A Comparative Performance Study of Sound Zoning Methods in a Reflective Environment

Marek Olik[1], Jon Francombe[2], Philip Coleman[1], Philip J.B. Jackson[1], Martin Olsen[3], Martin Møller[3], Russell Mason[2], and Søren Bech[3]

[1]*Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey, GU2 7XH, UK*

[2]*Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, UK*

[3]*Bang & Olufsen A/S, Peter Bangs Vej 15, 7600, Struer, Denmark*

Correspondence should be addressed to Marek Olik (`m.olik@surrey.ac.uk`)

ABSTRACT

Whilst sound zoning methods have typically been studied under anechoic conditions, it is desirable to evaluate the performance of various methods in a real room. Three control methods were implemented (delay and sum, DS; acoustic contrast control, ACC; and pressure matching, PM) on two regular 24-element loudspeaker arrays (line and circle). The acoustic contrast between two zones was evaluated and the reproduced sound fields compared for uniformity of energy distribution. ACC generated the highest contrast, whilst PM produced a uniform bright zone. Listening tests were also performed using monophonic auralisations from measured system responses to collect ratings of perceived distraction due to the alternate audio programme. Distraction ratings were affected by control method and programme material.

1. INTRODUCTION

Due to the increased popularity of personal electronic devices in recent years, the presence of conflicting audio streams in a single listening space has become a familiar problem affecting the experience of listening to the audio content delivered by such devices. Headphones are commonly used to minimise interruption from competing programmes, however in some situations avoiding headphones may be desirable (e.g. when listening to audio over long periods or to allow better audibility of background sounds). To tackle the problem, several signal processing methods have been proposed, which use interference between loudspeaker array signals to deliver audio content to specified regions whilst limiting audibility in other areas. Existing methods can be categorised into three major groups. In the first group, the zoning effect is achieved by focusing the sound energy towards the bright zone (beamforming); methods include delay and sum beamforming [1] and brightness control [2]. The second group concentrates on maximising the acoustic energy in the bright (target) zone, whilst minimising it in the dark (quiet) zone. Notable methods include acoustic contrast control [2] and acoustic energy difference maximisation [3]. The third method group has the additional capability of synthesising a desired sound field (e.g. a plane wave) in the zones. Such methods are based on analytical sound field synthesis [4] or direct optimisation of a sound field (pressure matching [5]).

Examples of sound zone method implementations in anechoic [6, 7, 3, 8] and reflective environments [9, 10, 11] can be found in the literature. However, there are limited examples of comparative evaluation of different methods [11, 12, 13]. In this paper, performance characteristics of each of the groups of methods are compared for a system implemented in a room. One representative method was chosen from each group: delay and sum (DS), acoustic contrast control (ACC) and pressure matching (PM). The system constituted two independent loudspeaker arrays based on configurations commonly used in previous studies: linear and circular arrays (each 24 elements). The system performance was evaluated in terms of the ratio of sound energy between the bright and dark zones (acoustic contrast) and uniformity of sound energy in the bright zone (qualitative assessment of SPL maps and the planarity metric [14]). To relate the acoustic contrast results to perceived separation, listening tests were also performed using monophonic auralisations made with system responses at the centre of each zone. Distraction ratings were collected using a multiple stimulus presentation.

In Section 2 the sound zone problem considered here is introduced and the physical metrics used for system evaluation are detailed. In Section 3 the theoretical background for the chosen sound zone methods is provided. Sections 4 and 5 contain details of the experimental setup and the sound zone reproduction process respectively. In Section 6 the performance results based on the physical metrics are discussed. The listening tests are detailed in Section 7.

2. BACKGROUND

In this section, the sound zone problem is introduced, along with physical evaluation metrics.

2.1. Sound Zone Problem Definition

Figure 1 shows an example sound zone system layout. Two audio programmes, A and B, are to be reproduced in zones A and B respectively. In each zone, only the target programme should be audible. For programme A, zone A is the "bright zone" and zone B is the "dark zone". Conversely, for programme B zone B is the "bright zone" and zone A is the "dark zone". The remainder of the room, zone C, is uncontrolled.

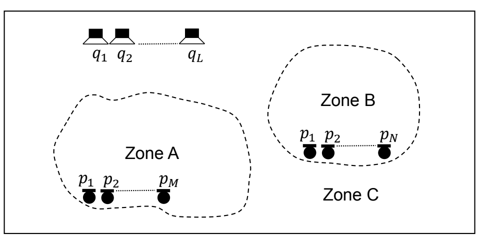


Fig. 1: Example of a sound zone system, with L loudspeakers and listening zones A and B defined by M and N microphones, respectively. Zone C is uncontrolled.

For a single frequency, the source weight vector is defined as $\mathbf{q} = [q_1\ q_2\ \cdots\ q_L]^T$, where q_l is the lth loudspeaker's complex source weight and L is the total number of loudspeakers. The vectors of complex pressures in zones A and B are $\mathbf{p}_A = [p_1\ p_2\ \cdots\ p_M]^T$ and $\mathbf{p}_B = [p_1\ p_2\ \cdots\ p_N]^T$ respectively, where p_m is the complex pressure at the mth microphone in zone A, p_n is the complex pressure at the nth microphone in zone B, and M and N are the number of microphones in zones A and B respectively.

The matrix of transfer functions between each loudspeaker and microphone in zone A is defined as

$$\mathbf{G}_A = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1L} \\ g_{21} & g_{22} & \cdots & g_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ g_{M1} & g_{M2} & \cdots & g_{ML} \end{pmatrix}, \qquad (1)$$

where g_{ml} is the transfer function between the mth microphone in zone A and the lth loudspeaker. The equivalent matrix $\mathbf{G}_B$ can be defined for zone B. The pressures and source weights are related as follows: $\mathbf{p}_A = \mathbf{G}_A\mathbf{q}$ and $\mathbf{p}_B = \mathbf{G}_B\mathbf{q}$.

2.2. Physical Evaluation Metrics

The sound zone methods were evaluated for two physical aspects of performance: acoustical separation between the zones and the spatial distribution of sound energy in the bright zone. The former is related to the perceived interference from the unwanted programme in the dark zone, and the latter may influence the perceived quality of the programme in the bright zone [15]. In particular, a spatially uniform bright zone is desirable to avoid changes in the perceived separation between the target and interfering programmes when the listener moves from one position in the zone to another.

2.2.1. Acoustic Contrast

The acoustical separation between the bright zone A and the dark zone B can be quantified as

$$contrast_{AB} = 10\log_{10}\left(\frac{N\mathbf{q}_{opt}^H\mathbf{G}_A^H\mathbf{G}_A\mathbf{q}_{opt}}{M\mathbf{q}_{opt}^H\mathbf{G}_B^H\mathbf{G}_B\mathbf{q}_{opt}}\right), \qquad (2)$$

where $\mathbf{q}_{opt}$ is the optimal source strength vector and the superscript H denotes conjugate transpose. The acoustic contrast is the ratio of the spatially averaged sound energies in the bright and dark zones, expressed in dB.

2.2.2. Planarity

The planarity metric, recently proposed by Jackson *et al.* [14], quantifies the extent to which the sound field resembles a plane wave. Here, the metric is used to evaluate the bright zone. For the considered systems, high planarity in this zone is indicative of a uniform distribution of sound energy across the zone.

The energy distribution at the microphone array in the bright zone (over incoming plane wave direction) is given by $w_i = \frac{1}{2}|\psi_i|^2$, where $\mathbf{w} = [w_1 \cdots w_i]$ are the energy components at the ith angle and $\psi = [\psi_1 \cdots \psi_i]$ are the plane wave components at the ith angle. The steering matrix $\mathbf{H}$ which maps between the observed pressures at the microphones and the plane wave components can then be defined:

$$\psi = \mathbf{Hp}. \tag{3}$$

A super-directive beamformer (ACC in this case) can be used to determine the steering matrix weights. The planarity of zone A can be defined as the ratio between the energy due to the largest plane wave component in this zone and the total energy flux of plane wave components:

$$planarity_A = \frac{\sum_i w_i \mathbf{u}_i . \mathbf{u}_{\hat{\imath}}}{\sum_i w_i}, \tag{4}$$

where $\mathbf{u}_i$ is the unit vector associated with the ith component's direction, $\mathbf{u}_{\hat{\imath}}$ is the unit vector in the $\hat{\imath}$th direction $\hat{\imath} = \arg\max_i w_i$, and . denotes the inner product.

3. SOUND ZONE METHOD THEORY

The three sound field control methods are introduced in the following sections.

3.1. Delay and Sum Beamforming

DS represents the beamforming approach [1], where the sound energy is focused towards the the bright zone. A delay is applied to each loudspeaker feed, so that the individual contributions combine in phase at the centre of the the bright zone, producing an energy peak. The optimal source strength vector is defined as

$$\mathbf{q}_{opt} = \begin{bmatrix} e^{(-j\omega\tau_1)} & e^{(-j\omega\tau_2)} & \cdots & e^{(-j\omega\tau_L)} \end{bmatrix}^T, \tag{5}$$

where $\tau_1, \tau_2, \ldots, \tau_L$ are time delays given by $\tau_l = (max\{r_l\}_{l=1}^{L} - r_l)/c$, where r_l is the distance between the lth loudspeaker and the bright zone centre and c is the speed of sound.

Outside the bright zone, the energy is partially cancelled due to destructive interference. For practical loudspeaker arrays, the phase differences between individual waves are small at low frequencies, so the non-target wave cancellation is limited.

3.2. Acoustic Contrast Control

ACC represents the sound energy control group of methods and aims to maximise energy in one region (bright zone A) while minimizing it in another (dark zone B), by maximizing the following cost function [2]:

$$C = \frac{N\mathbf{p}_A^H \mathbf{p}_A}{M\mathbf{p}_B^H \mathbf{p}_B} = \frac{N\mathbf{q}^H \mathbf{G}_A^H \mathbf{G}_A \mathbf{q}}{M\mathbf{q}^H \mathbf{G}_B^H \mathbf{G}_B \mathbf{q}}. \tag{6}$$

Choi and Kim [2] showed that the vector $\mathbf{q}_{opt}$ which maximises C is proportional to the eigenvector of the matrix $[\mathbf{G}_B^H \mathbf{G}_B]^{-1} \mathbf{G}_A^H \mathbf{G}_A$ that corresponds to its largest eigenvalue. For certain geometrical arrangements, the matrix $\mathbf{G}_B^H \mathbf{G}_B$ can become close to singular, which can cause large errors in the numerical inversion as well as impractical levels of loudspeaker driving signals. These problems can be alleviated by adding a regularisation parameter β to the main diagonal of this matrix, so the matrix for eigenvalue decomposition becomes $[\mathbf{G}_B^H \mathbf{G}_B + \beta\mathbf{I}]^{-1} \mathbf{G}_A^H \mathbf{G}_A$, where $\mathbf{I}$ is an $L \times L$ identity matrix.

3.3. Pressure Matching

PM aims to match the reproduced sound field with the desired one, specified for the bright and dark zones. For a typical sound zone scenario, the desired sound field $\mathbf{d}$ is defined as a plane wave [5], heavily attenuated in the dark zone. The optimisation cost function aims to minimise the error $\mathbf{e}$ between the desired and reproduced sound fields:

$$J = \mathbf{e}^H \mathbf{e} + \beta \mathbf{q}^H \mathbf{q}, \tag{7}$$

where $\mathbf{q}^H \mathbf{q}$ is a power constraint and β is the regularisation parameter. The power constraint is imposed to limit the loudspeaker drives and improve the robustness to errors in the solution. The optimal source strength vector can be found by solving for $\mathbf{q}$:

$$\mathbf{q}_{opt} = (\mathbf{G}^H \mathbf{G} + \beta \mathbf{I})^{-1} \mathbf{G}^H \mathbf{d}, \tag{8}$$

where $\mathbf{G} = [\mathbf{G}_A \mathbf{G}_B]^T$ is the matrix of system transfer functions.

4. EXPERIMENTAL SETUP

In this section, the experimental setup for the comparative evaluation of sound zone methods is detailed. The system was installed in a room (approx. volume 320 m^3) which had been acoustically treated to reduce the influence of surface reflections on method performance. The room's reverberation times (RT60), estimated in 250, 500, 1000, 2000 and 4000 Hz octave bands were 0.33, 0.30, 0.28, 0.27 and 0.25 s, respectively. The loudspeaker and zone layout is presented in Fig. 2. A linear and circular array, each used independently, were chosen to represent arrangements typically used in the literature for the sound zone methods considered (line for DS, circle for PM, and both arrangements for ACC). The line array comprised two rows of 12 loudspeakers (10 cm × 10 cm × 16 cm) fixed side to side. One row of units was placed on top of the other and shifted to obtain

5 cm spacing between the consecutive transducers. For the circular arrangement, 24 loudspeakers were mounted with regular spacing (45.7 cm) on a frame of 1.75 m radius. The zones were two 53 cm × 53 cm regions, located 1.51 m above the floor. A square grid of 144 microphones was used (3.2–5.0 cm spacing) in each zone to measure system responses required for source weight calculation, performance prediction and evaluation.

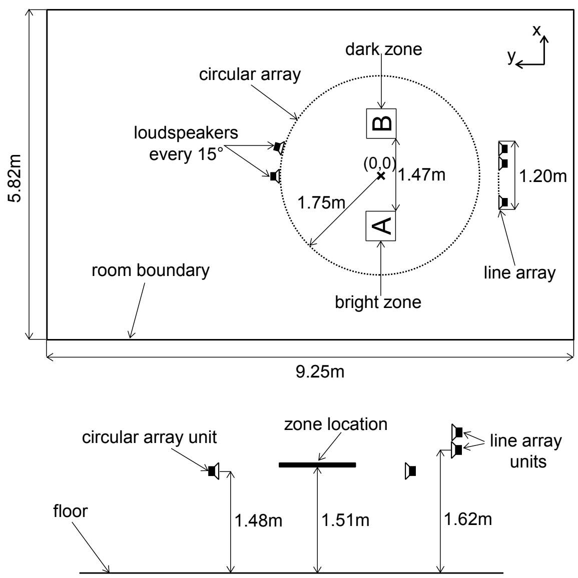


Fig. 2: Experimental sound zone system layout (not to scale). Top: plan view; bottom: cross-section.

5. SOUND ZONE REPRODUCTION PROCEDURE

The sound zone reproduction scenario examined here was the reproduction of a sound programme in zone A (bright zone) and its suppression in zone B (dark zone). For the PM method the specified sound field was a plane wave propagating in the direction of the positive y-axis in the bright zone, and a similar plane wave attenuated by 60 dB in the dark zone.

For the DS method, the optimal source strengths were derived analytically, based on the geometry of the arrangement and the speed of sound value ($c = 345$ m/s) estimated from the environmental conditions in the room. For the remaining cases, the responses between loudspeakers and microphones in the zone locations were measured using the maximum length sequence technique (MLS) and used to populate the transfer function matri-

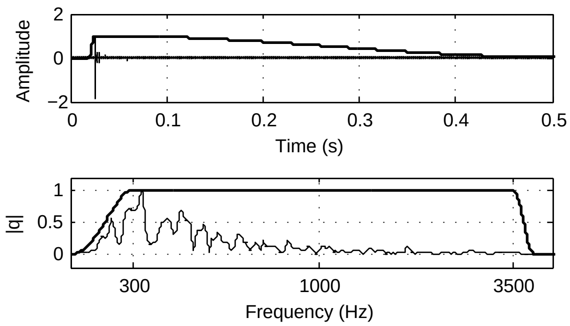


Fig. 3: A normalised impulse response measured in the room with a superimposed cropping window (top) and a normalised magnitude of source weight with a superimposed band pass filter window (bottom).

ces. Prior to the calculation of filter weights, the measured responses were processed in order to reduce noise, improve robustness to mismatch between setup and playback conditions, and limit the audibility of artefacts introduced to programme material by the sound zone filters (pre- and post-ringing [16]). Each impulse response was cropped 0.5 s after the direct sound by multiplying it with an asymmetrical raised-cosine window, as shown in Fig. 3 (top). Subsequently, each impulse was divided into segments with logarithmically increasing lengths and a low-pass filter (an 80 dB/decade slope) with a cut-off frequency decreasing in 1/12 octave steps was applied on a block by block basis. This filtered out high frequency noise occurring in the impulse tails. The source weights were calculated for individual frequency bins before being combined to form a filter frequency response. The ACC solutions were regularised with a frequency-independent regularisation parameter $\beta = 0.3$. For PM, a frequency-dependent regularisation parameter was determined using the L-curve method [17]. Each filter was then band-limited to the range 300–3500 Hz, as shown in Fig. 3 (bottom). The resulting filters were transformed to the time domain for convolution with programme material. For evaluation of performance, the filters were convolved with an MLS and the total system response, which included contributions from all sources active for a given arrangement, was measured at the microphone positions.

Since independent response measurement sets were unavailable for system setup and performance predictions, the responses from half of the microphones were used to calculate the source weights (for the ACC and PM

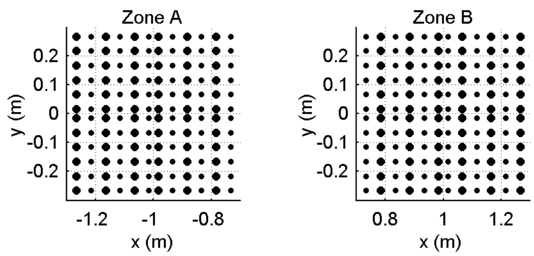


Fig. 4: Microphone positions in the zones. Large bullets indicate system setup locations. All positions were used for performance evaluation.

methods), as shown in Fig. 4. In this way, the numerically independent responses (from the remaining locations) could be included in the full response set used for predicting performance. This limited the bias in the predictions [18]. In evaluating the measured performance, all microphone locations were considered. Sensitivity of the room response inversion to geometrical mismatch between the setup and playback locations is a known problem [19]; including both setup and non-setup locations in the evaluation accounted for this sensitivity in the results.

6. PHYSICAL EVALUATION

In this section, the systems described above are evaluated using the physical metrics detailed in Sec. 2.2. Maps of sound pressure in the bright zone are also used to qualitatively assess the degree of sound field planarity and zone uniformity at a single frequency.

Figure 5 shows the predicted and measured acoustic contrast plotted against frequency. The predictions and measurements show good agreement. The ACC method implemented on the line array (ACC Line) achieved the best contrast in the majority of the considered frequency range. The highest contrast is obtained in the range 1–2.5 kHz, where it fluctuates in the proximity of 20 dB. Outside this range, a roll-off towards 15 dB and lower is observed. At low frequencies, this is due to the limited directivity of the array; in the higher range the method begins to lose control over the interference between the array element contributions as the spatial aliasing limit (3450 Hz) is approached. A similar contrast curve characteristic is observed for the DS method, however the low frequency drop is greater due to the limited ability to maintain a focused beam (see Sec. 3.1). The ACC method based on the circular array (ACC Circle) exhibited the best contrast in the low frequency range (below

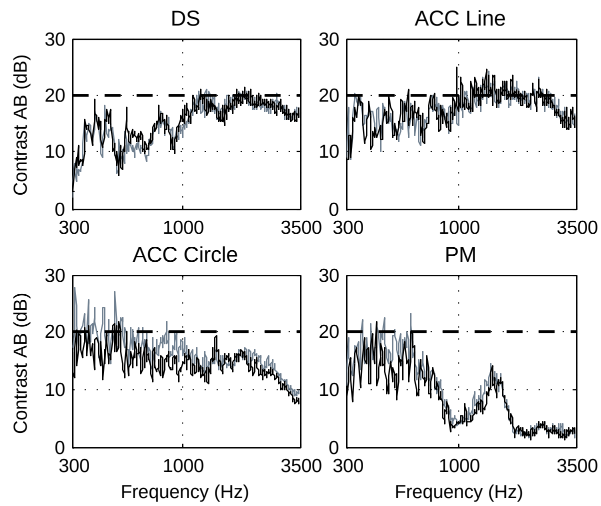


Fig. 5: Acoustic contrast predicted (light) and measured (dark) for all four cases: DS (top-left), ACC Line (top-right), ACC Circle (bottom-left) and PM (bottom-right). The dashed line indicates a reference 20 dB contrast.

500 Hz), but the contrast dropped gradually with frequency. Here, large spacing of the sources (45.7 cm) in the circular array facilitated destructive interference in the dark zone for long wavelengths, but hindered control for shorter wavelengths. For PM, the circular array imposed an upper limit of 508 Hz for reproduction of the desired sound field across the circular region, concentric to the array and extending over the zones (1.29 m radius) [20]. Consequently, the method was able to produce relatively high contrast up to approximately 700 Hz, where the contrast started to decrease rapidly. Interestingly, a contrast peak of 10–15 dB is observed at frequencies around 1.5 kHz. This can be attributed to an energy null due to uncontrolled destructive interference occurring locally in the dark zone.

Figure 6 shows the measured sound pressure maps (real part and level) in the bright zone, at 1.5 kHz for all four cases. Differences in the sound field characteristics can be observed. The DS and ACC Line methods tend to generate a highly planar wave propagating from the array in the direction of the bright zone, as can be observed from the wavefronts in Fig. 6a. Conversely, the ACC Circle method generates a standing wave pattern with pressure amplitudes rapidly changing throughout the zone. The PM method succeeds in creating a plane wave in the specified direction (along the y-axis). This shows that the method is able to control the bright zone sound field well, even above the spatial aliasing limit. The standing wave pattern observed for the ACC Circle method results from the multi-directional character of the sound field generated by this arrangement. The loudspeakers located at different directions with respect to the zone generate a sound field made up of plane-wave components arriving with similar amplitude from various directions. The interfering components yield a non-uniform energy distribution across the zone: sound pressure level (SPL) differences of up to 20 dB between some of the locations can be observed in Fig. 6b. In contrast, planar fields generated by the DS, ACC Line, and PM methods result in relatively homogeneous zones, also shown in Fig. 6b. In all three cases an energy beam is formed, covering the majority of the zone, but tending to lose intensity towards the zone margins. The energy roll-off is very rapid in the upper right corner of the zone for the DS method, where the SPL drops by nearly 20 dB in relation to the beam centre. This can be attributed to a mismatch between the actual loudspeaker and zone positions and those assumed when calculating the DS source weights, which led to a shifting of the beam from the centre of the zone.

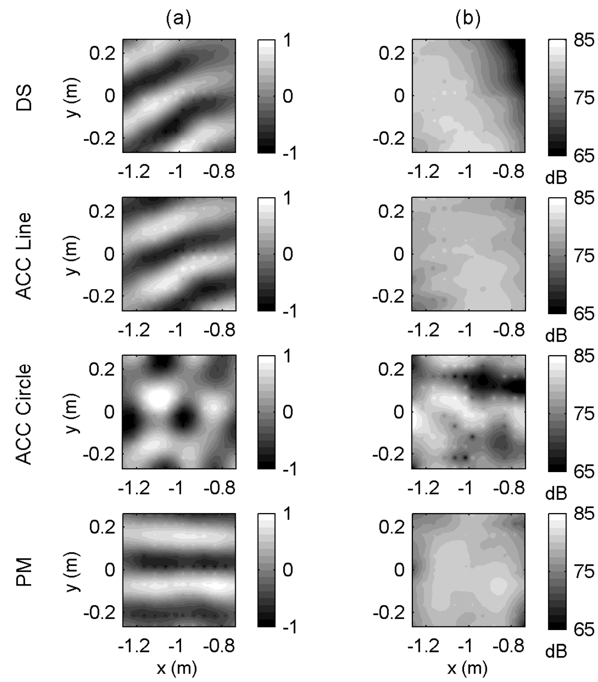


Fig. 6: Bright zone (a) sound pressure (real part) and (b) SPL maps at 1.5 kHz based on performance measurements. From top: DS, ACC Line, ACC Circle and PM.

Figure 7 shows the planarity scores in the bright zone based on predictions and performance measurements. A close match between the two sets of results can be observed. The ACC Circle method gives the lowest overall planarity result, with the highest score just above 80% and rapid variations of planarity across frequency. This indicates that the non-planar characteristics of the sound field generated by this method (observed in Fig. 6 at 1.5 kHz) are maintained at other frequencies. The DS and ACC Line methods obtain very similar planarity scores and exhibit a highly planar sound field in the range 600–3000 Hz (planarity over 80% in the majority of the range). Outside this range, two major dips can be observed in the proximity of 340 Hz and 520 Hz. These can be attributed to limited directivity of the line array at low frequencies, resulting in a strong first-order reflection from the wall nearest to the zone (see Fig. 2). At these two frequencies, the phase relationship between the direct wave and reflection encourages strong interference, generating a standing wave and thus limiting the influence of the planar direct part of sound field on the planarity score. The gradual loss of planarity above 3 kHz for both arrangements can be attributed to the increasing contribution from the side lobes as the aliasing frequency is approached (3450 Hz). Above this frequency, the contributions from individual array elements no longer combine into one dominant plane wave propagating towards the bright zone. Instead, multiple plane wave components are generated in the directions of the aliased lobes, causing the loss of sound field planarity in the zone.

The planarity score obtained by the PM method is generally lower than those for DS or ACC Line, with the exception of frequencies below approximately 500 Hz where a small improvement with respect to the other two methods is achieved. The planarity dips at 340 Hz and 520 Hz are less pronounced, which can be attributed to the fact that here the direct (plane) wave is directed along the side wall rather than towards it as it was the case for DS and ACC Line, thus reducing the amount of sound energy reflected into the zone. A relatively low planarity score outside the lowest frequency range can be related to the previously discussed limitations of the circular array.

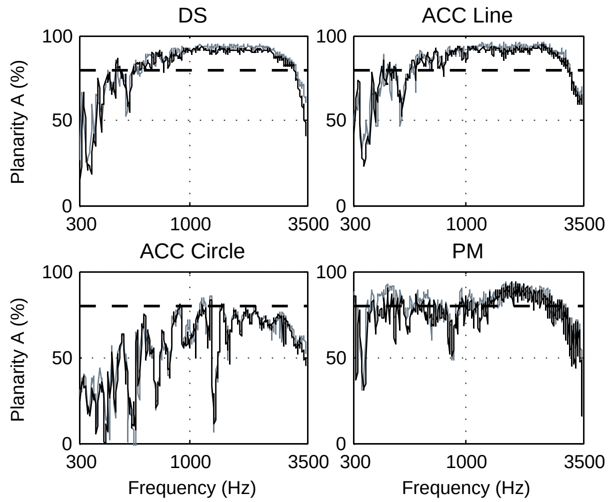


Fig. 7: Bright zone planarity predicted (light) and measured (dark) for all four cases: DS (top-left), ACC Line (top-right), ACC Circle (bottom-left) and PM (bottom-right). The dashed line indicates a reference 80% planarity.

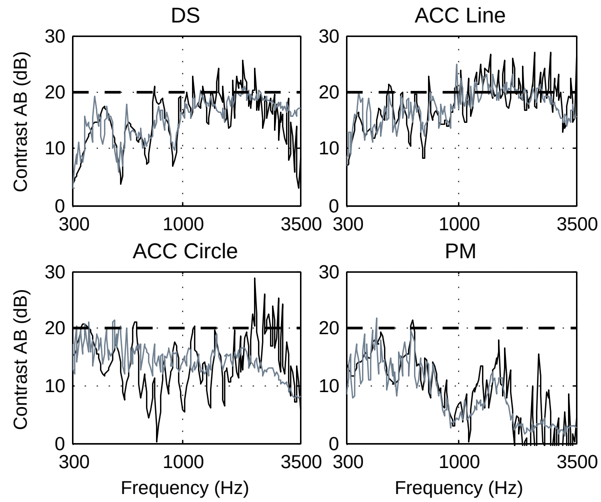


Fig. 8: Acoustic contrast based on the single microphone responses used for auralisations (dark) for all four cases: DS (top-left), ACC Line (top-right), ACC Circle (bottom-left) and PM (bottom-right). The measured spatially-averaged contrast is plotted for reference (light). For clarity at high frequencies, contrast is shown in 200 logarithmically spaced bins from 300–3500 Hz. The dashed line indicates a reference 20 dB contrast.

7. PERCEPTUAL EVALUATION

In addition to the physical metrics reported above, it is also necessary to consider the experience of a listener in a real sound zone system. A listening test was performed to collect ratings of perceived distraction due to the presence of an interfering audio signal; this attribute was found in a previous experiment to be most useful for rating audio-on-audio interference scenarios. Distraction was defined as *"how much the alternate audio pulls your attention or distracts you from the target audio"*, on a scale from *"not at all distracting"* to *"overpowering"*. The listening test results therefore consider the 'interference' domain, or separation between programmes, rather than quality or properties of the target sound field.

7.1. Listening Test Procedure

For each method, monophonic audio stimuli were generated by convolving programme material with the system response at the centre of zone A for the target programme and at the centre of zone B for the interferer programme. This method assumes symmetry in the system in order to simulate the listener experience in one zone. It must be noted that the use of monophonic auralisations limits the generality of the results to one point in the zone; this removes any spatial characteristics related to distribution of sound energy or variation in contrast throughout the zone. However, the listening test serves as a preliminary perceptual investigation of the experience of a listener in a real sound zone system, and can provide insight into the relationship between contrast and perceived distraction. The central points in the zones were selected as they displayed similar performance to the spatially-averaged contrast over the considered frequency range, as shown in Fig. 8. To reduce the influence of measurement noise on the result, contrast between the two auralisation points was calculated as the logartihmic ratio between power spectral densities at these points, estimated using the Welch's method [21] (each response had 65535 samples at 48 kHz sampling rate and was divided into 50 ms segments with a Hamming window, using a 50% overlap). The discrepancy between the single microphone and spatially-averaged contrast was more pronounced for the ACC Circle method due to large variations in the sound energy level across the microphone locations in the zones (as described in Section 6).

The programme items were chosen to cover a range of realistic listening scenarios and constituted 2 target programmes (sports commentary and pop music) and 3

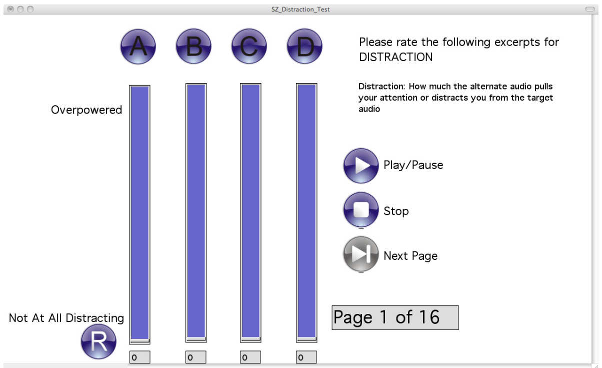


Fig. 9: Listening test interface for subjective tests, showing: buttons to play the reference (R) and test (A–D) stimuli including hidden reference; sliders to give a distraction score (from 0–100); and buttons to play/pause/stop the audio and move to the next trial.

interferer programmes (male radio speech, pop music, and classical music). Full-factorial combinations of the above factors gave a total of 24 test stimuli, each 30 seconds long.

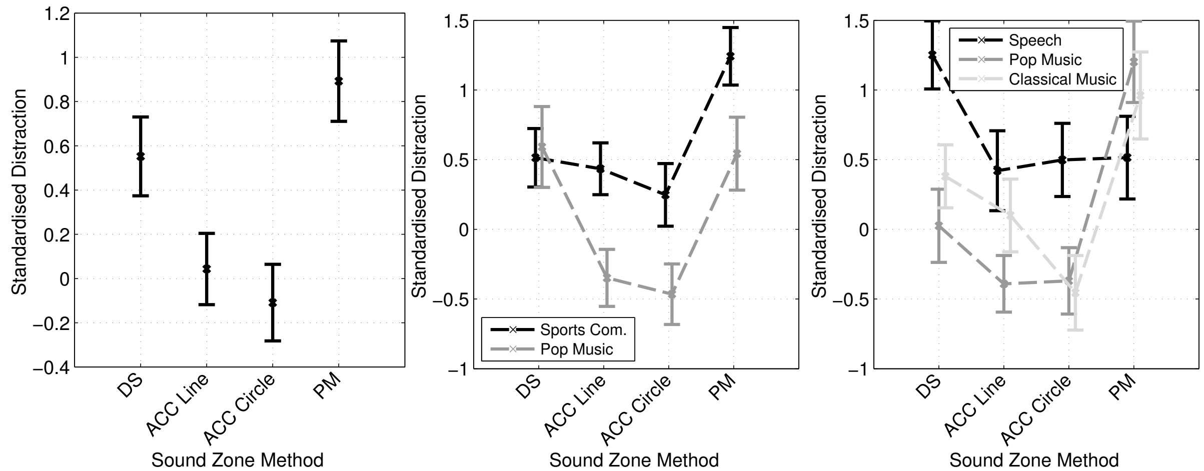
A multiple stimulus paradigm was used to collect distraction ratings using the interface shown in Figure 9. A reference stimulus (just the target audio with no interferer) was provided on each page alongside 4 test stimuli: the sound zone method and target audio programme were held constant on each page with interferer programme varied, and a hidden reference was also included to anchor the scale. Subjects were instructed to rate the hidden reference at 0. Each stimulus was presented twice giving a total of 48 ratings (64 including references) across 16 pages. The audio was presented over headphones with the same signal fed to the left and right channels. The level was calibrated using a binaural dummy head to replay at approximately 70 dB LA_{eq}. Seven subjects participated in the listening tests; all subjects were undergraduate students in Music and Sound Recording, or postgraduate students from the Institute of Sound Recording, University of Surrey, with experience of technical listening and listening test participation. The test was preceded by a familiarisation stage in which subjects were given the opportunity to audition a range of the stimuli.

7.2. Listening Test Results

The hidden reference was incorrectly identified (i.e. given a score > 0) in 17 out of 112 cases; in 16 of these cases, the control method was ACC on the line or circular array, and in 16 of the cases, the interferer was the pop music programme. This indicates that these factor levels produced similar performance to the reference (i.e. no audible interferer). The hidden reference ratings were removed in all further analysis. Observation of boxplots of subject scores indicated pronounced differences in median and range of the scores given by different subjects, therefore z-score noramlisation was performed in order to account for differences in scale use [22].

An ANOVA model was made including main effects and all interactions of target programme, interferer programme, and sound zone method as fixed factors and subject as a random factor. All effects that were not significant ($p > 0.05$) were removed and the model recalculated. The remaining significant factors were: sound zone method (partial eta-squared $\eta_p^2 = 0.290$), target programme ($\eta_p^2 = 0.152$), interferer programme ($\eta_p^2 = 0.126$), sound zone method * target programme ($\eta_p^2 = 0.073$), and sound zone method * interferer programme ($\eta_p^2 = 0.211$). The model showed a reasonable fit to the data (adjusted $r^2 = 0.496$); a Kolmogorov-Smirnov test on the standardised residuals suggested significant differences from normality ($D = 0.085, p < 0.01$), however, visual inspection of a histogram and q-q plot suggested that the ANOVA model was appropraite for the data.

The largest effect was caused by varying the sound zone method. Figure 10a shows mean standardised distraction (with 95% confidence intervals) for each method, showing the lowest distraction scores for the two ACC formulations (supporting the physical evaluation in which ACC produced the greatest contrast). However, the significant interactions between sound zone method and target/interferer programme must also be taken into account. The target programme * sound zone method interaction has a small effect size, and can be seen in Figure 10b; the ACC methods only perform better than DS for the pop music target, whilst the PM method performs less well for the speech target. The interferer programme * sound zone method interaction has a larger effect size; Figure 10c shows that the DS method produced particularly high distraction ratings for the speech interferer. It is likely that this interaction greatly contributes to the apparent difference between the ACC methods and DS. It can also be seen that for the speech interferer, the PM method did not produce greater distraction than the ACC methods. There is also a larger difference between the two ACC formulations for the classical interferer,



(a) Main effect of sound zone method (b) Sound zone method * target programme (c) Sound zone method * interferer programme

Fig. 10: Mean z-score normalised subjective distraction ratings (across subjects and repeats) with error bars showing 95% confidence intervals based on the normal distribution. Ratings are for monophonic auralisations made at a single point in the centre of the zone and so do not consider the performance of the methods across the whole zone area.

with ACC on the line array producing higher distraction scores.

Alongside the significant effect of sound zone method, the target and interferer programme also produce significantly different ratings. The pop music target produced lower distraction than the sports commentary (potentially due to the compressed and consistent nature of the pop music compared with the more fluctuating speech signal). On the other hand, the speech interferer was shown to produce higher distraction ratings than the 2 music interferers. These results highlight an important finding: the success of sound zone systems implemented in real environments with programme material is highly dependent on programme material.

8. CONCLUSIONS

A comparative study of sound zone methods implemented in a room was carried out. Three control methods were chosen: delay and sum (DS), acoustic contrast control (ACC) and pressure matching (PM), each representing a different approach to sound field control (beamforming, energy control and sound field synthesis, respectively). Two loudspeaker configurations, linear and circular, were used: the line array was used for DS, ACC was implemented on both array types, and PM was applied to the circle.

Physical attributes of the reproduced sound fields were assessed. The ACC method implemented on the line array achieved the highest overall acoustic contrast (above 15 dB in the majority of the considered frequency range) while reproducing a highly planar sound field in the bright zone and maintaining a relatively uniform level of sound energy across this zone. DS produced similar contrast, although with a decrease at low frequencies due to the limited ability to focus sound energy in this range. The circular array arrangements produced high contrast below approximately 500 Hz; lower contrast at higher frequencies was attributed to a relatively large spacing between array elements, which inhibited control over destructive interference in the dark zone. The PM technique maintained high bright zone planarity, particularly below the aliasing frequency (509 Hz for the chosen zone locations), and succeeded in reproducing a target plane wave from the specified direction. In contrast, the circular array implementation of the ACC method resulted in a low bright zone planarity, with large energy level variations across the zone.

A listening test was conducted using auralisations from the central positions in the bright and dark zones. Ratings of distraction due to the presence of an interfering audio signal were collected for combinations of 2 target and 3 interferer programmes. The results were found to generally agree with the physical metrics, with the two ACC arrangements providing the lowest distraction scores, particularly for the pop music target programme, and PM performing badly for the sports commentary programme. The programme material was found to have an overall effect with less distraction for the pop music targets, and more distraction for the speech interferer. The results highlight the importance of programme material for the success of a sound zone system, motivating the development of perceptual evaluation procedures.

9. ACKNOWLEDGEMENTS

This work forms part of the 'Perceptually Optimised Sound Zones' (POSZ) project funded by Bang & Olufsen. The authors would like to thank Adrian Celestinos, Jakob Dyreby, Kasper Thomsen, Patrick Hegarty, Morten Lydolf, Jan Abildgaard Pedersen, and other colleagues at Bang & Olufsen for discussions and help in implementing the system.

10. REFERENCES

[1] B. D. Van Veen and K. M. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5:4–24, 1988.

[2] J.-W. Choi and Y.-H. Kim. Generation of an acoustically bright zone with an illuminated region using multiple sources. *J. Acoust. Soc. Am*, 111(4):1695–1700, 2002.

[3] M. Shin, S.Q. Lee, F.M. Fazi, P.A. Nelson, D. Kim., S. Wang, K.-H. Park, and J. Seo. Maximization of acoustic energy difference between two spaces. *J. Acoust. Soc. Am.*, 128(1):121–131, 2010.

[4] Y.J. Wu and T.D. Abhayapala. Simultaneous soundfield reproduction at multiple spatial regions. In *Audio Engineering Society 128th Convention*, 2010.

[5] M. Poletti. An investigation of 2-d multizone surround sound systems. In *Audio Engineering Society 125th Convention*, 2008.

[6] 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.*, 125(4):2091–2097, 2009.

[7] S. Elliott, J. Cheer, H. Murfet, and K.R. Holland. Minimally radiating sources for personal audio. *J. Acoust. Soc. Am.*, 128(4):1721–1728, 2010.

[8] J.-Y. Park, J.-H. Chang, and Y.-H. Kim. Generation of independent bright zones for two-channel private audio system. *J. Audio Eng. Soc.*, 58(5):382–393, 2010.

[9] W.F. Druyvesteyn and J.Garas. Personal sound. *J. Audio Eng. Soc.*, 45(9):685–701, 1997.

[10] S. Elliott and M. Jones. An active headrest for personal audio. *J. Acoust. Soc. Am.*, 124(5):2702–2709, 2006.

[11] F. Jacobsen, M. Olsen, M. Møller, and F. Agerkvist. A comparison of two strategies for generating sound zones in a room. In *18th International Congress on Sound and Vibration*. International Institute of Acoustics and Vibration, 2011.

[12] M. Møller, M. Olsen, and F. Jacobsen. A hybrid method combining synthesis of a sound field and control of acoustic contrast. In *Audio Engineering Society 132nd Convention*, 2012.

[13] P. Coleman, P. Jackson, M. Olik, M. Olsen, M. Møller, and J. Pedersen. The influence of regularization on anechoic performance and robustness of sound zone methods. In *21st International Congress of Acoustics*, 2013.

[14] P. Jackson, F. Jacobsen, P. Coleman, and J. Pedersen. Sound field planarity characterized by superdirective beamforming. In *21st International Congress of Acoustics*, 2013.

[15] H. Wittek, F. Rumsey, and G. Theile. Perceptual enhancement of wavefield synthesis by stereophonic means. *J. Audio Eng. Soc.*, 55(9):723–751, 2007.

[16] S.G. Norcross, M. Bouchard, and G.A. Soulodre. Inverse filtering design using a minimal phase target function from regularization. In *Audio Engineering Society 121th Convention*, 2006.

[17] P.C. Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Review*, 34(4):561–580, 1992.

[18] M.A. Akeroyd, J. Chambers, D. Bullock, A.R. Palmer, A.Q. Summerfield, P.A. Nelson, and S. Gatehouse. The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback. *J. Acoust. Soc. Am.*, 121(2):1056–1069, 2007.

[19] F. Talantzis and D.B. Ward. Robustness of multichannel equalization in an acoustic reverberant environment. *J. Acoust. Soc. Am.*, 114(2):833–841, 2003.

[20] D.B. Ward and T. Abhayapala. Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Trans. Speech Audio Proc.*, 9(6):697–797, 2001.

[21] V.K. Madisetti. *Digital Signal Processing Fundamentals*. CRC Press, 2010.

[22] T. Næs, P. Brockhoff, and O. Tomic. *Statistics for Sensory and Consumer Science*. John Wiley & Sons, 2010.