

THE RELATIONSHIP BETWEEN TARGET QUALITY AND INTERFERENCE IN SOUND ZONES

Khan Baykaner, Philip Coleman, Russell Mason, Philip Jackson, Jon Francombe, Marek Olik, Søren Bech

December 13, 2014

Abstract

Sound zone systems aim to produce regions within a room where listeners may consume separate audio programs with minimal acoustical interference. Often, there is a trade-off between the acoustic contrast achieved between the zones, and the fidelity of the reproduced audio program (the target quality). An open question is whether reducing contrast (i.e. allowing greater interference) can improve target quality. The planarity control sound zoning method can be used to improve spatial reproduction, though at the expense of decreased contrast. Hence, this can be used to investigate the relationship between target quality (which is affected by the spatial presentation) and distraction (which is related to the perceived effect of interference). An experiment was conducted investigating target quality and distraction, and examining their relationship with overall quality within sound zones. Sound zones were reproduced using acoustic contrast control, planarity control and pressure matching applied to a circular loudspeaker array. Overall quality was related to target quality and distraction, each having a similar magnitude of effect; however, the result was dependent upon program combination. The highest mean overall quality was a compromise between distraction and target quality, with energy arriving from up to 15 degrees either side of the target direction.

1 INTRODUCTION

Sound zone systems aim to control sound fields in such a way that multiple listeners can enjoy different audio programs within the same room. Conceptually, the overall quality of the sound zone listening experience could be considered to be the result of some combination of the effect of the presence of an interferer program and the effect of any artefacts or degradations to the target program (i.e. target quality) caused by the sound zone processing. A similar conceptual framework was utilised in [1]. While the relationship between the effect of the interferer and the effect of target quality degradations is unclear, a considerable body of research exists on these topics individually.

Fields of research investigating the effect of auditory interferers include: the perception of environmental noise [2, 3], the perception of multiple talkers [4], source separation [5], and combinations of these [6]. These studies generally do not consider common domestic interferers, such as music or sound effects in films, and where they do, they either do not isolate the interferer effect or they include artefacts and degradations that may be specific to source separation algorithms.

In [7] a series of elicitation experiments were conducted to investigate terms describing auditory interference scenarios using ecologically valid programs (i.e. those that are commonly consumed in domestic environments). The results, and those of [8], showed that using the term ‘distraction’ produced good agreement between listeners, and that listener ratings made using this term were a good measure of the perceived effect of the interferer. It seems likely, therefore, that there would be some association between contrast and distraction.

The existing research investigating target quality includes objective measures of quality in telephony [9, 10], and measures of target quality for source separation algorithms [1]. However, these are not designed to address the degradations to target quality caused by reproducing programs using sound zoning systems. The types of degradations caused by sound zoning systems may include spatial degradations (due to uncontrolled phase and self-cancellation) [11], temporal degradations (such as ringing or pre-echo) [12], spectral coloration, and variation in all of these across the reproduction zone which may be audible with listener head movement.

A variety of approaches to controlling sound fields to create sound zones have been investigated [13, 14, 15, 16, 17]. Each approach enjoys differing degrees of success according to the physical measures of contrast (the acoustic separation between the zones), control effort (the energy required for sound attenuation), and planarity (the distribution of plane wave energy with respect to direction of arrival at the bright zone) [11]. Acoustic contrast control (ACC) [13] gives the maximum contrast between the zones, but does not attempt to control the phase of the resulting sound field. Least-squares optimization has therefore typically been used when control of the target field is necessary [12, 14, 18], at the cost of re-

duced acoustic contrast.

In recognition of the complex relationship between perception of contrast and reproduction error, recent work has aimed to increase the acoustic contrast between zones by allowing increased bright zone reproduction error. For instance, in [19] a weighting parameter was applied between the terms relating to the bright and dark zones, and in [20] an acoustic contrast constraint was imposed on a cost function that minimized the bright zone reproduction error. However, the target field must still be strictly specified, and any increase of the reproduction error incorporates magnitude and phase components averaged across the target zone.

The planarity control (PC) method [15] also relaxes the constraint on bright zone reproduction. Rather than allowing for increased reproduction error for a specific desired sound field, the sound energy arriving at the listener is placed (optimally for contrast) within an ‘angular pass range’. When loudspeakers surround the zones, varying the width of this pass range alters the spatial spread of sound energy impinging into the bright zone. For very wide pass ranges, PC behaves similarly to ACC and the array generates high contrast by focusing multiple energy beams in to the bright zone from various directions, at a cost of low planarity. For very narrow pass ranges, a planar sound field is reproduced at a cost of contrast. For moderately narrow pass ranges, the cancellation notches of ACC can be removed and a balance between contrast and planarity can be achieved [15].

The physical measures of contrast and planarity may be related to the listening experience within such sound zones, but they do not actually describe it in perceptual terms [17]. Moreover, the relative importance of these physical measures is unclear, making it nearly impossible to determine which sound zoning method would result in the highest overall quality of listening experience. PC offers a unique opportunity to investigate the relationship between target quality, distraction and overall quality in sound zone systems. The way in which the subjective attributes of distraction, target quality, and overall quality vary as the width of the PC target window changes is likely to give insight into the perception of sound zones as planarity and contrast are traded off against one another, and to illuminate the use of contrast and planarity as physical evaluation metrics. This leads us to address the question, ‘what is the relationship between distraction, target quality, and overall quality of listening experience in sound zones when the sound is allowed to arrive at the listener from a range of angles?’.

In Section 2 of this paper the notation and sound zone control methods are introduced. In Section 3, the experimental system and physical performance are described, and a listening test is outlined aiming to obtain subjective data describing the subjective measures of target quality, distraction and overall quality. Following this, in Section 4 the listening test results and subsequent analyses are presented. In Section 5,

the assumptions and limitations of the work are discussed, and the relationships between physical and perceptual metrics are explored. Finally conclusions are drawn and the work is summarized in Section 6.

2 BACKGROUND

In this section the sound zone system notation is introduced and the sound zone methods implemented in this study are described.

2.1 Notation

Reproduction of sound zones for two listeners requires superposition of two sets of source weights that each attempt to create a single bright zone and dark zone. The system is illustrated in Fig. 1, which shows the sound zone system notation and geometry. Zones A and B are in turn considered as the bright zone, and there are no constraints on the sound field outside of these regions. For clarity, the notation and theory considers a single set of filters creating bright zone and dark zone.

For each frequency, the optimal source weights $\mathbf{q} = [q^{(1)}, q^{(2)}, \dots, q^{(L)}]^T$ must be calculated, where there are L loudspeakers and $q^{(l)}$ is the complex source weight of the l th loudspeaker. The complex pressures at the control microphone positions in zones A and B are $\mathbf{p}_A = [p_A^{(1)}, p_A^{(2)}, \dots, p_A^{(N_A)}]^T$ and $\mathbf{p}_B = [p_B^{(1)}, p_B^{(2)}, \dots, p_B^{(N_B)}]^T$ respectively, where there are N_A control microphones in zone A and N_B in zone B, and $p_A^{(n)}$ and $p_B^{(n)}$ are the complex pressures at the n th microphones in each zone. The observed pressures at the monitor microphones in each zone are denoted as $\mathbf{o}_A = [o_A^{(1)}, o_A^{(2)}, \dots, o_A^{(M_A)}]^T$ and $\mathbf{o}_B = [o_B^{(1)}, o_B^{(2)}, \dots, o_B^{(M_B)}]^T$ respectively, where there are M_A monitor microphones in zone A and M_B in zone B, and the complex pressures at the m th microphones in each zone are $o_A^{(m)}$ and $o_B^{(m)}$. Microphones sample the zones in a uniform grid and are assigned alternately as control or monitor positions to reduce any bias arising from performance evaluation only at the control positions. The pressure vectors are related to the source weights by the summation of the contribution of the source weights at each microphone, written in vector form as $\mathbf{p}_A = \mathbf{G}_A \mathbf{q}$, $\mathbf{o}_A = \mathbf{\Omega}_A \mathbf{q}$, $\mathbf{p}_B = \mathbf{G}_B \mathbf{q}$ and $\mathbf{o}_B = \mathbf{\Omega}_B \mathbf{q}$, where \mathbf{G}_A and $\mathbf{\Omega}_A$ are the control and monitor microphone transfer function matrices, respectively, with respect to zone A, and \mathbf{G}_B and $\mathbf{\Omega}_B$ are the transfer function matrices with respect to zone B.

2.2 Acoustic contrast control

ACC [13] maximizes the contrast between the spatially averaged pressures in the target (bright) zone and the interferer (dark) zone. The cost function may be written to minimize

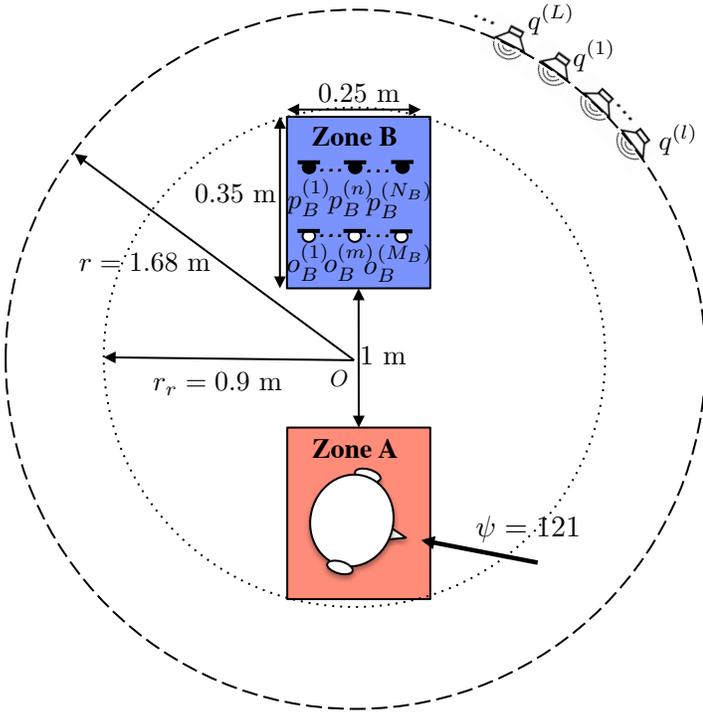


Figure 1: System layout with L loudspeakers, zone B comprising N_B control microphones (black) and M_B monitor microphones (white), and zone A containing a listener and showing the reference plane wave direction of $\psi = 121^\circ$. Control and monitor microphones are evenly distributed across both zones for the experiments.

the dark zone sound pressure while maintaining a certain sound pressure A in the bright zone, with the sum of squared source weights not exceeding Q [16]:

$$J = \mathbf{p}_B^H \mathbf{p}_B + \mu (\mathbf{p}_A^H \mathbf{p}_A - A) + \lambda (\mathbf{q}^H \mathbf{q} - Q), \quad (1)$$

where H denotes the Hermitian transpose, and μ and λ are Lagrange multipliers.

The cost function may be minimized by setting the derivatives with respect to \mathbf{q} , μ and λ to zero,

$$\begin{aligned} -(\mathbf{G}_A^H \mathbf{G}_A)^{-1} (\mathbf{G}_B^H \mathbf{G}_B + \lambda \mathbf{I}) \mathbf{q} &= \mu \mathbf{q}; \\ \mathbf{p}_A^H \mathbf{p}_A &= A; \quad \mathbf{q}^H \mathbf{q} = Q, \end{aligned} \quad (2)$$

where \mathbf{I} is the identity matrix and \mathbf{q} is proportional to the eigenvector $\hat{\mathbf{q}}$ corresponding to the maximum eigenvalue of $(\mathbf{G}_B^H \mathbf{G}_B + \lambda \mathbf{I})^{-1} (\mathbf{G}_A^H \mathbf{G}_A)$ [16]. The constraint that A equals a certain fixed value is enforced by scaling $\hat{\mathbf{q}}$, and the second Lagrange multiplier λ (which also acts as a regularization parameter for the matrix inversion) must be chosen such that the effort constraint is satisfied.

For this implementation, λ was first initialized such that the condition number of $(\mathbf{G}_B^H \mathbf{G}_B + \lambda \mathbf{I})$ did not exceed 10^{10} [11]. At higher frequencies where the condition number was already below this threshold, λ was instead initialized such that the condition number of $(\mathbf{G}_B^H \mathbf{G}_B + \lambda \mathbf{I})$ was reduced by a

factor of 10. This approach improved the robustness to errors and reduced ringing artefacts in the filter responses. Finally, a gradient descent search was used to increase λ such that $Q \geq \mathbf{q}^H \mathbf{q}$ when A has been fixed.

Since ACC purely maximizes the ratio of spatially averaged squared pressures between the zones, it tends to outperform other methods in terms of contrast [11]. However, ACC does not control the phase and so may result in confusing spatial cues for the listener. As a result, it may be expected that ACC would produce listening scenarios with a lower distraction score, but with a poorer target quality score than other sound zoning methods.

2.3 Pressure matching

Pressure matching (PM) minimizes the error in a least-squares sense between the desired and reproduced sound fields across both zones. A plane wave sound field can be written as $\mathbf{d}_A = D_A e^{jk\mathbf{r}_n \cdot \mathbf{u}_\varphi}$, for $n = 1, 2, \dots, N_A$, where D_A gives the pressure amplitude, \mathbf{r}_n is the position of the n th control microphone in zone A, \cdot denotes the inner product, and \mathbf{u}_φ is the unit vector in the direction of the incoming plane wave. The desired field for dark zone B is given by a vector of length N_B populated with zeros, $\mathbf{d}_B = \mathbf{0}$. The cost function, with a constraint on the sum of squared source weights Q , is [14]:

$$J = \mathbf{p}_B^H \mathbf{p}_B + (\mathbf{p}_A - \mathbf{d}_A)^H (\mathbf{p}_A - \mathbf{d}_A) + \lambda (\mathbf{q}^H \mathbf{q} - Q). \quad (3)$$

Using the method of Lagrange multipliers the solution can be found by taking the derivatives with respect to \mathbf{q} and λ :

$$\mathbf{q} = (\mathbf{G}_A^H \mathbf{G}_A + \mathbf{G}_B^H \mathbf{G}_B + \lambda \mathbf{I})^{-1} \mathbf{G}_A^H \mathbf{d}_A; \quad \mathbf{q}^H \mathbf{q} = Q. \quad (4)$$

The Lagrange multiplier λ is initialized as above and numerically chosen to satisfy the control effort constraint. It is assumed that the solution is appropriately scaled by setting $\mathbf{d}_A^H \mathbf{d}_A = A$.

As PM minimizes the error of the complex pressures in the reproduced sound field, the confusing spatial cues present in ACC implementations are avoided by specifying a suitable target field (typically a plane wave when applied on a circular array, due to the potential for superposition of solutions to create an arbitrary target scene). The strict target field does however result in poorer contrast than ACC [11], particularly at frequencies above the array aliasing limit. As a result, we might expect that PM would produce listening scenarios with a higher distraction score than ACC, but also with improved target quality.

2.4 Planarity control

PC [15] aims to avoid the self-cancellation artefacts of ACC, while allowing improved contrast with respect to PM by relaxing the requirement for reproduction of a specific sound

field. PC works by introducing a spatial filtering component to the ACC sound zone optimization. The cost function minimizes the dark zone pressures (as ACC) with the bright zone energy constraint enforced via a spatial domain (similar to [21]), and with an effort constraint:

$$J = \mathbf{p}_B^H \mathbf{p}_B + \mu (\mathbf{p}_A^H \mathbf{Y}_A^H \mathbf{\Gamma} \mathbf{Y}_A \mathbf{p}_A - A) + \lambda (\mathbf{q}^H \mathbf{q} - Q). \quad (5)$$

The steering matrix \mathbf{Y}_A of dimensions $I \times N_A$, with I steering angles, maps between the observed pressures at the microphones and the plane wave components and is populated by superdirective beamforming (as in [15]). The term $\mathbf{\Gamma}$ is a diagonal matrix allowing a weighting to be applied to the angular spectrum based on the desired incoming plane wave directions:

$$\mathbf{\Gamma} = \text{diag}[\gamma_1, \gamma_2, \dots, \gamma_I], \quad (6)$$

where $0 \leq \gamma_i \leq 1$ is the weighting corresponding to the i th steering angle. Energy will therefore be focused in the direction of the nonzero elements of $\mathbf{\Gamma}$.

The solution is found, as for ACC above, by setting to zero the derivatives with respect to \mathbf{q} and each of the Lagrange multipliers, and the optimal source weights are proportional to the eigenvector corresponding to the maximum eigenvalue of $(\mathbf{G}_B^H \mathbf{G}_B + \lambda \mathbf{I})^{-1} (\mathbf{G}_A^H \mathbf{Y}_A^H \mathbf{\Gamma} \mathbf{Y}_A \mathbf{G}_A)$. The values of the Lagrange multipliers are determined iteratively as above, where the sum of squared pressures (projected via the angular spectrum) is fixed to satisfy the constraint $A = \mathbf{p}_A^H \mathbf{Y}_A^H \mathbf{\Gamma} \mathbf{Y}_A \mathbf{p}_A$, and λ is initialized based on the matrix condition number and chosen such that the constraint on $\mathbf{q}^H \mathbf{q}$ is satisfied.

The design of the angular pass range $\mathbf{\Gamma}$, with weightings γ between zero and one, is a significant factor in PC implementation, and is exploited in this article. If the diagonal is filled with ones, then PC is identical to ACC (Eq. (1)), and energy may impinge on the target zone from any direction. If, on the other hand, the vector is populated with zeros apart from a single target direction, a plane wave impinging from that direction should be reproduced, acting in a similar manner to the wavenumber domain point focusing method of [21] (while maintaining the dark zone).

Designs of $\mathbf{\Gamma}$ between these two extremes can balance the freedom of the array to focus the sound from a certain direction against the freedom to create maximal acoustic contrast between the zones. By auditioning various widths of angular pass range, PC can be used to investigate the relative importance of distraction (which is related to contrast) and target quality (specifically, for degradations caused by self-cancellation or spatial spreading of the signal). With the results of such an investigation, it will be possible to infer the relationship between these quantities and their relationship to the overall quality of the listening scenario.

3 EXPERIMENT DESIGN

This section describes a listening test conducted to obtain subjective measures of distraction, target quality, and overall quality of listening experience for ecologically valid programs within a sound zoning system. The sound zoning reproduction system and physical performance is first described, before details of the listening test methodology are discussed.

3.1 Reproduction system realization

A reproduction and measurement system was designed and mounted on a bespoke spherical structure, the ‘‘Surrey Sound Sphere’’, placed in an acoustically treated room of dimensions $6.93 \times 7.81 \times 3.98$ m (RT60 217 ms averaged over 0.5 kHz, 1 kHz and 2 kHz octave bands). Loudspeakers (Genelec 8020b) were clamped to the equator of the sphere to form a 60 channel circular array (radius 1.68 m, as Fig. 1), and 48 microphones (Countryman B3 omni) were arranged as a 6-by-8 grid with 5 cm spacing. Four positions of the microphone stand were measured per zone to achieve a 25 cm \times 35 cm uniform grid of sampling points with 2.5 cm spacing. A photograph of the equipment is shown in Fig. 2. A computer running Matlab was used to play and record the signals via the playrec utility [22]. A 72 channel MOTU PCIe 424 sound card was used for the analogue to digital conversion, with the microphone inputs first passed through a pre-amplifier stage (PreSonus Digimax D8). Level differences between the input and output signal channels were compensated through calibration. Room impulse responses (RIRs) between each loudspeaker and each microphone position were measured using the maximum length sequence (MLS) approach (15th order). The RIRs for setup were cropped at 27 ms—determined in a pilot experiment to provide a good balance between contrast and sound quality—to ensure that the system did not attempt to compensate for reverberation beyond the first reflections.

Finite impulse response (FIR) filters were populated in the frequency domain based on source weights calculated at individual frequencies. The RIRs were first down-sampled to the sample rate of 16 kHz used to calculate the filters, and an 8192 point fast Fourier transform (FFT) was taken. Solutions were calculated up to the Nyquist frequency of 8 kHz regardless of the spatial aliasing effect due to the loudspeaker array. The source weights were collated across frequency, the negative frequency bins populated by complex conjugation, and the inverse FFT taken to obtain a time-domain filter. A 4096 sample modelling delay was applied to ensure causality. For the listening tests, the program material was convolved with the filter for each loudspeaker. Measurements of objective performance were made by convolving an MLS with each of the FIR control filters, simultaneously replaying them through all loudspeakers and sampling the reproduced

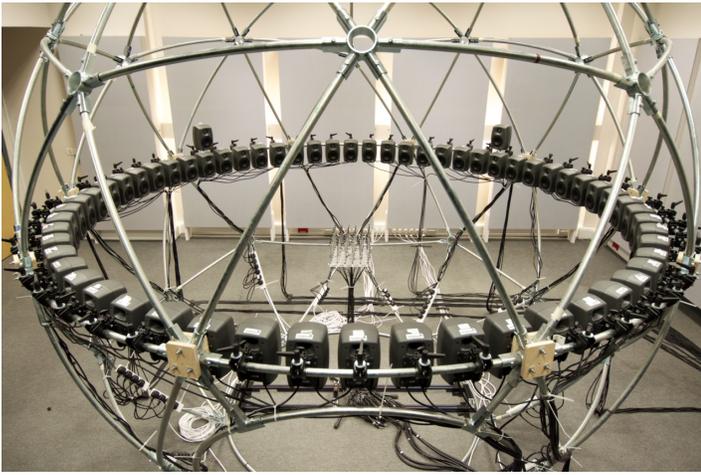


Figure 2: Photograph of the reproduction system showing the 60 channel circular loudspeaker array and microphone grid (centre).

sound pressures with the microphone array.

3.2 Physical performance

Measurements of contrast C and planarity η were made inside the sound zoning system to facilitate a comparison between the physical and perceptual metrics. These metrics are defined as [11]:

$$C = 10 \log_{10} \left(\frac{M_B \mathbf{o}_A^H \mathbf{o}_A}{M_A \mathbf{o}_B^H \mathbf{o}_B} \right); \quad \eta = \frac{\sum_i w_i \mathbf{u}_i \cdot \mathbf{u}_\alpha}{\sum_i w_i}, \quad (7)$$

where \mathbf{u}_i is the unit vector associated with the i th component's direction, \mathbf{u}_α is the unit vector in the principal direction $\alpha = \arg \max_i w_i$, the energy components w_i at each angle are elements of $\mathbf{w} = [w_1, w_2, \dots, w_I]^T = \frac{1}{2} |\mathbf{H}_A \mathbf{o}_A|^2$, and \cdot denotes the inner product. The steering matrix \mathbf{H}_A ($I \times M_A$) is populated by superdirective beamforming, as \mathbf{Y}_A but based on the monitor microphone positions.

The measured results, averaged across the frequency range 112–7079 Hz (i.e. one-third octave bands centered at 125–6300 Hz), are shown in Table 1. Although these results incorporate the effects of spatial aliasing, the overall trends expected among the methods are evident in the performance. In particular, ACC has the highest contrast and lowest planarity, and PM has the lowest contrast and highest planarity. Under each metric, the family of PC results fall between the ACC and PM values. There is little variation in contrast between the PC270–PC30 implementations, although there is a slight drop for PC0. On the other hand, there is a general trend for increasing planarity as the pass range is narrowed. The slight drop in PC0 planarity compared to PC30 is due to more significant aliasing lobes outside of the pass range.

Considering the physical results, one would expect the distraction scores among the PC methods to be similar, with

ACC the least distracting and PM the most distracting. The target quality scores would be expected to steadily increase as the angular pass range is tightened, with the plane wave reproduction of PM the highest quality.

3.3 Listening test design

Three multiple stimulus style listening tests, based on [23], were carried out within the sound zoning system to investigate distraction, target quality, and overall quality respectively. Each test featured the same set of stimuli but the page and order of stimuli was randomised for each test and each subject. Each page contained a known reference and nine test stimuli, including the hidden reference, with the remaining 8 stimuli produced using ACC, PM and 6 versions of PC. The 6 versions of PC were constructed each using a diagonal $\mathbf{\Gamma}$ which limited the target windows to 270° , 180° , 90° , 60° , 30° , and 0° (i.e. a single direction specified). These window widths were selected in order to cover as much of the range of distraction and target quality as possible; a pilot experiment was conducted and the consensus of the 3 listeners was used to determine the target windows to be included.

The reference and hidden reference signals consisted of the target program (without any interferer) replayed through a single loudspeaker. Subjects were asked to rate at least one stimulus per page at 100 (except for the distraction test for which the scale is reversed and subjects were required to rate at least one stimulus per page at 0). A target presentation level of 72 dB SPL, verified by taking measurements using a sound pressure level meter of noise replayed via each sound zone process, was used to equalize the level among the control methods. All stimuli were loudness matched against one another.

The listener was positioned in zone A, orientated towards the reference direction of 121° (Fig. 1). This direction corresponded to the installed loudspeaker closest to the angle of 115° which was found to be optimal for PC reproduction at 1 kHz in anechoic simulations of the reproduction system [15]). A single loudspeaker positioned at 121° with respect to the listener was used to replay the reference stimuli. The PC methods had pass ranges centred on 121° with respect to zone A, and the plane wave for PM was specified with $\varphi = 121^\circ$ as the angle of incidence. The difference between the target program location and the installed reference loudspeaker location was $\sim 0.3^\circ$, which is substantially less than the minimum audible angle [24]. The target direction of the zone B filters (i.e. the interferer as heard by the subjects) was designed to be symmetric to that of zone A about the axis equally dividing the zones, so that in principle the contrast between the zones was equivalent in both cases.

Program items demonstrating a range of spectro-temporal characteristics were used: pop/dance target with soft-pop interferer, classical target with pop interferer, and sports com-

	ACC	PC270	PC180	PC90	PC60	PC30	PC0	PM
Contrast (dB)	16.6	15.6	15.4	15.4	15.3	15.2	14.8	11.9
Planarity (%)	27.8	33.1	45.4	50.6	51.6	53.2	51.6	71.8

Table 1: Measured acoustic contrast and planarity scores for each method, averaged in one-third octave bands centered at 125–6300 Hz.

mentary target with pop interferer. Each test therefore had 3 pages (one per program combination). All programs were band pass filtered within the range 125 Hz to 6.3 kHz due to the limitations of the sound zone reproduction methods. It proved prohibitively difficult to find a common low anchor stimulus for all three rating scales and all three pairs of program material, hence a low anchor was not included in the experiment. Instead, a familiarisation page consisting of all stimuli was included at the start of the experiment to give listeners an impression of the overall scale range.

Subjects were directed to sit on a chair facing the angle of the reference loudspeaker, and were provided with a laptop computer which allowed them to interact remotely with a bespoke user interface modified from MUSHRAM [25]. The interface differed from MUSHRAM in that instead of using the Wavplay function, the playrec [22] function was utilised, along with a custom built buffering stage which allowed for relatively quick changes (< 100 ms) in stimulus playback. Using this interface, subjects were asked to make ratings of target quality, distraction, and overall quality. Instructions were given to the subjects which included descriptions of these quantities as follows:

‘Target quality is concerned with any and all degradations in the target program (relative to the reference). These could include degradations in spatial image, or in spectral or temporal aspects of the sound. Target quality is not concerned with how distracting you find the presence of the interferer program. Scores range from 100 (best target quality) to 0 (worst target quality).’

Distraction describes how much the alternate audio pulls your attention or distracts you from the target audio. Scores can range from 100 (overpowered) to 0 (not at all distracting).

In the overall quality of listening experience part of the test, please rate the overall quality of the listening experience including any and all aspects of the sound you considered to be important to making this judgement. Scores range from 100 (best overall quality) to 0 (worst overall quality).’

Ten subjects (8 male and 2 female) aged 21–38 reporting no hearing difficulties completed the listening tests for all three rating scales (target quality, distraction, and overall quality) on two occasions. This resulted in a total of 10 subjects \times 3 metrics \times 2 repeats \times 3 program combinations \times 8 stimuli per page, giving 1440 data points in total (excluding hidden references). Each test (of the three) required approximately 10 minutes to complete, giving a session time of approximately

30 minutes and a total time of one hour.

4 RESULTS AND ANALYSIS

In this section the results of the listening test are presented and analysed to investigate the relationship between target quality, distraction, and overall quality for the various sound zoning methods and program combinations.

4.1 Subject performance and consistency

In all cases, all subjects identified the hidden references correctly (i.e. they were rated at 100 for target quality and overall quality, and 0 for distraction). The hidden reference scores were therefore excluded from further analysis.

All absolute differences between repeats were calculated for each subject and for each rating scale. Target quality scores had slightly higher mean absolute repeat errors ranging from 8.8–30.2, with distraction and overall quality mean absolute repeat errors generally slightly lower ranging from 4.8–19.9 and from 4.9–21.4 respectively. The histograms for the within subject data were all negatively skewed (indicating that the mean absolute error will tend to overestimate the differences between repeats).

Based on these data, the subjects were assumed to be performing the task correctly.

4.2 Overview of mean scores

Figure 3 shows the scores for target quality, distraction, and overall quality for each sound zoning method averaged across subjects, repeats, and program combinations.

The scores show that as the PC window narrows the target quality increases steadily up to PC30, after which target quality remains approximately constant. By contrast, the distraction scores show approximately the reverse trend, but with very small differences between distraction scores for different widths of target window for PC. Since the target quality did not continue to improve beyond PC30, yet the distraction scores continued to increase, this seems to have resulted in the highest mean overall quality score at PC30.

4.3 ANOVAs

Shapiro-Wilk tests [26] were conducted to test for normality in the data grouped according to rating scale, program combi-

	dF	F	Sig.	Partial η^2
Intercept	1	739.332	<0.001	0.975
Program	2	7.246	0.006	0.475
SZ Method	7	23.288	<0.001	0.744
Prog * Sub	16	4.949	<0.001	0.188
SZ meth * Sub	56	1.509	0.015	0.198

Table 2: Statistics for the significant main effects in the ANOVA model for target quality.

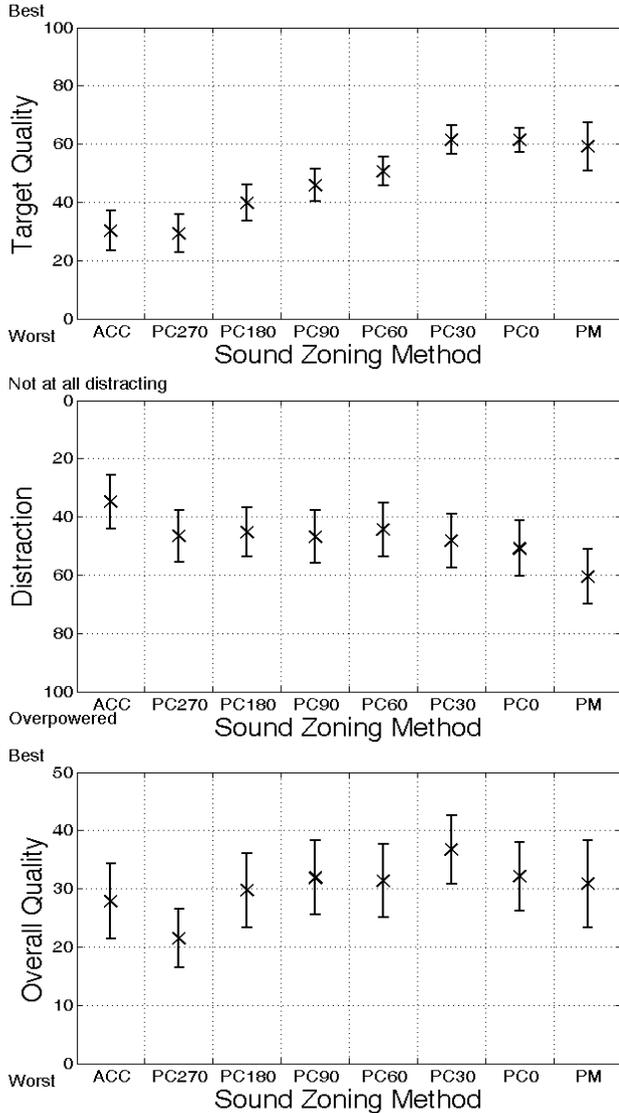


Figure 3: Target quality, distraction, and overall quality scores averaged across repeats, subjects and programs with 95% confidence intervals calculated with $n = 30$ based on 10 subjects and 3 program combinations. The scale is reversed for distraction scores so that for all plots a higher marker indicates a better score.

nation, and sound zoning method. For overall quality, only 3 of the 24 cases were not normally distributed, for distraction 10 of 24 cases were not normally distributed, and for target quality 4 of 24 cases were not normally distributed. An inspection of the histograms showed no strong indications of multimodal distributions, however, and based on the expectation that a violation of the normality assumption is likely to have only a small effect for parametric tests using $\alpha > 0.001$ [28], analyses of variance (ANOVAs) were conducted and are discussed in the following sub-sections.

4.3.1 Target quality

The target quality ANOVA (see Table 2) shows that (except for the intercept) the sound zoning method had the largest effect. The program had a moderate effect size and the two-way interactions between program and subject, and between sound zoning method and subject, had smaller significant effects.

In order to interpret significant effects for the sound zoning methods a post hoc Tukey's honest significant difference (HSD) test [27] was carried out; these results are shown in Table 3. The homogeneous subsets (groups) indicate which sound zoning methods did not have significantly different mean target quality scores; so for example although the PC270 method had the lowest mean target quality, it was not statistically significantly different from the mean target quality for the ACC method (and so both methods form group 1). The results show a general trend where as the target window narrows, the target quality improves. The highest scoring group included the PM, PC0, and PC30 sound zoning methods, and of these only the PM method could not be distinguished from PC60 (the method with the next highest mean target quality). This relationship between target window width and target quality follows from the control method design, as wider target windows allow sound energy to arrive at the zone from many directions, leading to lower spatial quality relative to the reference case (represented by a single loudspeaker).

SZ Method	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
PC270	30.70					
ACC	32.00	32.00				
PC180		40.39	40.39			
PC90			45.13	45.13		
PC60				49.69	49.69	
PM					58.24	58.24
PC0						61.07
PC30						62.11
Sig.	1.000	0.096	0.757	0.793	0.083	0.900

Table 3: Homogeneous subsets based on a Tukey HSD post hoc test of the sound zoning methods for target quality. ACC represents the acoustic contrast control sound zoning method, PM represents the pressure matching sound zoning method, and PC followed by a number represents the planarity control sound zoning method with a specified angular pass range.

	dF	F	Sig.	Partial η^2
Intercept	1	64.422	<0.001	0.877
Program	2	18.391	<0.001	0.671
SZ Method	7	15.997	<0.001	0.640
Subject	9	7.406	<0.001	0.776
Prog * Sub	18	21.009	<0.001	0.499
SZ meth * Sub	63	1.742	0.001	0.224

Table 4: Statistics for the significant main effects in the ANOVA model for distraction.

4.3.2 Distraction

Table 4 shows the ANOVA for distraction scores. As with the target quality scores, the two-way interactions featuring subject were significant with small to moderate effect sizes. The program combination and sound zoning method main effects had similar and large significant main effects.

Since the sound zone differences were of primary interest, a post hoc Tukey’s HSD test was conducted on these (see Table 5). The results show that the PM method produced significantly more distracting sound zones than all other methods, and that the ACC method produced significantly less distracting sound zones than all other methods. PC60 was also found to be significantly less distracting than PC0; the reason for this distinction is unclear, however this effect was very small (6 points). It was expected that the relative loudness of the target and interferer programs would play an important role in distraction. It is therefore plausible that ACC resulted in the lowest distraction because it optimises for contrast only, whereas PM has the fewest degrees of freedom for cancellation among the methods, resulting in the poorest contrast and therefore the highest distraction. These methods were distinct from the family of PC methods, which provide a greater spatial constraint in the algorithm design compared to ACC but still do not strictly control the phase, compared to PM. The distraction scores correspond well to the expected results outlined in Section 3.2.

SZ Meth	Group 1	Group 2	Group 3	Group 4
ACC	35.33			
PC60		44.93		
PC180		45.62	45.62	
PC270		47.07	47.07	
PC90		47.35	47.35	
PC30		48.62	48.62	
PC0			50.93	
PM				60.48
Sig.	1.000	0.502	0.087	1.000

Table 5: Homogeneous subsets based on a Tukey HSD post hoc test of the sound zoning methods for distraction. ACC represents the acoustic contrast control sound zoning method, PM represents the pressure matching sound zoning method, and PC followed by a number represents the planarity control sound zoning method with a specified angular pass range.

4.3.3 Overall quality

Table 6 shows the ANOVA for overall quality. As with distraction, for the overall quality both interactions including the subject were significant with small to moderate effect size, and the program combinations and sound zoning method had similar effect sizes, with subject and intercept having the largest effect size. Overall quality differs, however, in that the interaction between program combination and sound zoning method was significant, and the effect size was similar to that of the main effects.

In this case, where the main effects under consideration also have significant interactions with reasonable effect sizes, it is inappropriate to rely on a post hoc test of the sound zoning method. Figure 4 shows the two-way interaction between program combination and sound zoning method. As can be seen, the pop target and sports commentary targets have similar trends, with PC30, PC0 and PM producing the highest mean overall quality scores, and scores decreasing as the target window is widened. Conversely, for the classical target

	dF	F	Sig.	Partial η^2
Intercept	1	62.116	<0.001	0.873
Program	2	3.954	0.038	0.305
SZ Method	7	4.551	<0.001	0.336
Subject	9	6.413	<0.001	0.715
Prog * Sub	18	8.638	<0.001	0.298
SZ meth * Sub	63	2.218	<0.001	0.276
Prog * SZ meth	14	7.852	<0.001	0.231

Table 6: Statistics for the significant main effects in the ANOVA model for overall quality.

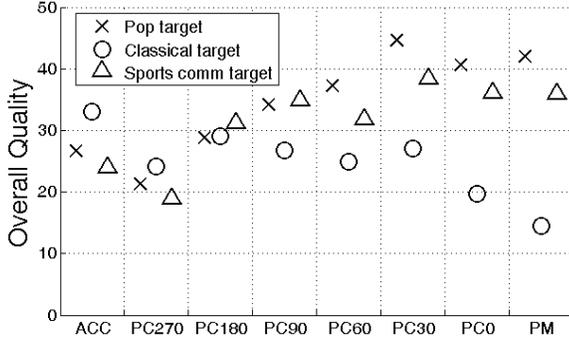


Figure 4: Mean overall quality scores separated by sound zoning method and program combination.

the ACC method produces the highest mean overall quality scores, with the PC270, PC180, PC90, PC60, and PC30 methods producing slightly lower scores, and with PC0 and PM producing the lowest scores. The trend for the pop and sports commentary data is likely to be due to these being relatively robust to interference, and as a result, target quality was a higher priority for listeners. Conversely, classical music was not very robust to interference, so for this the listeners prioritized a higher contrast.

Subjects reported finding it fairly difficult to rate overall quality, noting that it can be difficult to decide how to aggregate multiple aspects of the listening experience into a single value. As a result the confidence intervals are fairly wide. Despite this, a general trend for the interaction between sound zoning method and program combination is still apparent.

It was noted in Section 4.2 that the PC30 had the highest mean overall quality scores. The interaction between overall quality and program combination explains this result: since the pop and sports commentary target were particularly robust to interference, the benefits of improved contrast offered by ACC were relatively less important than the improved target quality offered by the narrower PC and PM methods. Conversely, the classical music was not particularly robust to interference, and so the ACC performed best, yet the PC30 had no disadvantage in target quality (relative to PM) while

	Target qual.	Distraction	Overall qual.
Contrast	-0.54	-0.42	-0.12
Planarity	0.74	0.40	0.32

Table 7: Correlation coefficients between physical metrics (averaged in one-third octave bands centered at 125 Hz to 6.3 kHz), and subjective scores (averaged across subjects).

maintaining some of the benefit of the improved contrast of ACC.

5 EXPERIMENTAL FINDINGS

In this section, assumptions and limitations of the work are first stated. Then, the relationship between the physical and perceptual evaluation is considered and the relationships among the perceptual metrics are discussed.

5.1 Assumptions and limitations

It is worth briefly considering the limitations to the scope of this work, and the assumptions upon which the conclusions depend.

Firstly, although the authors expect that the work presented here gives a good indication of target quality, distraction, and overall quality for sound zoning in general, this work is necessarily limited to the SZ methods tested. Subjective scores for radically different sound zoning methods may not necessarily conform to the conclusions of this work; this is particularly likely for any sound zoning method which tends to produced characteristically different target quality degradations.

Secondly, it should be noted that this work investigated subjective attributes of the listening experience by varying the width of the PC angular pass range; yet the optimal width may vary across frequency. An investigation into this relationship is beyond the scope of this work, however it is important to note that there may be variations in target quality scores across program combinations which are caused by the differences in spectra of programs.

Finally it is worth noting that the subjective measure ‘overall quality’ is not identical to ‘preference’, and so the sound zoning method with the highest quality of listening experience may not always be the preferred listening experience, but should be the scenario which listeners find most closely corresponds with the reference case.

5.2 Relationships amongst physical and subjective measures

Table 7 shows the correlation coefficients between the mean planarity scores and mean contrast (averaged from 125 Hz to 6.3 kHz) and the three subjective measures ($n = 24$, averaged

across subjects). As expected, a strong positive correlation ($R = 0.74$) was found between the mean planarity and target quality scores, and a negative correlation ($R = -0.42$) was found between the mean contrast and distraction scores. This seems to indicate that planarity plays an important role in the perception of target quality, and contrast plays an important role in distraction.

Conclusions based on these correlations should be interpreted cautiously, however, since the positive correlation ($R = 0.40$) found between planarity and distraction, and the negative correlation ($R = -0.54$) found between contrast and target quality seem spurious. It is likely that these correlations are the result of covariation effects caused by the experiment design; specifically, in this experiment planarity was traded off against contrast, resulting in a strong negative correlation between the physical metrics ($R = -0.89$).

The correlations between both physical measures and the overall quality were fairly low ($R \leq 0.32$). As described in section Section 4.3.3, the overall quality was dependent on both the sound zoning system employed and the program item combination. Hence, as the physical measures only represent differences caused by the sound zoning system and are not dependent on the program items, there was no strong correlation found between these and the overall quality.

5.3 Relationships amongst target quality, distraction, and overall quality

A linear regression model was constructed to investigate the relationship between target quality and distraction, and overall quality. Since this regression model is based on a sample size of only 24 (8 sound zoning methods \times 3 program combinations), it should be considered indicative rather than definitive; nonetheless, the model should give a reasonable indication of the relative importance of these subjective attributes for these sound zoning methods and programme items.

A linear regression to overall quality was calculated using the Matlab regress function; using z standardised target quality and distraction data, the resulting model was:

$$Q = 4.86T - 6.16D + 30.24 \quad (8)$$

where Q represents overall quality, T represents target quality, and D represent distraction. The model had a fit of $R = 0.69$ and all factors were significant with $p = 0.0033$ for target quality, $p = 0.0004$ for distraction, and $p < 0.0001$ for the constant term.

The coefficients indicate that the distraction and target quality were of approximately equal importance to the overall quality. Since the correlation between target quality and distraction was $R = 0.55$, the approximately equal coefficient sizes cannot be explained by target quality and distraction being precisely equal and opposite across sound zoning conditions.

Another regression model was calculated, this time including the interaction term; the model was:

$$Q = 16.41T + 8.02D - 23.06TD + 30.24. \quad (9)$$

The model had a fit of $R = 0.88$. For this model the constant term was not significant ($p = 0.42$), however the target quality, distraction, and interaction terms were all significant ($p < 0.0001$, $p = 0.0142$, and $p = 0.0001$ respectively).

The interaction term had the coefficient with the largest value, however all coefficients were within an order of magnitude, indicating that all terms were of similar levels of importance for overall quality.

6 SUMMARY AND CONCLUSIONS

An investigation into the relationship between target quality, distraction, and overall quality of listening scenarios in sound zones was conducted, with a specific focus on the effect of constraining the width of the angle from which target program energy was reproduced. A listening test was carried out to gather the subjective data for programs processed using PC with a range of window widths as well as ACC and PM.

The results indicated that as a general rule, as the width of the pass band was more tightly constrained the target quality scores increased, whereas the distraction scores remained approximately constant across most widths with the exception of the extreme cases, ACC and PC0. PM produced sound zones with target quality matching that of the narrowest two methods tested (PC30 and PC0), and had higher distraction scores than all other methods tested. For overall quality, PC30 had the best average scores (although the confidence intervals overlapped), as it offered a good compromise between high target quality and reasonably low distraction. As a result, the PC30 method seemed to be most robust to the differing priorities for different program combinations. For program combinations which were more robust to interference, such as the pop and sports commentary targets, the sound zoning methods producing better average target quality (PC30, PC0, and PM) resulted in the highest overall quality scores. For the classical target with pop interferer, however, the interference was prominent and minimizing the distraction seemed to be more important than target quality.

A regression model was constructed to investigate the relationship between distraction and target quality, and overall quality, across sound zoning methods and program combinations. The model had $R = 0.69$, and the coefficients for target quality and distraction were 4.86 and -6.16 respectively, indicating that these quantities were broadly of equivalent importance to the determination of overall quality for these stimuli.

A positive correlation between target quality and planarity was found with $R = 0.74$, indicating that planarity is likely

to be an important aspect of target quality within sound zone scenarios. A smaller negative correlation was found between contrast and distraction $R = -0.42$. As previous work suggests, contrast is one of many important aspects of distraction.

Acknowledgment

This research forms part of the POSZ project (www.posz.org), supported by Bang & Olufsen AS and EPSRC DTG, and was also supported by EPSRC Grant Ref. EP/L000539/1 (www.s3a-spatialaudio.org).

References

- [1] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation", *IEEE Trans. Aud., Sp. and Lang. Proc.*, Vol 19 (7), 2046–2057, 2011
- [2] C.E. Williams, K.N. Stevens, M.Klatt, "Judgments of the acceptability of aircraft noise in the presence of speech", *J. of Sound and Vib.*, Vol 9 (2), 263–275, 1969
- [3] K.S. Pearsons, R.L. Bennett, S.Fidel, "Environmental health effects research series - speech levels in various noise environments", *Office of Health and Ecological Effects - U.S. Environmental Protection Agency*, 1977
- [4] D.S. Brungart, B.D. Simpson, M.A. Ericson, and K.R. Scott, "Informational and energetic masking effects in the perception of multiple simultaneous talkers", *J. of the Acoust. Soc. of Am.*, Vol 110 (5), 2527–2538, 2001
- [5] Y. Li, and D. Wang, "On the optimality of ideal binary time-frequency masks", *Speech Communication*, Vol 51 (3), 230–239, 2009
- [6] D.S. Brungart, P.S. Chang, B.D. Simpson, D. Wang, "Multitalker speech perception with ideal time-frequency segregation: effects of voice characteristics and number of talkers", *J. of the Acoust. Soc. of Am.*, Vol 125 (6), 4006–4022, 2009
- [7] J. Francombe, "Perceptual evaluation of audio-on-audio interference in a personal sound zone system", *PhD. Thesis*, University of Surrey, UK., August 2014.
- [8] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, "Modelling listener distraction resulting from audio-on-audio interference", *Proc. of Meet. on Acoust.*, Vol 19 (1), 2013
- [9] International Telecommunications Union - Telecommunications P.862, "Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", 2001
- [10] International Telecommunications Union - Telecommunications P.862, "Perceptual objective listening quality assessment", 2011
- [11] P. Coleman, P.J.B. Jackson, M. Olik, M. Møller, M. Olsen, and J.A. Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array", *J. of the Acoust. Soc. of Am.* Vol 135 (4), 1929–1940, 2014
- [12] M. F. Simón Gálvez, S. J. Elliott, J. Cheer, "A superdirective array of phase shift sources", *J. of the Acoust. Soc. of Am.*, Vol 132 (2), 746–756, 2012
- [13] J. Choi and Y. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. of the Acoust. Soc. of Am.*, Vol. 111 (4), 1695–1700, 2002
- [14] M. Poletti, "An investigation of 2-D multizone surround sound systems", in *Proc. 125th Aud. Eng. Soc. Conv.*, San Francisco, October, 2008
- [15] P. Coleman, P.J.B. Jackson, M. Olik, and J.A. Pedersen, "Personal audio with a planar bright zone", *J. of the Acoust. Soc. of Am.*, Vol. 136 (4), 1725–1735, 2014
- [16] S.J. Elliott, J. Cheer, J.W. Choi, and Y. Kim, "Robustness and regularization of Personal Audio systems", *IEEE Trans. on Aud. Sp. and Lang. Proc.*, Vol. 20 (7), 2123–2133, 2012
- [17] M. Olik, J. Francombe, P.J.B. Jackson, P. Coleman, M. Olsen, M. Møller, R. Mason, and S. Bech, "A comparative performance study of sound zoning methods in a reflective environment", *52nd Aud. Eng. Soc. Conf.*, Guildford, September, 2013
- [18] J. Cheer, S. J. Elliott, M. F. Simón Gálvez, "Design and implementation of a car cabin personal audio system", *J. of the Audio Eng. Soc.*, Vol 61 (6), 412–424, 2013
- [19] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers", *J. of the Acoust. Soc. of Am.*, Vol 131 (6), 4518–4525, 2012
- [20] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint", *J. of the Acoust. Soc. of Am.*, Vol 135 (2), 734–741, 2014
- [21] J.H. Chang, J. Choi, and Y. Kim, "A plane wave generation method by wave number domain point focusing", *J. of the Acoust. Soc. of Am.*, Vol 128 (5), 2010

- [22] R. Humphrey, "Playrec, Multi-channel Matlab Audio", <http://www.playrec.co.uk/>, accessed June 2014
- [23] ITU-R, "BS.1534-1 Method for the subjective assessment of intermediate quality level of coding systems", 2003
- [24] A.W. Mills, "On the minimum audible angle", *J. of the Acoust. Soc. of Am.*, Vol 30, 1958
- [25] E. Vincent, "MUSHRAM: A Matlab interface for MUSHRA listening tests", <http://c4dm.eecs.qmul.ac.uk/objectcoding/>, 2005
- [26] S.S. Shapiro, M.B. Wilk, "An analysis of variance test for normality (complete samples)", *Biometrika*, Volume 52 (3-4), 591-611, 1965
- [27] J.W. Tukey, "The problem of multiple comparisons", *The collected works of John W. Tukey VIII. multiple comparisons: 1948-1983*, 1-300, 1953
- [28] G. Glass, P. Peckham, J. Sanders, "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance", *Review of educational research*, Volume 42 (3), 1972