MAXIMISING PERCEIVED DIFFUSENESS IN LOUDSPEAKER SYSTEMS WITH HEIGHT USING OPTIMISED RELATIVE LOUDSPEAKER LEVELS

Michael P. Cousins Institute of Sound and Vibration Research, University of Southampton, UK
Filippo Maria Fazi Institute of Sound and Vibration Research, University of Southampton, UK
Stefan Bleeck Institute of Sound and Vibration Research, University of Southampton, UK
Frank Melchior BBC Research & Development, Media City, Salford, UK

1 INTRODUCTION

Most loudspeaker systems are two-dimensional (2D). There is a potential with three-dimensional (3D) loudspeaker layouts to position elevated sources but also increase the perceived diffuseness. In concert hall acoustics, diffuse reverberation is perceived as desirable spatial attributes, namely, apparent source width and listener envelopment[1]. In reproduced sound, adding elevated sources has been shown to increase desirable spatial attributes e.g. listener envelopment, spatial impression[2], engulfment[3] and surroundedness[4]. However, it is not immediately clear at what amplitude additional elevated loudspeakers should be. Whilst there is research indicating the benefits of 3D layouts[3], theoretically adding a single Voice-Of-God (V.O.G.) loudspeaker above the listener would allow a 2D layout to become 3D. It was noted informally that adding a V.O.G. loudspeaker to a 2D layout with many loudspeakers made very little difference to the perceived diffuseness as the sound from the elevated loudspeaker was mostly masked by that from the many head-height loudspeakers. However, making the V.O.G loudspeaker the same sound pressure level (SPL) as the sum of the head-height loudspeakers had the opposite effect with the V.OG. loudspeaker now easily localised. From this it was hypothesised that there is an optimal amplitude distribution between the head-height and non-head-height loudspeakers that would maximise the perceived diffuseness.

This paper presents the results of two listening tests. Firstly, listeners performed an adjustment task in which the relative amplitude between a subset of loudspeakers at head-height and a subset of loudspeakers not at head-height could be varied to the point they determined most perceptually diffuse (section 2). This preferred relative amplitude difference (Inter-Subset Level Difference, ISLD) was then validated as part of a separate listening test based on the Multiple Stimulus with Hidden References and Anchor (MUSHRA)[5] methodology (section 3).

2 ADJUSTMENT EXPERIMENT

This section describes an adjustment task that was conducted, allowing listeners to choose the most perceptually diffuse level distribution for different arrangements of loudspeakers. Each arrangement of loudspeakers comprised of a head-height and a non-head-height subset and listeners could vary the Inter-Subset Level Difference (ISLD) to the point they determined most perceptually diffuse.

2.1 Listening Room

The “Audio Lab" at the University of Southampton (4.80m×3.97m×2.56m) has a reverberation time of 0.12s ±0.02s in 1/3 octave bands between 125Hz and 8kHz. An arrangement of 37 Kef HTS3001SE loudspeakers were mounted to a Bosch Rexroth mounting frame on the walls and ceiling of the listening room at the angles given in table 1.

Loudspeaker Positions			Head-Height Subsets			Non-Head-Height Subsets						
	Azimuth	Elevation	$n = 12$	$n = 6$	$n = 4$	$12/n/13$	$8/0/8$	$8/0/0$	$0/n/8$	$0/n/4w$	$0/n/4$	$0/n/1$
Floor	±45°	-56°				✓						
	±135°	-56°				✓						
Lower	0°	-20°				✓	✓	✓				
	±45°	-17°				✓	✓	✓				
	±90°	-24°				✓	✓	✓				
	±135°	-17°				✓	✓	✓				
	180°	-20°				✓	✓	✓				
Middle	0°	0°	✓									
	±30°	0°	✓	✓	✓							
	± 60°	0°	✓									
	±90°	0°	✓	✓								
	±120°	0°	✓		✓							
	±150°	0°	✓	✓								
	180°	0°	✓									
Upper	0°	27°				✓	✓		✓			
	±45°	24°				✓	✓		✓	✓		
	±90°	32°				✓	✓		✓			
	±135°	24°				✓	✓		✓	✓		
	180°	27°				✓	✓		✓			
Ceiling	±45°	52°				✓					✓	
	±135°	52°				✓					✓	
V.O.G	0°	90°				✓						✓

Table 1. Loudspeaker positions for all subsets. Each stimulus consisted of a head-height and a non-head-height subset.

2.2 Stimulus Material

Pink noise was used to avoid bias with frequency and material preference. Each individual loudspeaker was uncorrelated with every other loudspeaker and aligned in time and level (in 1/6 octave bands). Digital delay applied to each loudspeaker compensated for the differing propagation delay to the listening position (±10µs). Level alignment in 1/6 octave bands equalised each loudspeaker to within ±0.5dB from 95Hz to 20kHz (the working region of the loudspeakers).

2.3 Subjects

Listeners were 16 postgraduate or undergraduate students at the University of Southampton with self-reported normal hearing.

2.4 Stimuli

Each loudspeaker layout was divided into a head-height subset and a non-head-height subset of loudspeakers which the listener could crossfade between. The subsets were chosen to investigate the following factors:

- The number of loudspeakers at head-height.
- The number of loudspeakers not at head-height.
- The effect of different elevations of a non-head-height layer (wide or narrow span).
- The effect of placing a non-head-height layer above and/or below the head-height layer.

All the subsets are either, distributed evenly in azimuth, or symmetrical front-back and left-right if the subset could not be even in azimuth (n=4). This was done to minimise the number of variables and to prevent the relative level from the front and the back varying as the listener crossfaded between the subsets.

These factors led to 3 head-height subsets of loudspeakers, and 7 non-head-height subsets of loudspeakers (table 1). Stimuli are labelled using the convention $m_b/n/m_a$ where m_b and m_a are the number of loudspeakers below and above head-height respectively and *n* is the number of loudspeakers at head-height.

All 21 combinations of head-height and non-head-height subsets were adjusted twice by each listener in two 20 minute sessions separated with a 5 minute break.

2.5 Listener Response Method

The user interface in figure 1 was implemented in Max 6.1 and listeners were asked to,

> *"Move the slider from left to right in order to vary the spatial attributes of a noise stimulus. Your task is to find the point at which you perceive the sound field most diffuse. Ideal diffuseness is defined in this experiment as the sound coming from all directions with equal intensity. Therefore, the sound should ideally be impossible to localise and without any gaps (areas you perceive there is no sound coming from) in three dimensions."*

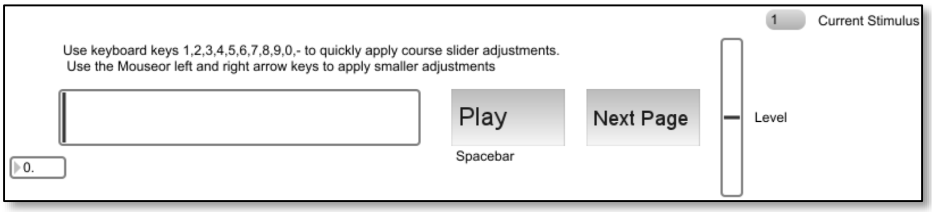

Figure 1. User interface for the adjustment task.

Every subset was adjusted to 70dBSPL A-Weighted at the central listening position. The slider implemented a -3dB constant power crossfade between a head-height subset and a non-head-height subset. This ensured the total SPL at the listening position was 70dBSPL A-Weighted independently of the slider position or the loudspeaker arrangement. The presentation order was randomised as was which of the two subsets was presented on the left/right of the slider.

The slider value s_s ranged from 0 to 1 for left to right respectively. This crossfaded between two subsets. The gains in dB for the head-height subset and the non-head-height subset are given by *X* and *Y* respectively.

$$X = 10\log_{10}(s) \quad (1)$$

and

$$Y = 10\log_{10}(1 - s) \quad (2)$$

respectively, where $s = s_s$, if the non-head-height subset is on the left, or $s = 1 - s_s$, if the head-height subset is on the left. This leads to an Inter-Subset Level Difference (*ISLD*) in dB of,

$$ISLD = 10\log_{10}(s/(1 - s)) \quad (3)$$

Therefore, positive *ISLD* values ($0 \leq s < 0.5$) indicate how many dB the head-height subset is louder than the non-head-height subset. An ISLD of zero ($s = 0.5$) means both subsets have the same level. Negative *ISLD* ($0.5 < s \leq 1$) values indicate the non-head-height subset is louder.

Keyboard shortcuts allowed quick coarse adjustments and a level fader allowed ±2dB of gain.

2.6 Post-screening

Listeners were ranked based on the mean absolute differences between the two adjustments of the same stimulus. A cut-off of 20% of the total scale excluded five listeners. This cut-off was chosen to remove the least consistent listeners whilst retaining an adequate sample size. The two repeats for each of the remaining 11 listeners were averaged and the data is shown in figure 2. Outliers (labelled with subject identifiers) are defined as listeners whose average rating for a stimulus extended more than 1.5 times the interquartile range (IQR) either, below the lower quartile ($< Q_1 + 1.5 \times IQR$), or above the upper quartile ($> Q_3 + 1.5 \times IQR$). The outliers are not far from the data medians and no further listeners were excluded. The whiskers indicate the range of the mean adjustment values excluding outliers.

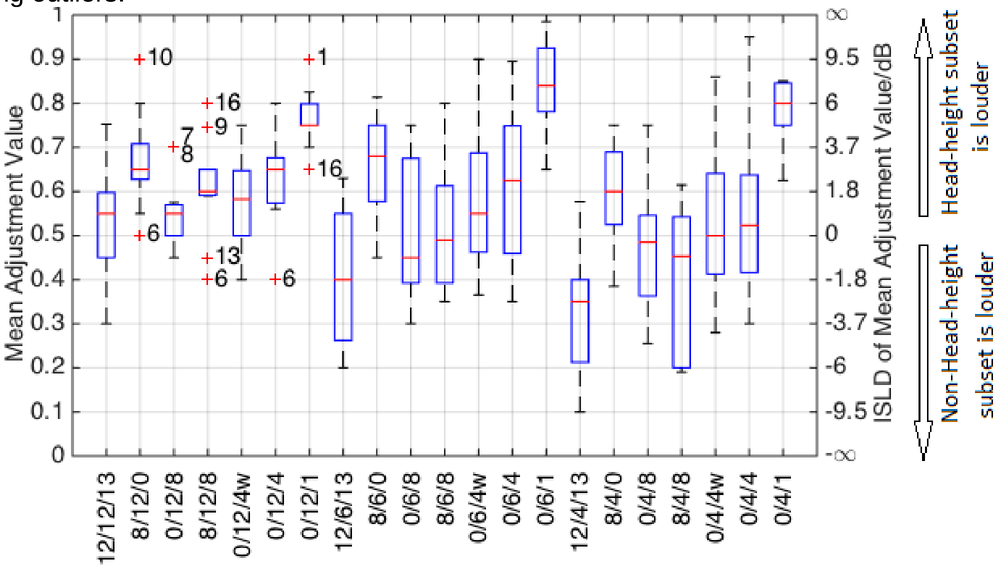

Figure 2. Boxplot of data from consistent listeners. Associated ISLD are shown on the right axis.

2.7 Results and Discussion

Figure 2 shows the median, to vary notably for different loudspeaker arrangements. It also shows a wide range, consistent with the difficulty of the task. In the *0/6/1* case, with 6 head-height loudspeakers and a V.O.G loudspeaker, one listener adjusted the slider to the end of the scale in both repeats indicating that effectively muting the V.O.G gave the maximum perceptual diffuseness. The mean adjustment values from all consistent listeners are converted to an ISLD using equation 3. These $ISLD_{mean}$ values are plotted in figure 3.

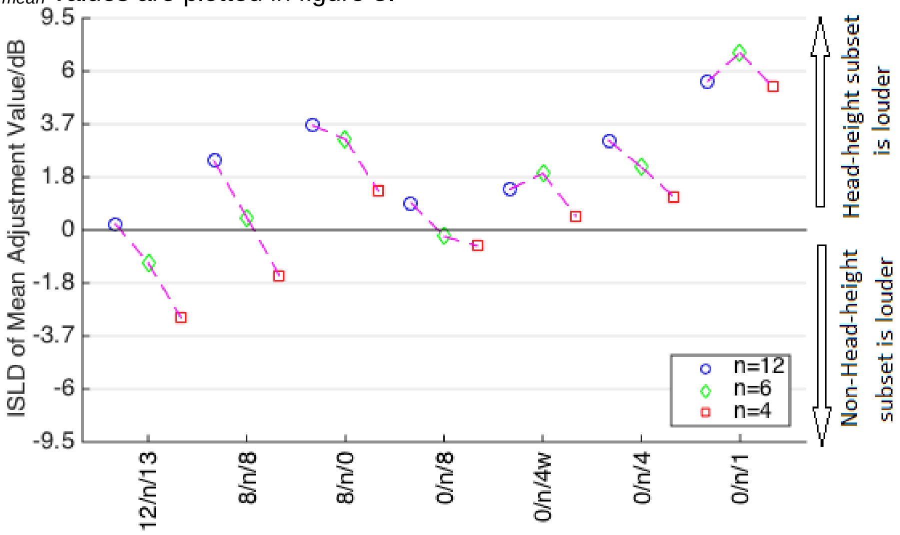

Figure 3. *ISLD* of mean adjusted *s* value ($ISLD_{mean}$), across all consistent listeners, for all stimuli.

Referring to the factors in section 1.4:

- The $ISLD_{mean}$ is positively correlated to the number of head-height loudspeakers. For stimuli with more head-height loudspeakers (e.g. $n = 12$), the head-height layer is adjusted to a higher level than for stimuli with fewer head-height loudspeakers (e.g. $n = 4$) to give the maximum perceived diffuseness.
- The $ISLD_{mean}$ is negatively correlated to the number of non-head-height loudspeakers. For stimuli with more non-head-height loudspeakers (e.g. *12/n/13)*, the non-head-height layer is adjusted to a higher level than for stimuli with fewer non-head-height loudspeakers (e.g. *0/n/1*) to give the maximum perceived diffuseness.
- The $ISLD_{mean}$ appears not to vary with the elevation of the non-head-height subset with both *0/n/4w* and *0/n/4* adjusted similarly.
- The $ISLD_{mean}$ varies when placing a non-head-height layer above and/or below the head-height layer with *8/n/8*, *8/n/0* and *0/n/8* all rated differently.

Repeated measures ANalysis Of VAriance (ANOVA) shows the choice of both head-height and non-head-height subset to be significant factors with *p*-values less than 0.05 for the 95% confidence interval and high *F*-statistics indicating the explained variance is greater than the unexplained variance meaning the result is unlikely to be due to chance. The significance value for the factor of head-height subset is p=0.017 (F(2,9)=6.674) and for the factor of non-head-height subset is p=0.004 (F(6,5)=16.815).

Pairwise comparison (tables 2 and 3) reveals, in general, insignificant differences between similar subsets. This is consistent with the range of the adjustments and the difficulty of the task. *0/n/1* is always significantly different from the other non-head-height subsets and is likely because this has the fewest loudspeakers and therefore the greatest variation in perceived diffuseness along the length of the slider.

	$n=6$	$n=12$
$n=4$	**0.128**	0.016
$n=6$		**0.200**

Table 2. Significance values from pairwise comparisons between head-height subsets. Significance has Bonferroni adjustment for multiple comparisons.

	$8/n/8$	$8/n/0$	$0/n/8$	$0/n/4w$	$0/n/4$	$0/n/1$
$12/n/13$	**0.337**	0.008	**0.271**	**0.322**	0.032	0.000
$8/n/8$		**0.073**	**1.000**	**1.000**	**0.925**	0.000
$8/n/0$			**0.242**	**1.000**	**1.000**	0.006
$0/n/8$				**1.000**	**0.860**	0.000
$0/n/4w$					**1.000**	0.008
$0/n/4$						0.002

Table 3. Significance values from pairwise comparisons between non-head-height subsets. Significance has Bonferroni adjustment for multiple comparisons.

Observing the first two factors shows listeners “turning down" the subset with fewer loudspeakers similar to attempting to reduce the level difference between an individual loudspeaker in the head-height subset and an individual loudspeaker in the non-head-height subset. The Inter-Subset Channel Level Difference (ISCLD) is an alternate measure of the relative level between subsets and is the difference in level between a single channel in the head-height subset and a single channel in the non-head-height subset. Figure 4 shows the $ISLD_{mean}$ values plotted against the ratio of the number of loudspeakers in the two subsets ($n/(m_a + m_b)$) as well as a comparison curve showing the ISLD had the listeners chosen that maintaining equal loudness from every loudspeaker is the most perceptually diffuse level distribution (*ISCLD=0*).

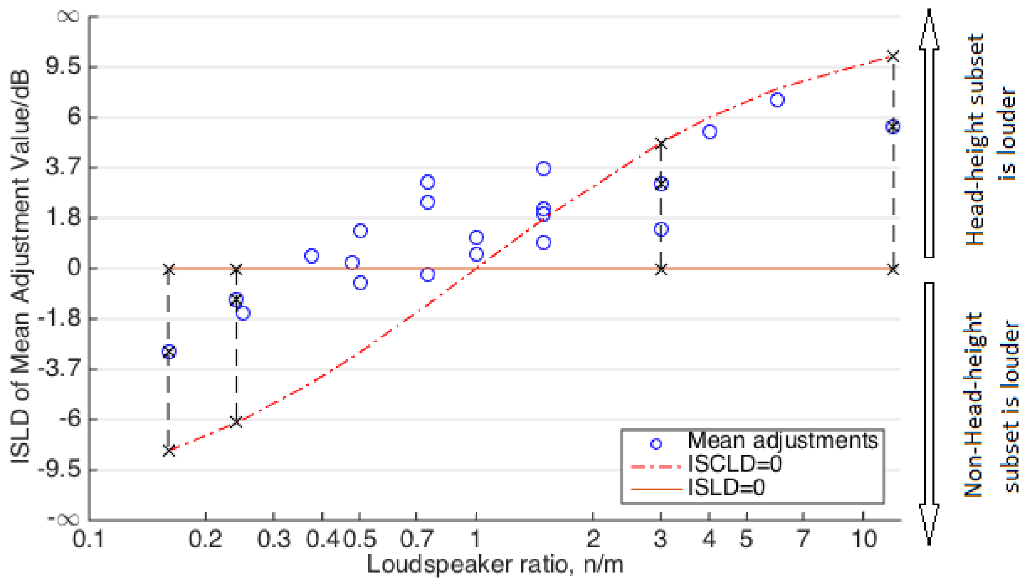

Figure 4. *ISLD* of mean adjusted *s* value plotted against the ratio of the number of head-height to non-head-height loudspeakers. Black crosses indicate the choice of stimuli for the validation experiment.

There is a slight trend towards the *ISCLD=0* curve. However, factors such as whether the non-head-height loudspeakers are below or above the head-height layer are not taken into account in any metric based on the loudspeaker ratio.

An initial hypothesis predicted the choice of slider value would be the value that minimised the InterAural Cross-correlation Coefficient (IACC). Simulations of the IACC using binaural impulse responses of each loudspeaker measured in the Audio Lab revealed this to not be the case. The IACC was found to be low for all the subsets except *0/n/1* and to vary by less than the just noticeable difference of IACC over the length of the slider.

Whilst there are clear trends in the data, the underlying psychoacoustic phenomenon behind listeners adjustment of ISLD is not fully understood and so the optimal ISLD cannot be found for an arbitrary arrangement of loudspeakers. However, it is also not clear how important using an optimised ISLD is in terms of absolute perceived diffuseness. The $ISLD_{mean}$ can be assumed to be the optimal ISLD but how much more perceptually diffuse is it than using *ISLD=0* or *ISCLD=0* requires further investigation. The validation experiment (section 3) was designed to investigate the absolute difference in perceived diffuseness for different ISLD and to test whether using $ISLD_{mean}$ increases the perceived diffuseness.

3 VALIDATION EXPERIMENT

Some stimuli were chosen to validate the adjustment task by testing and quantifying improvements in the perceived diffuseness when using an optimised ISLD. These were included as part of a larger listening test designed to assess the perceptual diffuseness of a range of stimuli. The details and results of this experiment are reported in [6] but the relevant parts are summarised here.

3.1 Experiment

A MUSHRA[5]-style listening test was chosen to make direct comparison between seven stimuli with moderate differences at once and allow for consistent absolute assessments between stimuli.

The proposed optimal ISLD ($ISLD_{mean}$) was compared to *ISCLD=0* and *ISLD=0* ISLD options (notated with subscript *m*, *c* and *l* respectively) to test whether the perceptual diffuseness can be improved by selecting an optimal ISLD.

The layouts were chosen based on figure 4 to be the stimuli with the most extreme ratio of head-height to non-head-height loudspeakers where the three ISLD options were not close to equivalent. These stimuli were *0/12/1*, *0/12/4*, *12/4/13* and *12/6/13*. These loudspeaker arrangements had the greatest range of ISLD.

These four layouts with three ISLD options gave 12 independent stimuli. A MUSHRA-style listening test was conducted testing 44 different stimuli and is explained in detail in reference [6]. Of these 44 stimuli, 14 relevant stimuli are described here (the 12 stimuli plus *12/6/13*$_c$ and *12/6/13*$_m$ in an off-centre listening position 80cm to the right of centre). Each of the 16 listeners rated the perceived diffuseness (same definition as for the adjustment task) for all 44 stimuli three times.

3.2 Post-screening

Three listeners were removed on the basis of inconsistency between repeats of the same stimulus using the standard deviation between repeats of the same stimulus as a metric for consistency. The three repeats of the same stimulus were then averaged and a further two listeners with three or more outliers over all 44 stimuli were also removed leaving 11 consistent, congruent listeners.

3.3 Results and Discussion

The mean perceived diffuseness ratings are shown in figure 5. They show the $ISLD_{mean}$ to be rated slightly more perceptually diffuse than the other ISLD with the exception of *12/4/13* where the perceived diffuseness for all three ISLD is similar. *0/12/1*$_L$ is notably less perceptually diffuse than the other stimuli. Repeated measures ANOVA reveals significant differences between the *ISLD* ($F(2,9)=32.517$, $p<0.0005$) and significant differences between the loudspeaker layouts ($F(3,8)=19.282$, $p=0.001$ respectively) as well as a significant interaction between them ($p=0.005$). Pairwise comparison (table 4) shows *ISLD=0* is significantly less perceptually diffuse than the other level distributions.

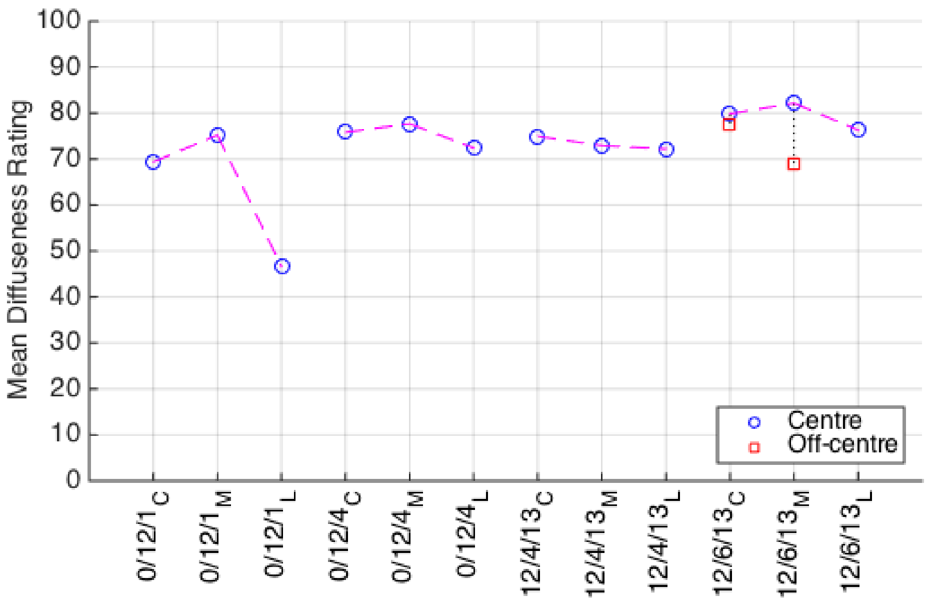

Figure 5. Mean perceived diffuseness of stimuli that assess the effect of *ISLD*. Identical layouts are connected by dashed magenta lines. Only *12/6/13*$_c$ and *12/6/13*$_m$ were evaluated off-centre.

	$ISLD{=}ISLD_{Mean}$	$ISLD{=}0$
$ISCLD=0$	**1.000**	0.001
$ISLD{=}ISLD_{Mean}$		0.000

Table 4. Significance values from pairwise comparisons between different *ISLD*. Significance has Bonferroni adjustment for multiple comparisons.

When moving off-centre, $12/6/13_m$ is no longer more perceptually diffuse than $12/6/13_c$. This is likely due to the level of the head-height loudspeakers when moving off-centre. As the listener moves off-centre, the head-height loudspeaker nearest the listener gets louder, easier to localize, and therefore the perceived diffuseness decreases. In $12/6/13_m$, the head-height layer is louder (-3.6dB of gain) than in $12/6/13_c$ (-7.1dB of gain). This makes moving off-centre more of an important factor in $12/6/13_m$. However, this is unlikely to happen in the *0/12/1* and *0/12/4* layouts where the opposite occurs, as the head-height loudspeakers are at a lower level in the $ISLD_{mean}$ case than in the *ISCLD=0* case.

4 CONCLUSION

In conclusion, the ISLD affects the perceived diffuseness and optimising the ISLD can increase perceived diffuseness. However, these improvements are only small and not statistically significant. Maintaining equal loudness from each individual loudspeaker is rated similarly to the optimised level distribution. From informal conversation with the listeners, the ISLD seems to make most difference when there is a large difference in the number of head-height and non-head-height loudspeakers (when it is difficult to detect the subset with fewer loudspeakers over the louder subset with many loudspeakers). But also, the ISLD makes most difference when the total number of loudspeakers is small (with diminishing returns with highly diffuse layouts). However, these two conditions are mutually exclusive, in order to have a high ratio of number of loudspeakers, one of the subsets must have many loudspeakers leading to minimal absolute improvements when optimising the ISLD.

Nonetheless, because the ISLD can vary over a range of values without affecting the perceived diffuseness adversely for a centred listening position, the ISLD could be optimised for other factors e.g. moving off-centre, without degrading the perceived diffuseness on-centre.

5 AKNOWLEDGEMENTS

This research has been partially funded by the Engineering and Physical Sciences Research Council and by the BBC R&D. The authors would also like to thank Andrew Mason for his additional guidance on the project.

6 REFERENCES

1. D. Nyberg and J. Berg, 'Listener envelopment: What has been done and what future research is needed?,' Presented at AES 124th Convention. Amsterdam (May 2008).
2. F. Rumsey, 'Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm,' J.Audio.Eng.Soc. 50(9) 651-666, (2002)
3. G. Paine, R. Sazdov and K. Stevens, 'Perceptual investigation into envelopment, spatial clarity and engulfment in reproduced multichannel audio,' Presented at AES 31st International Conference. London (2007).
4. J. Berg, 'The contrasting and conflicting definitions of envelopment,' Presented at AES 126th Convention. Munich (2009).
5. 'Rec. ITU-R BS.1534-2:2014 Method for the Subjective Assessment of Intermediate Quality level of Audio Systems.' Int. Telecommun. Union, Geneva, Switzerland. (2014)
6. M. Cousins, F.M. Fazi, S. Bleeck and F. Melchior, 'Subjective diffuseness in layer-based loudspeaker systems with height. ' Presented at AES 139th Convention. New York (2015).