Perceptually optimised loudspeaker selection for the creation of personal sound zones

Jon Francombe[1], Philip Coleman[2], Marek Olik[2], Khan Baykaner[1], Philip J.B. Jackson[2], Russell Mason[1], Martin Dewhirst[1], Søren Bech[3], and Jan Abildgaard Pedersen[3]

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

[2] *Centre for Vision, Speech and Signal Processing, 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 Jon Francombe (`j.francombe@surrey.ac.uk`)

ABSTRACT

Sound field control methods can be used to create multiple zones of audio in the same room. The separation achieved by such systems has classically been evaluated using physical metrics including acoustic contrast and target-to-interferer ratio (TIR). However, to optimise the experience for a listener it is desirable to account for perceptual factors. A search procedure was used to select 5 loudspeakers for production of 2 sound zones using acoustic contrast control. Comparisons were made between searches driven by physical (programme-independent TIR) and perceptual (predicted distraction from a perceptual model) cost functions. Performance was evaluated on TIR and predicted distraction; results suggested that the perceptual cost function showed some benefits over physical optimisation although the perceptual model needs further work.

1. INTRODUCTION

One of the many potential uses for sound field control methods is the production of two or more 'zones' of audio in the same space, enabling simultaneous replay of more than one audio programme to multiple listeners with little or no interference between programmes. A number of methods have been developed to facilitate such systems, based on control of the energy between the zones (e.g. [1, 2]) or synthesis of a sound field with attenuated regions (e.g. [3, 4, 5]). Areas of high and low pressure (or 'bright' and 'dark' zones) are produced, giving the potential for the resultant sound fields to be superposed to produce multiple audio zones. Usually, evaluation is based on a 'single-sided' case, considering the attenuation achieved in the dark zone with respect to the bright zone and, in certain circumstances, the accuracy of reproduction in the target zone. For example, in a system comprising two zones A and B: acoustic contrast is the ratio of sound pressure in zone A due to programme A to sound pressure in zone B due to programme A. The control effort, the electrical power required by the array for reproduction, is also commonly reported. It is also beneficial to consider the situation from the perspective of a single zone: target-to-interferer ratio (TIR), the ratio of sound pressure in zone A due to programme A and sound pressure in zone A due to programme B, was used in [6] to evaluate channel separation with respect to a single zone when multiple targets were reproduced.

Whilst such metrics are undeniably useful for quantifying performance in the physical domain, they are insufficient for evaluating performance of systems intended for real-world applications as they do not account for the experience of a listener. Perhaps due to the difficulties involved in establishing real implementations of sound zone systems, there has been little perceptual research in this area, although the required separation between two programme items has been considered [7, 8] and occasionally comments from informal listening are reported. However, to improve the performance of sound field control methods from a listener-orientated perspective it is desirable to integrate perceptual evaluation. Consequently, recent work has focussed on modelling the 'distraction' experienced by a listener in the presence of multiple audio programmes [9]. The ability to model the listener experience in a sound zone also presents the opportunity to make system design decisions in a perceptually optimal way. In [10], a sequential forward-backward search procedure was used to select a number of loudspeaker positions from candidate locations, motivated by the physical constraints on loudspeaker positions in an automotive system. The search was driven by a cost function including weighted terms related to acoustic contrast, matrix condition number, and control effort. Here, in order to establish the potential benefits of combining perceptual and physical optimisation of a sound zone system, a similar selection procedure was performed using physical and perceptual cost functions to select 5 loudspeakers from 40 candidate positions.

The design and results of this experiment are described in Sections 2 and 3 respectively. In order to validate the selection, a subjective listening test was performed; this is reported in Section 4.

2. EXPERIMENT SETUP AND DESIGN

As discussed above, an experiment was performed in order to compare sound zone systems optimised with physical and perceptual cost functions. Two cost functions were used to drive a search in order to produce optimal sets of 5 loudspeakers. The experiment setup is detailed in the following sections.

2.1. Physical setup

The candidate source array comprised 40 loudspeakers (Genelec 8020b), distributed evenly around a circle of radius 1.63 m at the acoustic centre of the loudspeakers and at a height of 1.6 m. The two sound zones 'A' and 'B' of dimensions 0.3x0.4 m and separated by 1.4 m (centre-to-centre) were defined based on impulse responses measured by placing a 6x8 array of microphones (Countryman B3 omni), with 5 cm spacing, in multiple positions in the zones. A subset of 96 microphone positions in each zone, with 3.5 cm spacing, was used to calculate the sound zone filters, and a single microphone in a central position in each zone, spatially mismatched from those used for setup, was used for performance predictions.

2.2. Sound zone filter calculation

The underlying sound field optimisation used was acoustic contrast control (ACC) [1], which has been shown to be very effective at creating regions of cancellation [11, 2], although the search procedure detailed below could theoretically be applied to any control method.

As in [12], ACC is written as a constrained optimisation problem, where for a single frequency the pressure in the dark zone is minimised whilst constraining the pressure in the bright zone to be as large as possible, with an added effort term to regularise the solution:

$$\alpha = \mathbf{p}_d^H \mathbf{p}_d + \lambda_1 (\mathbf{p}_b^H \mathbf{p}_b - B) + \lambda_2 (\mathbf{q}^H \mathbf{q} - E), \quad (1)$$

where $\mathbf{p} = \mathbf{Gq}$ is the vector of complex pressures at the microphones, $\mathbf{G}$ is the plant matrix, $\mathbf{q}$ is the vector of complex source strengths, and the subscripts $\cdot_b$ and $\cdot_d$ denote bright and dark zones, respectively. The target sound pressure level B is here taken to be 76 dB, and the control effort limit E is 0 dB. It is shown in [12] that the optimal solution can be obtained by taking the eigenvector corresponding to the maximum eigenvalue of $(\mathbf{G}_d^H \mathbf{G}_d + \lambda_2 \mathbf{I})^{-1} (\mathbf{G}_b^H \mathbf{G}_b)$.

Implementation of ACC filters for replaying music requires the source weights to be optimised at multiple frequencies. In [13], it is shown that the control effort constraint E imposes a physical limit on system effort but does not necessarily ensure that the inversion of $\mathbf{G}_d^H \mathbf{G}_d$ is well conditioned, particularly at high frequencies. Therefore, λ_2 was set by reducing the matrix condition number by a factor of 10 at all frequencies, before enforcing E if necessary.

The plant matrices $\mathbf{G}$ were populated by transforming the measured impulse responses into1 the frequency domain. First, the impulse responses were captured using the maximum length sequence (MLS) technique, in each case using two averages of two MLS sequences. The time domain impulses were then downsampled to 16 kHz for the filter calculations, and truncated to the room's reverberation time of 225 ms. Finally, a 8192 point fast Fourier transform (FFT) was used to transform into the frequency domain. The filter weights $\mathbf{q}$ were subsequently calculated for all frequency bins up to the Nyquist bin, and FIR filters constructed for each loudspeaker by appropriate treatment of the negative frequency components and applying the inverse FFT with a 4096 sample modelling delay.

2.3. Cost functions

Two cost functions were used to drive the loudspeaker search: a simple physical cost function and a perceptual cost function.

The physical cost function (PhysTIR) optimised mean programme-independent TIR equally across all 4096 frequency bins (up to 8 kHz) at single positions in the centre of each zone.

The perceptual cost function (DistOpt) optimised predictions of distraction made using a statistical model (described in [9]). The model was trained on distraction ratings of 54 audio-on-audio interference situations (full-factorial combinations of 3 target programmes, 3 interferer programmes at 3 levels, with and without road noise); distraction was defined as *"how much the alternate audio pulls your attention or distracts you from the target audio"* on a scale from 0 to 100 with endpoints labelled *"not at all distracting"* to *"overpowering"*. For each stimulus combination, time-frequency TIR maps were created from monophonic recordings using the Computational Auditory Signal-Processing and Perception (CASP) auditory pre-processor [14]. Simple statistical features were extracted from the TIR maps and a principal component analysis (PCA) was performed, from which the first 17 components were used as features in a linear regression model.

The two cost functions meant that a total of 5 loudspeaker sets were determined: 1 for the PhysTIR cost function, and 4 for each programme combination for the DistOpt cost function.

2.4. Loudspeaker search

A sequential forward-backward search (*plus l-take away r* [15, p. 220]) was used to select 5 loudspeakers from the 40 candidate positions, based on the cost functions. This algorithm comprises l iterations of the sequential forward search (SFS) algorithm, followed by r iterations of the sequential backward search (SBS) algorithm. Here, as in [10], $l = 2$ and $r = 1$. For the SFS algorithm the selected set J_k contains k features (loudspeakers) from the full set X. The features ξ_x in the candidate set $X - J_k$ are ranked according to their performance Y under each cost function such that

$$Y(J_k + \xi_1) \geq Y(J_k + \xi_2) \geq \cdots \geq Y(J_k + \xi_{X-k}), \quad (2)$$

and the feature set J_{k+1}, initialised with $J_0 = \emptyset$, becomes $J_{k+1} = J_k + \xi_1$. In order to produce acceptable performance in both zones, the ranking of Y was based on the minimum of the zone A and zone B scores,

$$Y(J_k + \xi_x) = \min\{Y_A(J_k + \xi_x), Y_B(J_k + \xi_x)\}. \quad (3)$$

In this way, selection of loudspeaker sets that produced good performance in one zone at the cost of the other zone was avoided. The SBS algorithm operates in a similar manner, reducing the feature set on each iteration. The candidate features ξ for removal from J_k are ranked such that

$$Y(J_k - \xi_1) \geq Y(J_k - \xi_2) \geq \cdots \geq Y(J_k - \xi_{X-k}), \quad (4)$$

and the feature set becomes $J_{k+1} = J_k - \xi_1$.

For the DistOpt cost function, where the score should be minimised through the choice of loudspeakers, the sets selected for the forward and backward steps become $J_{k+1} = J_k + \xi_{X-k}$ and $J_{k+1} = J_k - \xi_{X-k}$, respectively, and the maximin criterion (Eq. 3) becomes the minimax,

$$Y(J_k + \xi_x) = \max\{Y_A(J_k + \xi_x), Y_B(J_k + \xi_x)\}. \quad (5)$$

With only 5 active loudspeakers, the separation that can be achieved between zones is limited, however it was felt that this was an appropriate system on which to demonstrate the principle of integration with perceptual optimisation, and relevant to realistic situations as 5 loudspeakers may be present in current domestic systems (e.g. 5.1 systems).

2.5. Stimuli

The DistOpt cost function requires reproductions of audio programme items through the sound zone system. Two target programme items (pop and classical music) were used in each zone. The distraction model used in the DistOpt cost function is still in early development; the model was shown to fit well to the training data but perform relatively poorly on a independently collected validation set [9]. For this reason, stimuli on which the model was trained were used for the programme items. The zone A targets were pop music (The Killers *"On Top"*) and classical music (Brahms *"Hungarian Dance No. 18"* for string orchestra); the zone B targets were pop music (The Bravery *"Give In"*) and classical music (Mahler *"Symphony No. 5 Mov. 4"*, string section). This gave a total of 4 combinations. The specific programme items were chosen to cover a range of realistic music programmes. The stimuli were perceptually loudness matched for equal long-term loudness (LTL) level (details in [8]).

The audio for the DistOpt cost function was produced by convolving the predicted filtered system response at 1 microphone in the centre of zones A and B with the respective target programmes.

2.6. Evaluation metrics

As well as observation of the selected loudspeakers, two metrics were used to evaluate performance. The metrics were essentially those on which the cost functions were based: physical (programme-independent) TIR, and predictions of distraction made using the perceptual model.

The results based on these metrics are described in Section 3. Alongside these evaluation metrics, a listening test was performed in order to validate the perceptual metric. This procedure is discussed in Section 4.

3. PERFORMANCE EVALUATION

The selected loudspeakers and performance on the evaluation metrics are described in the following sections.

3.1. Selected loudspeakers

Figure 1 shows the 5 selected loudspeakers for the PhysTIR cost function. 5 adjacent loudspeakers approximately equidistant from the two zones were selected. This is unsurprising given the requirement for optimal performance in both zones.

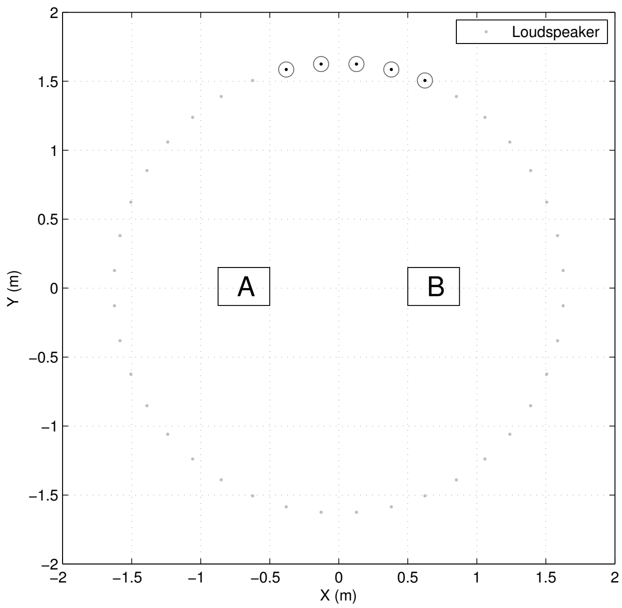


Fig. 1: Selected loudspeakers for PhysTIR cost function. Selected loudspeakers are circled and darker.

The loudspeakers selected for the DistOpt cost function are shown in Figure 2. There is considerable variation between selected loudspeakers, indicating a strong dependancy on programme material.

The loudspeaker selection appears to be linked to the combination of programme material. When the programme material is the same in both zones, the DistOpt cost function selects a spread of loudspeakers; both loudspeaker sets for this condition include a 3 loudspeakers in close proximity in addition to 2 outlying loudspeakers. When the programme material differs between zones, the DistOpt cost function selects small clusters of loudspeakers (or individual loudspeakers). There is a degree of reciprocity evident between the sets selected for the pop/classical and classical/pop cases, suggesting that programmes with similar time-frequency content require similar loudspeaker sets for optimal performance. This reciprocity is also exhibited in the two cases with simlay programme material in each zone. Determining an aggregate set of loudspeakers that would optimise performance for a range of loudspeakers would be an interesting extension of this work.

3.2. Physical TIR

Figure 3 shows mean physical TIR results (from performance predictions) for each cost function and programme combination as well as the result in each zone. The average physical TIR for the PhysTIR cost function is 9 dB, with very similar performance in the two zones (this is ensured by the minimax criterion as specified in the optimisation search). In every case, the loudspeakers selected by the DistOpt cost function produced lower average physical TIR between the two zones, although in some cases the TIR is higher in one zone or the other (this discrepancy is most pronounced for the two cases in which the zone A target is classical music). When the programme content is taken into account (i.e. using the DistOpt cost function), Figures 3b and 3c suggest that the zone with the classical music target is prioritised in terms of physical TIR. This result indicates that the programme content has a pronounced effect on the listening experience, and that it may be beneficial to account for this in optimisation.

3.3. Predicted distraction

Figure 4 shows mean distraction predictions (using predicted audio) for each cost function and programme combination as well as the result in each zone. In each case, the loudspeakers selected with the DistOpt cost function produced the lowest distraction scores. The fact that these selections did not produce the lowest physical TIR scores provides evidence that use of perceptual metrics is important for optimising the listener experience in a sound zone system. In addition to reducing the overall distraction prediction, use of the DistOpt cost function also provided the benefit of reducing the discrepancy in predicted distraction between the two zones.

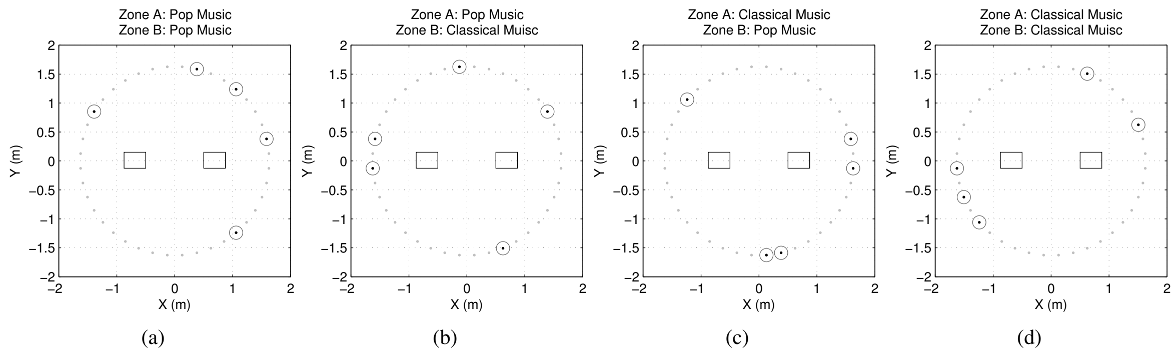


Fig. 2: Loudspeakers selected using the DistOpt cost function, for each programme combination

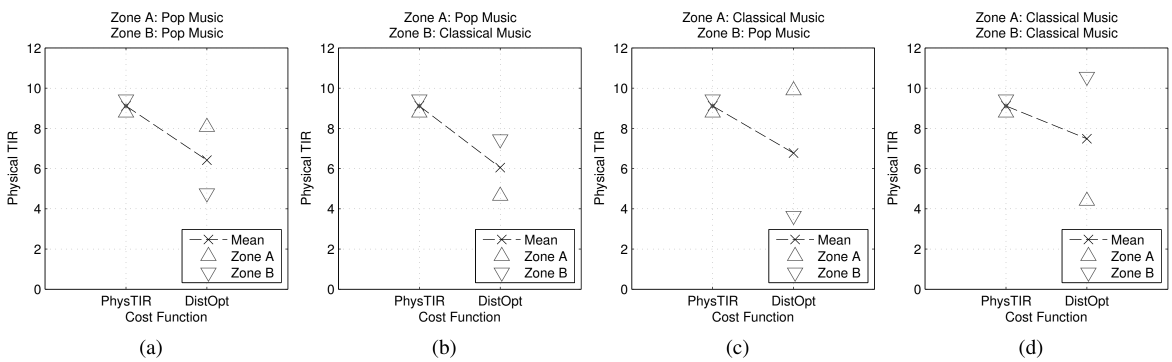


Fig. 3: Mean (predicted) physical TIR for zones A and B separately and combined, for each cost function and stimulus combination.

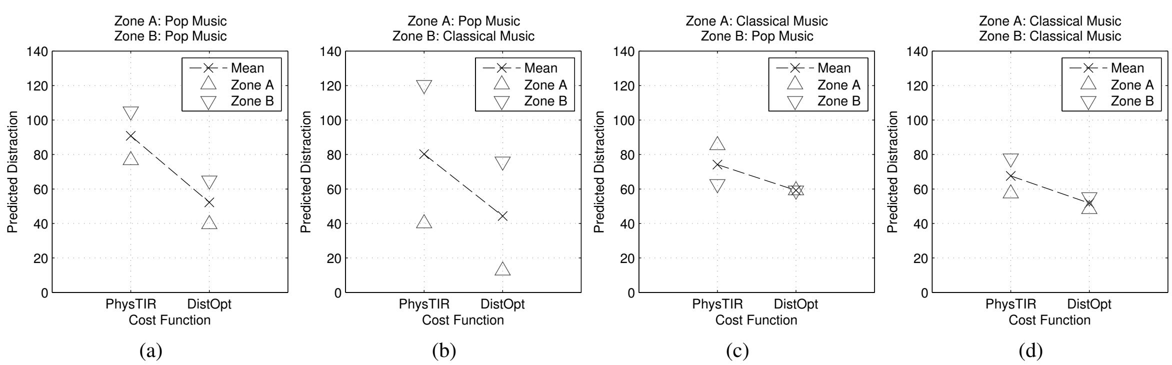


Fig. 4: Predicted distraction (from simulated audio) in zone A, B, and overall, for each cost function and stimulus combination.

4. VALIDATION EXPERIMENT

The results presented above suggested that there is potentially a perceptual advantage to be gained by using a perceptually motivated cost function. In order to validate this finding, listening tests were performed using recordings made in the sound zone system described above.

4.1. Experiment design

Listeners were asked to rate distraction (as defined above) for each programme combination (4) and each cost function (2) in zones A and B. In order to replicate the audio used by the distraction model, the listening test was performed with monophonic recordings made in the centre of each zone, replayed over headphones with the same mono signal fed to the left and right channels. This therefore limits the generality of the results, as spatial factors can not be accounted for, producing an incomplete evaluation of the listening experience. However, this allows the closest comparison with the measured results, and hence can be used to validate the use of a perceptual cost function and the performance of the distraction model.

The 8 stimuli (and a repeat of each judgement to give a total of 16 trials for each zone) were presented individually in a random order alongside a reference stimulus (just the target programme for the zone under test). The replay level was calibrated to approximately 70 dBLA$_{eq(15s)}$ using a binaural dummy head. Participants were asked to make ratings of distraction on a single scale using the interface shown in Figure 5. The test was preceded by a familiarisation stage in which subjects were given the opportunity to audition a range of the stimuli.

Eight listeners participated in the test. Listeners were undergraduate students on the Music and Sound Recording degree course at the University of Surrey, or postgraduate students in the Institute of Sound Recording; all students had experience of technical listening and listening test participation. The order of participation in zone A and B was balanced so that an equal number of subjects performed the first test in each zone.

4.2. Results

In order to evaluate the performance in both zones, a dummy variable 'overall distraction' was calculated for each programme combination and subject by taking the mean across both zones and repeats.

Figure 6 shows overall distraction (with 95% confidence intervals calculated using the t-distribution), alongside

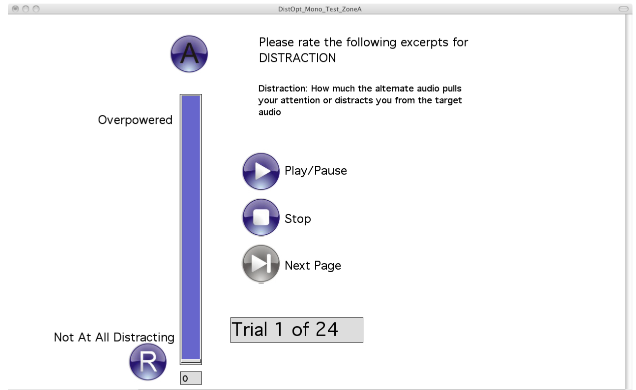


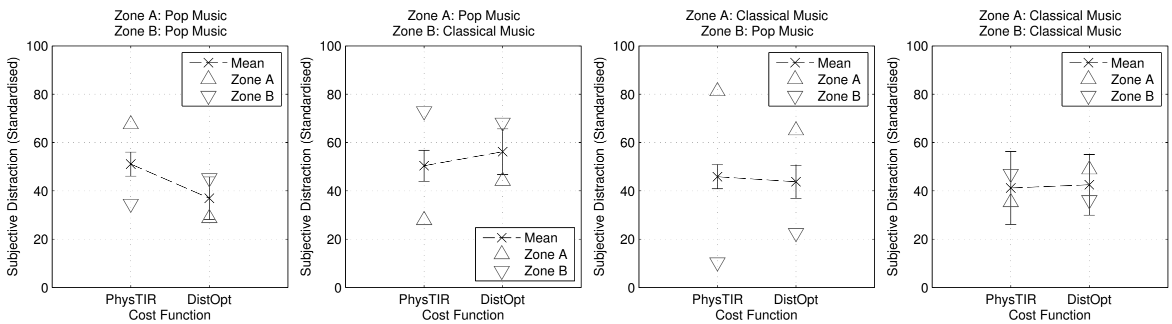
Fig. 5: Listening test interface

the mean distraction in each zone. There are generally not pronounced differences between the subjective distraction scores produced by the two cost functions with the exception of the pop-on-pop case (Figure 6a) for which the DistOpt cost function produces significantly lower distraction. In the other cases, mean distraction produced by the PhysTIR and DistOpt cost functions tends to be very similar. However, the DistOpt cost function does tend to reduce the discrepancy between zones; for the first three stimulus combinations the worst zone for the DistOpt cost function always performs better than the worst zone for the other cost functions. This is important as in a real system it is essential that performance in both zones is optimised. The results suggest that there are potential benefits of using the perceptual cost function in order to produce optimal perceptual performance.

4.2.1. Model Performance

As stated above, the distraction model used in the DistOpt cost function is still undergoing development and was shown in [9] to fit poorly to stimuli on which it was not trained. In order to assess model performance, the fit between predictions and observations of distraction for the 8 stimuli (2 cost functions, 2 target programmes, and 2 interferer programmes) in each zone was calculated.

Figure 7 shows model fit for zones A and B. In zone A, the model prediction is reasonable (root-mean-square error (RMSE) = 16.4) and the linearity of predictions is good ($r = 0.72$). However, in zone B the model fit is poor, particularly in terms of the magnitude of the scores (which are all overpredicted). When designing the op-



(a) (b) (c) (d)

Fig. 6: Overview of perceived distraction for each zone and the mean of both zones. Error bars show 95% confidence intervals calculated using the t-distribution for the overall mean distraction from both zones for the specified programme and cost function.

timisation experiment, stimuli that the model had been trained on were selected in order to make the model predictions more accurate. However, the model was trained on the one-sided cases (i.e. just the zone A combinations) and therefore it is unsurprising that the zone B fit is less accurate.

Even with the poor fit, there is still a strong positive correlation between observations and predictions ($r = 0.58$) suggesting that the model has some ability particularly in ranking the distraction of audio-on-audio interference situations; as the DistOpt optimisation is performed based on distraction rankings (Section 2.4), it is still feasible for the distraction-based cost function to show a perceptual advantage.

These results suggest that improvements in the perceptual model would increase the benefit that can be achieved by including such a model in an optimisation procedure.

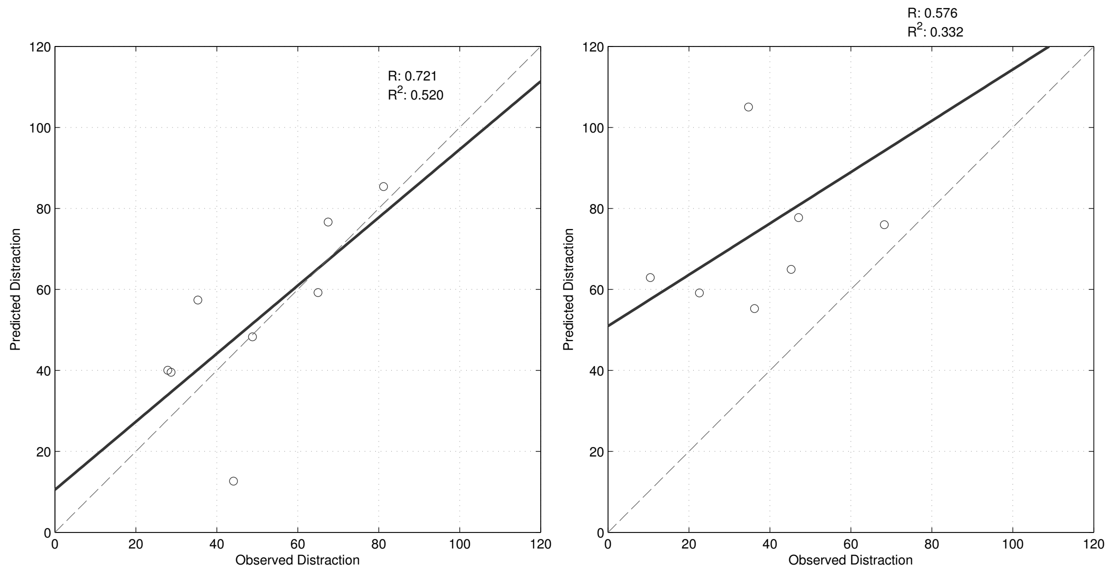
5. CONCLUSIONS AND FURTHER WORK

An experiment was performed in which a sequential forward-backward search was used to select 5 loudspeakers from a candidate set of 40 positions in a circular array for production of 2 sound zones using acoustic contrast control. The optimisation was based on two cost functions: physical TIR (PhysTIR) and predictions of distraction from a perceptual model (DistOpt). Four programme combinations (two targets in each of two sound zones) were used for the DistOpt cost function.

The loudspeakers selected by the DistOpt cost function were found to vary substantially between programme material combinations. Some patterns were observed: the loudspeakers selected seemed to relate to the combination of programme material. This work could be extended by searching for an optimal set of loudspeakers for all programme material rather than optimising individually for separate programme items.

The physical TIR was suggested to be insufficient for perceptual optimisation of the system, as the loudspeakers selected by the PhysTIR cost function produced higher distraction predictions than those selected using the DistOpt cost function. The programme material content was found to greatly influence the situation.

In order to evaluate the results, a listening test was performed in which monophonic recordings of the stimuli reproduced through the selected loudspeakers were rated for distraction by 8 listeners. The differences in the ratings were found to be small, but it was found that the DistOpt cost function could potentially reduce perceived distraction or even out the perceptual performance of the two zones. The distraction model was found to make reasonably accurate predictions for the zone A stimuli (i.e. the same programme material on which it had been trained) but predict poorly for the zone B stimuli (i.e. the reverse cases). The absolute values of the predictions were poor (overpredicted) although the ranking appeared reasonably accurate ($r = 0.58$). These findings suggest that development of perceptual models is necessary before further advantage can be gained from including per-



(a) Zone A (b) Zone B

Fig. 7: Predicted distraction (from the perceptual model) against observed distraction (mean listening test score) for all stimuli in each zone

ceptual models in the optimisation stage of a sound zone setup.

Currently the model works for monophonic input files, but the fact that the selected loudspeakers vary considerably in positions suggests that spatial effects are likely to affect the situation. Therefore, development of a binaural model is desirable. It is also clear that the model must be trained on a wider variety of programme material in order to generalise well to new stimuli.

However, regardless of weaknesses in the perceptual model, the procedure outlined above has highlighted differences in physical and perceptual evaluation and optimisation of sound zone systems; where such differences exist, it is desirable to optimise in a perceptually relevant manner in order to attain optimal performance from the experience of a listener.

6. ACKNOWLEDGEMENTS

This work forms part of the ‘Perceptually Optimised Sound Zones’ (POSZ) project funded by Bang & Olufsen. The authors would like to acknowledge valuable discussions with members of the POSZ project team.

7. REFERENCES

[1] Joung-Woo Choi and Yang-Hann Kim. Generation of an acoustically bright zone with an illuminated region using multiple sources. *J. Acoust. Soc. Am.*, 111(4):1695–1700, April 2002.

[2] Mincheol Shin, Sung Q. Lee, Filippo M. Fazi, Philip A. Nelson, Daesung Kim, Semyung Wang, Kang Ho Park, and Jeongil Seo. Maximization of acoustic energy difference between two spaces. *J. Acoust. Soc. Am.*, 128(1):121–131, 2010.

[3] Y.J. Wu and T.D. Abhayapala. Spatial multizone soundfield reproduction: Theory and design. *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(6):1711–1720, 2011.

[4] Mark Poletti. An investigation of 2-d multizone surround sound systems. In *125th Audio Engineering Society Convention, San Francisco, CA, USA*, 10 2008.

[5] N. Radmanesh and I. S. Burnett. Generation of isolated wideband sound fields using a combined two-stage lasso-ls algorithm. *Audio, Speech,*

and Language Processing, IEEE Transactions on, 21(2):378 –387, feb. 2013.

[6] Jin-Young Park, Ji-Ho Chang, and Yang-Hann Kim. Generation of independent bright zones for a two-channel private audio system. *J. Audio Eng. Soc*, 58(5):382–393, 2010.

[7] W.F. Druyvesteyn, R.M. Aarts, A.J. Asbury, P. Gelat, and A. Ruxton. Personal sound. *Proceedings of the Institute of Acoustics*, 16(2):571–585, 1994.

[8] J. Francombe, R. Mason, M. Dewhirst, and S. Bech. Determining the threshold of acceptability for an interfering audio programme. In *132nd Audio Engineering Society Convention, Budapest, Hungary*, 2012.

[9] Jon Francombe, Russell Mason, Martin Dewhirst, and Søren Bech. Modelling listener distraction resulting from audio-on-audio interference. In *Proceedings of 21st International Congress on Acoustics, Proceedings of Meetings on Acoustics (POMA)*, Montreal, Quebec, Canada, June 2013. Presented at ICA2013, Montreal, 2-7 June 2013.

[10] Philip Coleman, Martin Moller, Martin Olsen, Marek Olik, Philip Jackson, and Jan Abildgaard Pedersen. Performance of optimized sound field control techniques in simulated and real acoustic environments. *J. Acoust. Soc. Am.*, 131(4):3465–3465, 2012.

[11] Ji-Ho Chang, Chan-Hui Lee, Jin-Young Park, and Yang-Hann Kim. A realization of sound focused personal audio system using acoustic contrast control. *J. Acoust. Soc. Am.*, 125(4):2091–2097, 2009.

[12] Stephen J. Elliott, Jordan Cheer, Jung-Woo Choi, and Youngtae Kim. Robustness and regularization of personal audio systems. *IEEE Trans. Audio, Speech and Lang. Proc.*, 20(7):2123–2133, 2012.

[13] P. Coleman, P.J.B. Jackson, M. Olik, M. Olsen, M. Møller, and J.A. Pedersen. The influence of regularization on anechoic performance and robustness of sound zone methods. In *Proceedings of 21st International Congress on Acoustics, Proceedings of Meetings on Acoustics (POMA)*, 2013. Presented at ICA2013, Montreal, 2-7 June 2013.

[14] Morten L. Jepsen, Stephan D. Ewert, and Torsten Dau. A computational model of human auditory signal processing and perception. *J. Acoust. Soc. Am.*, 124(1):422–438, 2008.

[15] P. A. Devijver and J. Kittler. *Pattern recognition: A statistical approach*. NJ, USA: Prentice Hall, 1982.