotes the Hermitian transpose. The filters $\boldsymbol{u}$ are determined frequency-by-frequency as the eigenvector corresponding to the largest eigenvalue of the matrix

$$[\boldsymbol{G}_d^H \boldsymbol{G}_d + \lambda \boldsymbol{I}]^{-1} [\boldsymbol{G}_b^H \boldsymbol{G}_b], \tag{2}$$

where λ is a regularisation parameter, selected to reduce numerical errors that originate from the inverse of $[\boldsymbol{G}_d^H \boldsymbol{G}_d]$, which may be ill-conditioned. This is particularly significant for compact loudspeaker arrays at low frequencies, as the columns of the transfer responses in $\boldsymbol{G}_d$ are almost equal. The true transfer responses $\boldsymbol{G}$ in Eqs. 1 and 2 may be nonlinear and time-variant, and in practice, $\boldsymbol{u}$ is generated using estimates of the transfer response matrices $\widehat{\boldsymbol{G}}$. These may be captured using impulse response measurements or analytically using source models.

The present paper demonstrates a simple, reproducible design for a personal audio loudspeaker array and evaluates its performance when both analytical and measured transfer responses are used to optimise the array filters. The array offers control over horizontal directivity and enhanced vertical directivity over arrays with a similar footprint, to reduce the detrimental effects of floor and ceiling reflections.

2. ARRAY DESIGN

To produce sound zones using a single, compact loudspeaker array, each loudspeaker in the array must be independently controllable, and positioned relative to the others in a way that facilitates sound field control in the plane(s) of interest. To give a prototypical example, a uniform linear array of point-sources can be processed to radiate directionally, with the maximum response forming a conical shell, coaxial with the array. The spacing between the array elements, d, limits the maximum frequency, f_{max}, that the array can reliably beamform without incurring spatial aliasing; $d \leq 0.5\ c/f_{max}$, where c is the speed of sound. For c = 343 m/s, and speech bandwidth up to 4 kHz, this corresponds to a maximum inter-element spacing of 43 mm.

If array-based sound field control is limited to the horizontal plane, control in the vertical direction is naturally compromised, and in reverberant environments, this can lead to a loss of performance. Excitation of the reverberant field reduces acoustic contrast by raising the background noise level outside of the direct field of the array, and in certain room geometries, uncontrolled specular reflections from the ceiling and floor can leak into the dark zone, again reducing acoustic contrast [2], [11]. Therefore, it is prudent to also consider a means for vertical directivity control. Although this can be achieved using a 2-dimensional array of loudspeakers, this is not cost efficient.

An increase in the vertical directivity can alternatively be realised through the geometrical properties of the loudspeakers used in the array. Figure 1 shows the theoretical vertical directivity patterns of a series of narrow rectangular pistons, representing loudspeaker diaphragms mounted vertically in an infinite baffle [12]. The key parameter in the predicted vertical directivity is the product of the acoustic wavenumber k and the vertical dimension of the piston a. When $ka < 1$, the source approximately radiates with a monopole directivity, while increasing the size of the radiator, or the frequency, yields a significant increase in the vertical directivity. In the case of the practical array design, the vertical directivity can thus be increased by using loudspeakers with a large diaphragm. This, however, potentially conflicts with the small spacing required between loudspeakers to increase the spatial aliasing frequency with respect to beamforming in the horizontal plane. However, loudspeaker drivers intended for fitment in televisions, such as those depicted in Figure 2, often have the desired high aspect ratio form-factor, with the added benefits of broad availability and wide audio bandwidth in a single device – this simplifies the acoustic modelling of these drivers. Multiple drivers could also be stacked vertically and driven in parallel to increase the vertical directivity at lower frequencies [2].

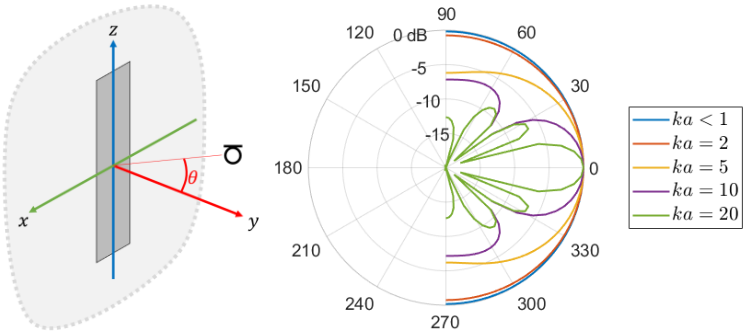

Figure 1: Geometry and vertical directivity pattern of a narrow rectangular piston in an infinite baffle.

Figure 2: Selection of high aspect ratio loudspeaker drivers designed for fitment in televisions. Left to Right: Toshiba EAS16S15A (used in the presented design) Toshiba AA-31816F-XX2G42, Samsung BN83-01803A, Dayton Audio CE140-30S-8, Dayton Audio HARB252-8 LG EBR71208201.

Typically, custom-designed loudspeaker array systems are costly to implement. Whilst the loudspeaker drivers themselves can be acquired relatively cheaply, there is a pronounced increase in the price of digital-to-analogue converters and amplification as the channel count increases. This is due to the requirement to use more capable multi-channel audio communication standards such as MADI or Dante. The loudspeaker array presented in this paper uses a consumer-grade 7.1-channel USB audio card, whose outputs are routed to Class-D amplifier circuits. Figure 3 shows a block diagram of the equipment used in the presented design.

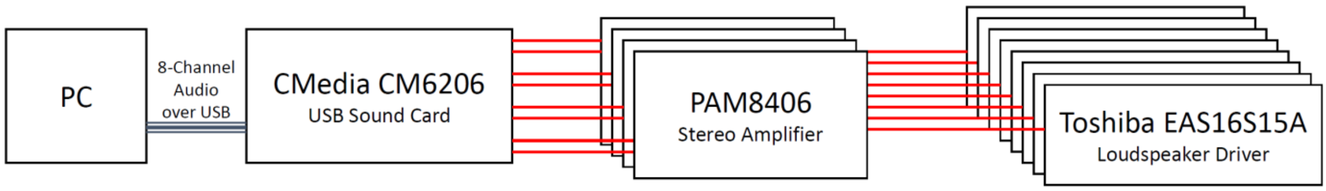

Figure 3: Loudspeaker array system block diagram.

The 8-channel array presented in this paper is designed to be simple to construct, using readily available consumer electronics and a single 1220 x 610mm sheet of 6mm plywood, leading to a total material cost of under £100. Figure 4 shows the construction of the cabinet; each loudspeaker is enclosed within its own sealed cabinet within an outer box. The presented design uses Toshiba EAS16S15A drivers, with baffle cut-out dimensions of 145 mm × 28 mm. The minimum horizontal distance between these drivers using a 6 mm cabinet wall is 42 mm, satisfying the inter-element spacing requirement for preventing spatial aliasing up to 4 kHz. The outer enclosure also houses the USB DAC, amplifiers and power supply, shown in the right panel of Figure 5. The only external connections to the device are a Type B USB socket and an IEC power inlet. The class-compliant USB DAC used in the design requires no additional drivers, enabling the individual channels to be driven using a standard digital audio workstation or programming environment. A parametric SolidWorks design that can be adapted to other driver sizes is available at https://doi.org/10.5258/SOTON/D1218.

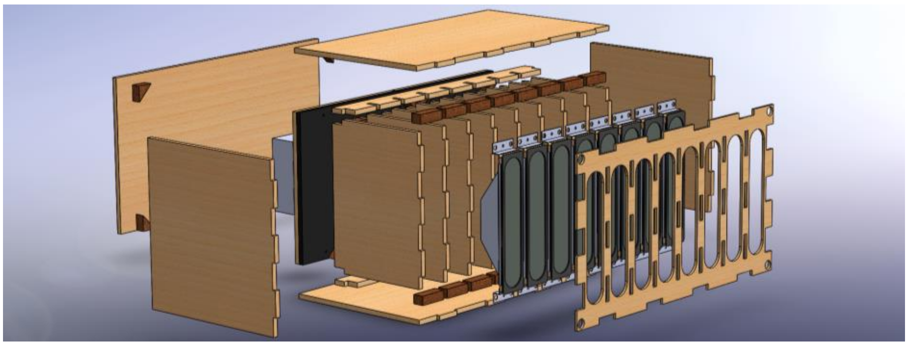

Figure 4: Exploded view of the loudspeaker array design.

Figure 5: 8-Channel loudspeaker array. Left: Front view; Right: Rear view with back panel removed.

3. TRANSFER RESPONSE MEASUREMENTS AND MODELLING

The Acoustic Contrast Control process requires an estimate of the transfer responses between the loudspeaker array elements and microphones situated in each sound zone, $\widehat{\boldsymbol{G}}_b$ and $\widehat{\boldsymbol{G}}_d$. With all other factors held constant, the more accurately that $\widehat{\boldsymbol{G}}$ matches the true responses $\boldsymbol{G}$, the greater the level of control over the sound field. Impulse response measurements capture all relevant physical information, such as the true directivity of the sources and any mismatches in sensitivity between drivers at the time of recording. These are expensive and time-consuming to obtain and are inflexible if different zone positions are required at a later stage. Acoustic source models based on the geometry of the loudspeaker array offer this flexibility and are also naturally free from noise, which can result in higher audio reproduction quality at the expense of acoustic contrast [13], [14]. In this study, two acoustic models of different complexities are used to provide the modelled responses; the array sources are modelled as either point monopoles radiating into a free field environment or rectangular pistons, mounted in an infinite baffle, with the expectation that the rectangular piston model will provide a better approximation to the true directivity at high frequencies.

In order to test the array in optimal conditions, ground truth transfer response measurements between the 8-channel loudspeaker array and a 20-channel array of omnidirectional measurement microphones were acquired in the ISVR Large Anechoic Chamber. Angular measurements were simplified by placing the loudspeaker array on a turntable, and were captured using the swept-sine method [15]. Figure 6 shows the measurement microphone locations relative to the loudspeaker array for the horizontal and vertical directivity measurements – the vertical measurements were carried out with the loudspeaker array in portrait orientation on the turntable. A coarser grid of positions was used to the rear of the array as in most envisaged practical situations, finer grained angular control in front of the array is required. Furthermore, attenuation of rearward radiation at high frequencies will intrinsically be provided by the sealed cabinet design.

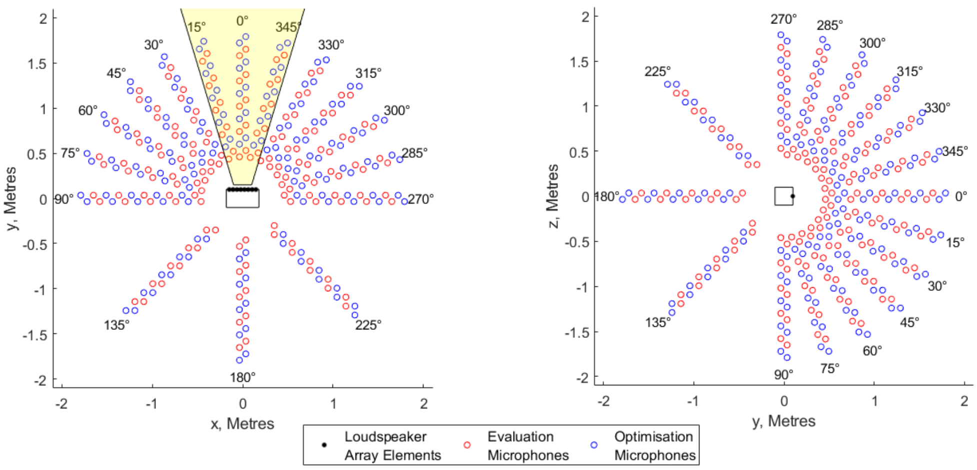

Figure 6: Microphone array positions for horizontal (Left) and vertical (Right) directivity measurements.

In the results that follow, the bright zone is defined by all horizontal optimisation microphones that subtend a 30° angle from the centre of the array, namely measurement lines 0°, 15° and 345°, as indicated in Figure 6 by the shaded region. All other microphones in the horizontal plane form the dark zone. Separate evaluation microphones, shown in red in Figure 6, are used to measure the reproduced sound field, to avoid bias in the estimates of acoustic contrast and directivity, compared to re-using the optimisation measurement positions [16].

4. RESULTS

The loudspeaker array is configured to generate a single "on-axis" zone of width 30 degrees, and to minimise horizontal radiation at other angles. The sound pressure spectra at each microphone location is calculated as $\boldsymbol{p} = \widehat{\boldsymbol{G}}\boldsymbol{u}$, where $\widehat{\boldsymbol{G}}$ is the matrix of measured evaluation transfer responses. These spatial and spectral results are visualised in Figures 7-9.

Figure 7 shows the octave band-averaged horizontal directivity from 125 Hz to 8 kHz when three different transfer response estimates are used to generate the sound zoning filters. Figure 8 shows the accompanying vertical directivity, and Figure 9 shows the variation in the acoustic contrast and on-axis sound pressure level with frequency. The greatest horizontal directivity is achieved when measured transfer responses are used in the acoustic contrast control process, and there are only marginal differences between the monopole and baffled rectangular piston models for the presented zonal geometry, across the frequency range. The rear radiation from the array is attenuated by 5 dB at 500 Hz and 10 dB at 2 kHz. The vertical directivity shown in

Figure 8 is unaffected by the choice of transfer response estimate as the generation of zonal filters uses no information from the vertically oriented measurement microphones. Extending the bright zone to a three-dimensional, 30° cone and including the off-axis vertical measurement positions in the dark zone would provide additional constraints to the beamforming process in a direction in which the array has no beamforming capability. These additional constraints would only serve to increase the condition number of $\boldsymbol{G}_d^H \boldsymbol{G}_d$ in Eq. 2 and thus may reduce robustness to uncertainty.

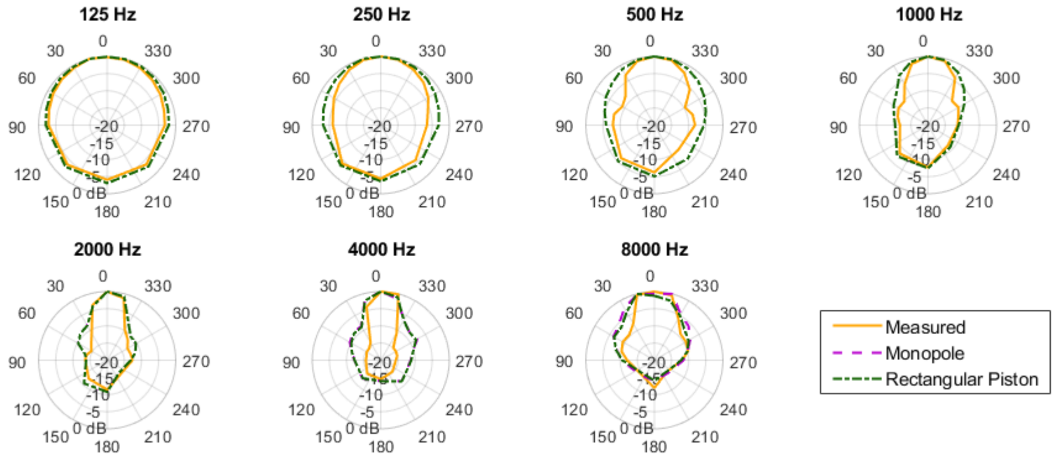

Figure 7: Horizontal directivity with a single forward zone.
Each plot shows directivity results averaged over an octave band.

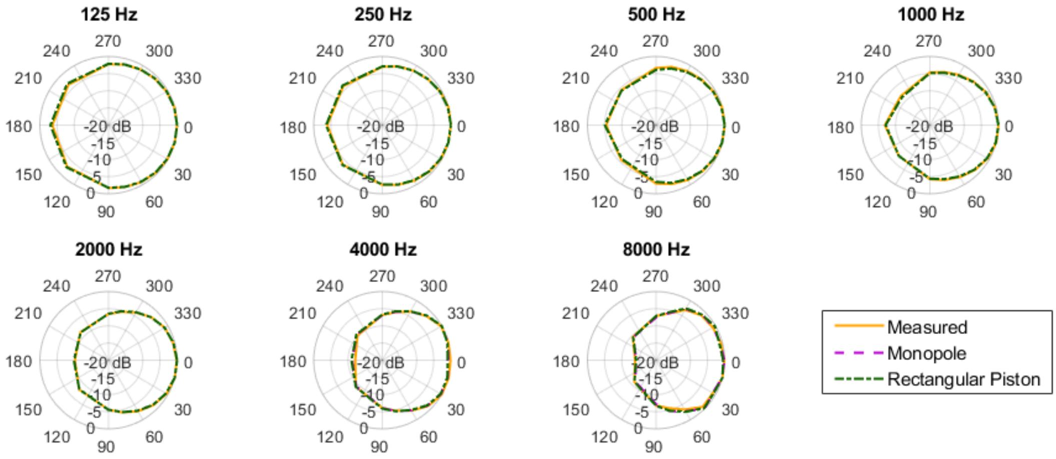

Figure 8: Vertical directivity with a single forward zone.
Each plot shows directivity results averaged over an octave band

Figure 9a exemplifies that the most pronounced differences between the measured and modelled results occur at frequencies greater than 3 kHz. Below this frequency, in terms of acoustic contrast, the system based on measured responses outperforms those using modelled responses by around 5 dB. This is caused by mismatches in the sensitivity and the frequency response between the array elements. The impact of this mismatch was minimised by individually adjusting the amplifier gains such that broadband noise emitted from each loudspeaker, recorded at an on-axis point in the far field, matched within 1 dB SPL prior to the transfer response measurements being taken. At low frequencies, acoustic contrast is limited by the overall aperture of the array; for wavelengths significantly longer than this dimension, the entire array appears as a monopole source. Some potential extensions to low-frequency control are possible with non-uniform array spacing, at the expense of beam-width at higher frequencies [17].

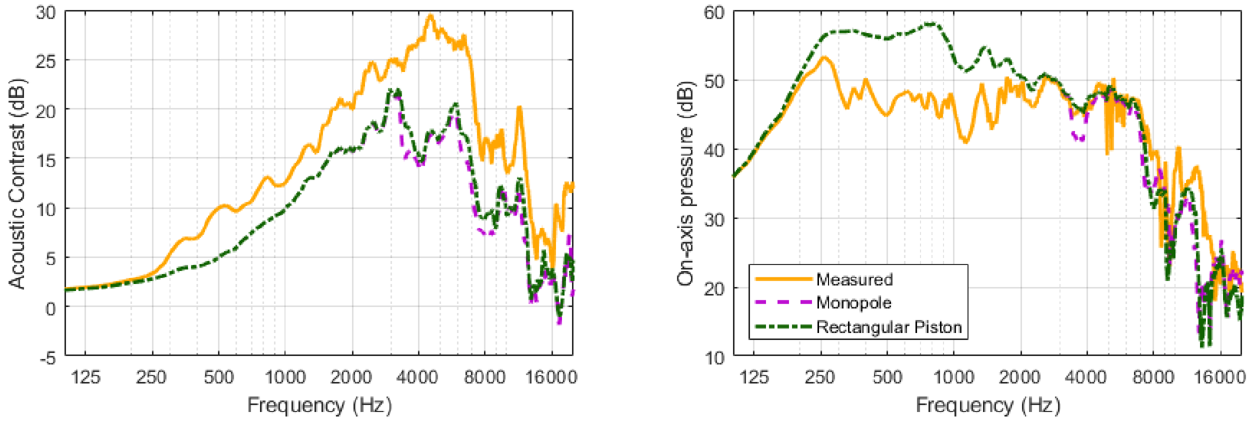

Figure 9: Left: Experimental acoustic contrast using measured and modelled responses to optimise array filters. Right: On-axis (bright zone) sound pressure level from array using measured and modelled responses.

In addition to the sensitivity issues described above, at frequencies above 3 kHz, discrepancies between the modelled and true transfer responses cause the acoustic contrast to decrease, while the measured contrast continues to increase up to 4.4 kHz, whereupon spatially aliased beams begin to impinge on the dark zone. At 3 kHz, the wavelength of sound in air is comparable to the dimensions of the loudspeaker cabinet, meaning that both the monopole model and the model of the rectangular piston in an infinite baffle poorly approximate the true transfer response matrix. Given the marginal performance increase when moving from the simple monopole source model to the more complex rectangular piston model exhibited in Figures 7-9, it is hypothesized that the mismatch between array elements is the dominant source of error between the modelled transfer responses, which assumes matched drivers, and the true transfer responses – accordingly, without measurements and subsequent digital correction for this mismatch, predicting the radiation pattern using a more detailed technique such as finite element modelling is unlikely to yield significantly better performance. In practical, reverberant environments, lower levels of acoustic contrast are expected overall due to increased leakage from the bright zone into the dark zone. Additionally, without re-measuring the responses in-situ, the difference in acoustic contrast between using modelled and measured responses may be less significant in these environments, as the mis-match between the reverberant $\boldsymbol{G}$ and anechoic $\widehat{\boldsymbol{G}}$ is greater.

5. CONCLUSIONS

The intention of the presented paper is to introduce a reproducible, low-barrier-to-entry testbed for personal audio research that does not require specialist acoustical measurement equipment or facilities to deliver high performance. To this end, a design for a compact loudspeaker array has been presented. The device can be constructed using low-cost components and is driven using the ubiquitous USB audio protocol. High aspect ratio loudspeaker drivers are used to improve vertical directivity over circular drivers with the same inter-element horizontal spacing. This aims to reduce excitation of the reverberant field by reducing the impact of floor and ceiling reflections.

The performance of the array has been evaluated in anechoic conditions, using both measured and modelled transfer responses to optimise acoustic contrast control filters. The highest performance is achieved using measured responses, as this captures the true directivity and any inherent mismatch in sensitivity and frequency response between the loudspeaker drivers, though comparable performance is achieved without the use of any acoustical measurements. Using only the geometry of the loudspeaker array and the desired positions of the sound zones, acoustic contrast in excess of 10 dB can be achieved between the bright and dark zones from 1-10 kHz. The limiting factor in the acoustic contrast is the matching between loudspeaker drivers; with compensation for this, or by selecting drivers with better matching, the performance deficit compared to using measured responses is expected to be narrowed.

6. REFERENCES

[1] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, “Personal Sound Zones,” *IEEE Signal Process. Mag.*, no. March, pp. 81–91, 2015.

[2] M. F. Simon Galvez, S. J. Elliott, and J. Cheer, “Personal Audio Loudspeaker Array as a Complementary TV Sound System for the Hard of Hearing,” *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E97.A, no. 9, pp. 1824–1831, 2014.

[3] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, “A realization of sound focused personal audio system using acoustic contrast control.,” *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2091–2097, 2009.

[4] K. R. Baykaner, C. Hummersone, R. Mason, and S. Bech, “The acceptability of speech with interfering radio program material,” *AES 136th Conv.*, pp. 1–9, 2014.

[5] J. Cheer, S. J. Elliott, and M. F. S. Gálvez, “Design and implementation of a car cabin personal audio system,” *AES J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 412–424, 2013.

[6] S. J. Elliott and M. Jones, “An active headrest for personal audio,” *J. Acoust. Soc. Am.*, vol. 119, no. 5, p. 2702, 2006.

[7] S. J. Elliott, J. Cheer, H. Murfet, and K. R. Holland, “Minimally radiating sources for personal audio.,” *J. Acoust. Soc. Am.*, vol. 128, no. 4, pp. 1721–8, 2010.

[8] J. Donley, C. H. Ritz, and W. B. Kleijn, “Multizone Soundfield Reproduction With Privacy and Quality Based Speech Masking Filters,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 4, pp. 1–15, 2018.

[9] D. Wallace and J. Cheer, “Optimisation of Personal Audio Systems for Intelligibility Contrast,” in *Proc. 144th Audio Engineering Society Convention*, 2018.

[10] J.-W. Choi and Y.-H. Kim, “Generation of an acoustically bright zone with an illuminated region using multiple sources,” *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700, 2002.

[11] M. F. Simón-Gálvez, S. J. Elliott, and J. Cheer, “The effect of reverberation on personal audio devices,” *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2654–2663, 2014.

[12] P. P. Morse and K. U. Ingard, *Theoretical Acoustics*, 1st ed. New York: McGraw-Hill, 1968.

[13] D. Wallace and J. Cheer, “The Design of Personal Audio Systems for Speech Transmission using Analytical and Measured Responses,” in *Proc. 44th International Conference on Acoustics, Speech and Signal Processing*, 2019.

[14] 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,” *Proc. 52nd Audio Eng. Soc. Int. Conf.*, pp. 1–10, 2013.

[15] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *Proc. AES 108th conv, Paris, France*, 2000.

[16] M. A. Akeroyd *et al.*, “The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics,” *J. Acoust. Soc. Am.*, vol. 121, no. 2, pp. 1056–1069, 2007.

[17] M. Van Der Wal, E. W. Start, and D. De Vries, “Design of Logarithmically Spaced Constant - Directivity Transducer Arrays,” *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 497–507, 1996.