Buenos Aires – 5 to 9 September 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Isotropy and Diffuseness in Room Acoustics: Paper 703

Relation between acoustic measurements and the perceived diffuseness of a synthesised sound field

Michael P. Cousins[a], **Stefan Bleeck**[a], **Frank Melchior**[b], **Filippo Maria Fazi**[a]

[a]Institute of Sound and Vibration Research, Southampton University, UK,
mpc1r13@soton.ac.uk
[b]BBC R&D, UK

Abstract:

This paper describes an investigation of different objective metrics for predicting the perceived diffuseness of reproduced sound and why common metrics, such as Interaural Cross-Correlation Coefficient (IACC), sound pressure level uniformity, and the diffuseness calculation used in Directional Audio Coding (DirAC), may be less appropriate for analysing the perceived diffuseness of a reproduced field than they are for architectural acoustics applications. A listening test was conducted to elicit the perceived diffuseness of sound fields of uncorrelated pink noise signals replayed over 19 different loudspeaker arrangements. Listeners rated how diffuse they perceived each stimulus. A range of different measurements of the sound field were then compared to the subjective test results. The data show that objective metrics do not always correlate well with the perceived diffuseness, especially for specific loudspeaker arrangements. Possible explanations of these results are discussed.

Keywords: Diffuseness, Perception, Loudspeaker Reproduction

Relation between acoustic measurements and the perceived diffuseness of a synthesised sound field

1 Introduction

Diffuse is a term often used when describing late reflections in reverberant spaces, reverberation rooms and concert halls. A sound field is truly diffuse when it is isotropic and homogeneous, as would be generated by an infinite number of uncorrelated noise sources far from the sampling position [1]. This is a useful property for making measurements of absorption, microphone directivity index or total power output of sound sources [2]. In a reverberation room, isotropy and homogeneity is achieved by having a long reverberation time and randomising the phase and direction of arrival of the late reflections [3]. If the reflections are all derived from the same source they are not truly uncorrelated so the reflections add coherently and an interference pattern is created [3]. However, the length of the reverberation time and the randomisation of the phase means that the reflections are sufficiently random to add as if incoherent. This generates a sound field with properties sufficiently similar to the theoretical diffuse sound field for the purpose of the measurements.

In concert hall acoustics and reproduced sound, the sound field is not used for measurement and so the advantages of diffuse sound are the improvements in Apparent Source Width (ASW) and Listener EnVelopment (LEV) [4]. In reproduced audio, the sound field is composed of a small, finite number of truly uncorrelated components. Although the number of loudspeakers is low in comparison to the number of reflections in a concert hall, the sources can be completely uncorrelated and so always add incoherently. With many loudspeakers surrounding the listener the sound field becomes more isotropic and more homogeneous approximating a diffuse sound field. Unfortunately this approximately diffuse sound field is difficult to measure as a sound field is either diffuse (both homogeneous and isotropic) or not diffuse (either not homogeneous and/or not isotropic) [1]. The most useful objective metric of diffuseness in these cases is one that correlates well with the desirable spatial attributes, ASW and LEV. In reproduced sound this would allow for easy evaluation of different arrangements of loudspeakers, compression algorithms and microphone techniques. However, because the sound field is generated in a different way in reproduced sound than in architectural acoustics, existing metrics are not always appropriate.

In this paper, a range of uncorrelated pink noise stimuli described in section 2 were chosen to compare the objective measured diffuseness of the sound field with the subjective perceived diffuseness. The perceived diffuseness was elicited from a group of listeners using a listening test (section 3). These subjective ratings were then compared to; the diffuseness estimation from Directional Audio Coding (DirAC) [5]; the standard deviation of the Sound Pressure Level (SPL); and the InterAural Cross-correlation Coefficient (IACC) [6, 7] in section 4. These metrics are measurements of the isotropy, homogeneity, and the correlation between points respectively and are used to demonstrate in which cases biases can arise when attempting to predict the perceived diffuseness from simple sound field descriptors.

2 Stimuli

A truly diffuse sound field synthesised using loudspeakers would require an infinite number of loudspeakers from all directions and far from the listening position. In reality there are a finite number of loudspeakers placed at specific locations with specific gains. In order to thoroughly evaluate any objective measure of the sound field it was important to test a wide range of synthesised sound fields. The stimuli were uncorrelated pink noise signals and the following factors were independently varied; the number of loudspeakers at head-height; the number of loudspeakers not at head-height; placing a horizontal 2D layer of loudspeaker above, below or both above and below head-height; placing the loudspeakers evenly in azimuth around the listener or in common standardised arrangement [8]; and the relative SPL generated by the subset of loudspeakers at head-height and the subset of loudspeakers not at head-height.

Table 1: **Active loudspeakers for all stimuli and choice(s) of relative level between the head-height and non-head-height loudspeakers. *high and low hidden anchors.**

	Azimuth	Elevation	Stereo*	5.0	9.0	9.0b	22.0	0/4/0	0/6/0	0/8/0	0/12/0	0/6/1	0/6/4	0/6/8	8/0/0	0/0/8	0/12/1	0/12/4	12/4/13	12/6/13	12/12/13*
4 × Floor	±45°	-56°																	✓	✓	✓
	±135°	-56°																	✓	✓	✓
8 × Lower	0°	-20°					✓								✓				✓	✓	✓
	±45°	-17°					✓								✓				✓	✓	✓
	±90°	-24°													✓				✓	✓	✓
	±135°	-17°													✓				✓	✓	✓
	180°	-20°													✓				✓	✓	✓
12 × Head-height	0°	0°		✓	✓	✓	✓			✓	✓						✓	✓			✓
	±30°	0°	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			✓	✓	✓	✓	✓
	±60°	0°					✓			✓	✓						✓	✓			✓
	±90°	0°					✓		✓	✓	✓	✓	✓	✓			✓	✓		✓	✓
	±120°	0 °		✓	✓	✓	✓			✓	✓						✓	✓			✓
	±150°	0°						✓	✓		✓	✓	✓	✓			✓	✓	✓	✓	✓
	180°	0°					✓			✓	✓						✓	✓			✓
10 × Upper	0°	27					✓							✓		✓			✓	✓	✓
	±30°	26°			✓																
	±45	24°					✓						✓	✓		✓			✓	✓	✓
	±90°	32°					✓							✓		✓			✓	✓	✓
	±135°	24°			✓		✓						✓	✓		✓			✓	✓	✓
	180°	27°					✓							✓		✓			✓	✓	✓
5 × Ceiling	±45°	52°				✓												✓	✓	✓	✓
	±135°	52°				✓												✓	✓	✓	✓
	0°	90°					✓					✓					✓		✓	✓	✓
Head-height and non-head-height subsets generate same SPL (A)																	✓	✓	✓	✓	
SPL difference between head-height and non-head-height subsets is optimised (B)												✓	✓	✓			✓	✓	✓	✓	✓
All loudspeakers reproduce the same SPL at the the listening position (C)			✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	

The gains for each loudspeaker in the 19 loudspeaker layouts was determined by one or more of the three following options; the combined SPL of all the loudspeakers at head-height is the same as the combined SPL of all the loudspeakers not at head-height (label A); the relative level of the loudspeakers at head-height and those not at head-height is set to optimise the perceived diffuseness based on the results of a subjective adjustment task documented in [9]; or every individual loudspeaker generates the same SPL at the listening position (label C).

On and off-centre listening positions were also investigated however in this paper we will only present the results of the on-centre listening position.

The 27 stimuli (table 1) are labelled either as standard layouts (5.0, 9.0, 22.0), or in the form $m_b/n/m_a$ where m_b is the number of loudspeakers below head-height, n is the number of loudspeakers at head-height and m_a is the number of loudspeakers above head-height.

3 Listening test

The listening test used a MUltiple Stimulus with Hidden Reference and Anchor (MUSHRA) [10] style listening test to elicit the perceived diffuseness for all the stimuli. A user interface designed in Max 6.0 allowed listeners to rate how diffuse they perceived each of the 7 stimuli per trial. Listeners were given the following description of diffuseness,

> "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."

The Audio Lab at the University of Southampton is 4.80m×3.97m×2.56m (RT$_{60}$ of 0.12s, ±0.02s, 125Hz-8kHz in 1/3 octave bands) and features a wall mounting system for the 39 KEF HTS3001SE loudspeakers. Each loudspeaker was time aligned digitally to $\pm 20\mu$s at the central listening position and the SPL was equalised in 1/6 octave bands (±0.5dB, 95Hz-20kHz) to avoid colouration. All stimuli were replayed at 65dBSPL (A-Weighted) at the central listening position using set of RME D/A converters. The order of the stimuli was randomised and both $12/12/13_B$ and stereo were both included on every page as hidden high and low anchors. There was no explicit high reference. Each stimulus was rated 3 times by 16 postgraduate and undergraduate students at the University of Southampton and a full description of the listening test and results is given in [9, 11].

3.1 Results: Perceived diffuseness

Post screening removed 5 listeners, 3 for high deviation between repeats of the same stimulus and a further 2 listeners for having more than 2 outliers with respect to the rest of the listeners. The 3 repeats for the remaining 11 listeners were averaged and plotted in the boxplot in figure 1 along with the means for each stimulus. The mean ratings are referred to throughout the rest of the paper as the perceived diffuseness for each stimulus.

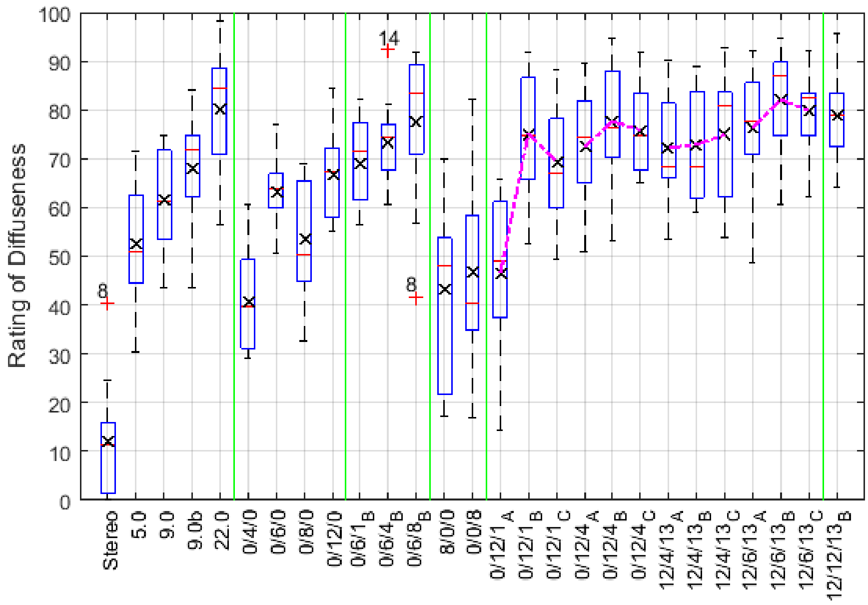


Figure 1: **Boxplot of data for 11 listeners after screening. Means are shown with a cross. Green lines separate stimuli that test different factors. Magenta lines join together stimuli that use the same layout but different relative levels between the head-height and non-head-height loudspeakers.**

4 Objective metrics

This section compares different metrics from the literature that measure the homogeneity, the isotropy or the correlation between spatially separated points with the subjective data.

4.1 Diffuseness estimation from Directional Audio Coding (DirAC): Isotropy

The diffuseness estimation $\psi(t,f)$ used in DirAC, uses B-format signals to estimate the time averaged intensity and time averaged energy density for each time window and frequency bin in the Short Term Fourier Transform (STFT) domain [5]. An A-format Tetramic was used to record the B-format signals. The components of the instantaneous intensity vector are I_X, I_Y and I_Z and can be calculated using,

$$I_X(t,f) = \frac{1}{\sqrt{2}Z_0}\mathrm{Re}\{W^*(t,f)\cdot X(t,f)\}, \tag{1}$$

where $W(t,f)$, and $X(t,f)$ are the Short Time Fourier Transform (STFT) spectra of the respective B-format signals and * denotes the complex conjugate. The expected energy density given by,

$$E(t,f) = \frac{1}{2}\rho_0 Z_0^{-2}\left(|W(t,f)|^2 + \frac{|X(t,f)|^2 + |Y(t,f)|^2 + |Z(t,f)|^2}{2}\right), \tag{2}$$

where ρ_0 is the mean density of air and Z_0 is the acoustic impedance of air. The diffuseness estimation is given by,

$$\psi(t,f) = 1 - \frac{||E\{[I_X(t,f),\, I_Y(t,f),\, I_Z(t,f)]^T\}||}{cE\{E(t,f)\}}, \tag{3}$$

where $E\{.\}$ is the expectation operator and c is the speed of sound. The diffuseness coefficient ranges from 0 to 1 for each frequency bin and time window. Averaging across frequency bins and windows gives a single diffuseness value for each stimulus and these values are plotted in figure 2 versus the perceived diffuseness.

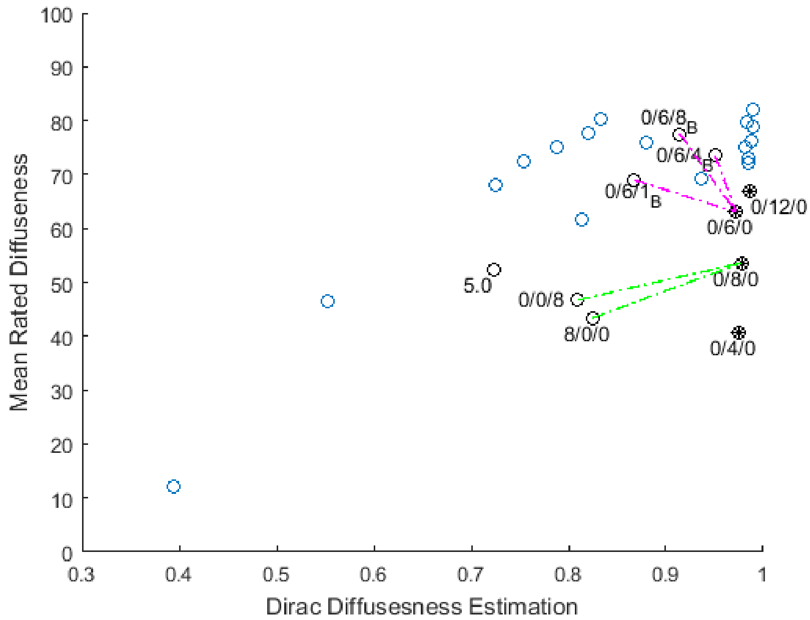


Figure 2: **Mean diffuseness estimation for all stimuli plotted versus the mean perceived diffuseness from the listening test.**

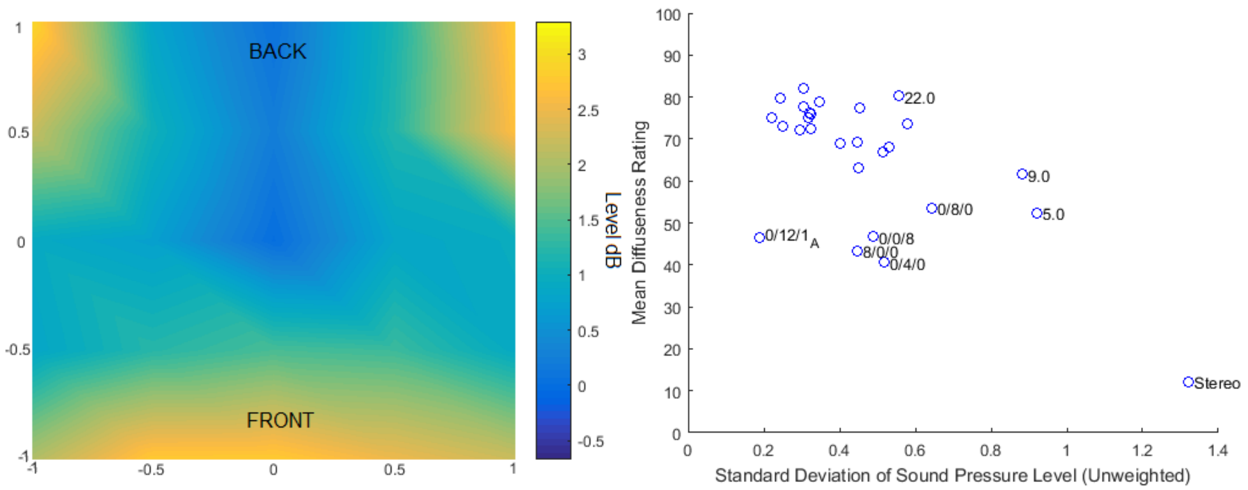
The two dimensional layouts $(0/n/0)$, all have a diffuseness close to 1 because the loudspeakers are positioned opposite the measurement position. With equal energy coming from opposite directions the net flow of energy is zero and this is independent of the number of loudspeakers. The layout $0/4/0$ is predicted as diffuse as $0/12/0$ despite a large perceptual difference.

The magenta lines shows the effect of adding elevated loudspeakers to a 2D layout. There are more loudspeakers and the array is now three dimensional increasing the perceived diffuseness but more energy is coming from above the measurement position leading to some intensity in the Z dimension. This gives a lower diffuseness estimation despite being perceived as more diffuse. Whilst this might imply that some intensity in the Z dimension increases the perceived diffuseness, we also see that layouts such as $0/0/8$ and $8/0/0$, are rated as less diffuse than $0/8/0$ (green lines) despite having a greater absolute intensity vector due to changing the height of the 2D $0/8/0$. When comparing $0/4/0$ to 5.0 we see this same bias but not in the Z dimension. The layout 5.0 has an additional loudspeaker which increases the perceived diffuseness but also increases the intensity as more energy comes from the front than from the back.

These results show the intensity to be related to the uniformity of the arrangement of loudspeakers but this does not seem to correlate with the perceived diffuseness.

4.2 Standard deviation of SPL: Homogeneity

The SPL was measured for each of the stimuli using a calibrated B&K microphone moved across a horizontal 5×5 (2m×2m) grid at head-height centred at the listening position. The SPL at each of the points is shown in figure 3a, for 5.0 with linear interpolation between points to show a colour map of the SPL. The standard deviation of SPL for all stimuli is plotted versus the perceived diffuseness in figure 3b.



(a) **Example levels in dB for 5.0**

(b) **Standard deviation of SPL across entire array.**

Figure 3: **Sound Pressure Level over 5×5 array of mics 2m×2m**

We would expect a negative correlation with a smaller standard deviation in SPL associated with high homogeneity and therefore a more diffuse sound field perceptually. However, several stimuli do not follow this trend. The stimuli $0/0/8$ and $8/0/0$ both do not have any head-height loudspeakers and have a lower standard deviation of SPL than their head-height counterpart ($0/8/0$) despite being rated less diffuse. The stimulus $0/12/1_A$ has a single loudspeaker above the listening position that is at a very loud level relative to any single head-height loudspeaker. In these cases the majority of the measured SPL comes from loudspeakers far from the 2D head-height microphone array. The layout $0/4/0$ also has loudspeakers that are generally further from the array because the array is square whereas the room and loudspeaker array was rectangular. All these stimuli appear overestimated whereas the layouts of stereo, 5.0, 9.0 and 22.0 are all underestimated. These standardised layouts have more loudspeakers concentrated at the front than at the rear increasing the SPL variation despite the perceived diffuseness being seemingly unaffected by the unevenness of the loudspeaker array.

4.3 Inter Aural Cross-correlation Coefficient

The Inter Aural Cross-correlation Coefficient (IACC) is a perceptually motivated measure of the diffuseness. It is a measure of the similarity between the two ear signals and is given by the maximum value of the normalised interaural cross correlation function (equations 4 and 5) [7].

$$IACC_{t1,t2} = max|IACF_{t_1,t_2}| \tag{4}$$

for -1ms$<\tau<$+1ms, where,

$$IACF_{t_1,t_2}(\tau) = \frac{\int_{t_1}^{t_2} p_l.p_r(t+\tau)dt}{\sqrt{\int_{t_1}^{t_2} p_l^2(t)dt \int_{t_1}^{t_2} p_r^2(t)dt}} \tag{5}$$

and p_l and p_r are the impulse responses for the right and left ears respectively.

It is usually computed in octave bands and performed on room impulse responses. Early reflections ($t_1 = 0$, $t_2 = 80$ms) fuse with the direct sound and the late diffuse reflections ($t_1 = 80$ms, $t_2 <$RT60) are perceived as part of the environment. Early and late IACC correlate therefore correlate with apparent source width (ASW) and listener envelopment (LEV) respectively. The IACC$_3$ of Hidaka et al. [6] uses the mean IACC from the middle frequency bands 500Hz, 1kHz and 2kHz to give a single value for the IACC$_3$. In reproduced sound there is no difference in the amount of diffuseness between the early and late reflections and so the IACC can be calculated using steady state full-bandwidth noise dummy head recordings. The measured IACC$_3$ is plotted versus the perceived diffuseness in figure 4.

Interestingly, stereo and 0/4/0 have near identical IACC$_3$ values. The dummy head is static and in both cases the sources lie in the same 2 cones of confusion at 60° and 120° from the interaural axis. However, listeners are able to separate loudspeakers in the same cone of confusion using head rotation. The same can be seen in $0/6/0$ and 5.0 where 5.0 has the loudspeakers in the front at different angles to the loudspeakers in the rear. Whereas $0/6/0$ has loudspeakers symmetrical about the interaural axis and therefore has 4 cones of confusion (at 0°, 60°, 120° and 180° relative to the interaural axis) versus the 5 cones of confusion for 5.0 (30°, 60°, 90°, 120° and 150°).

In studies by other authors where the range of stimuli is smaller, IACC has been found to correlate with the perceived envelopment [12] although the findings of this paper imply that the IACC is not a complete metric of the perceived diffuseness and the ignoring of head rotation is likely to be the cause.

5 Conclusions

Pink noise stimuli were replayed over loudspeakers and different methods of measuring the diffuseness of a sound field were compared to the rated perceived diffuseness. All the methods tested have issues that arise when using them for reproduced sound that may not appear in

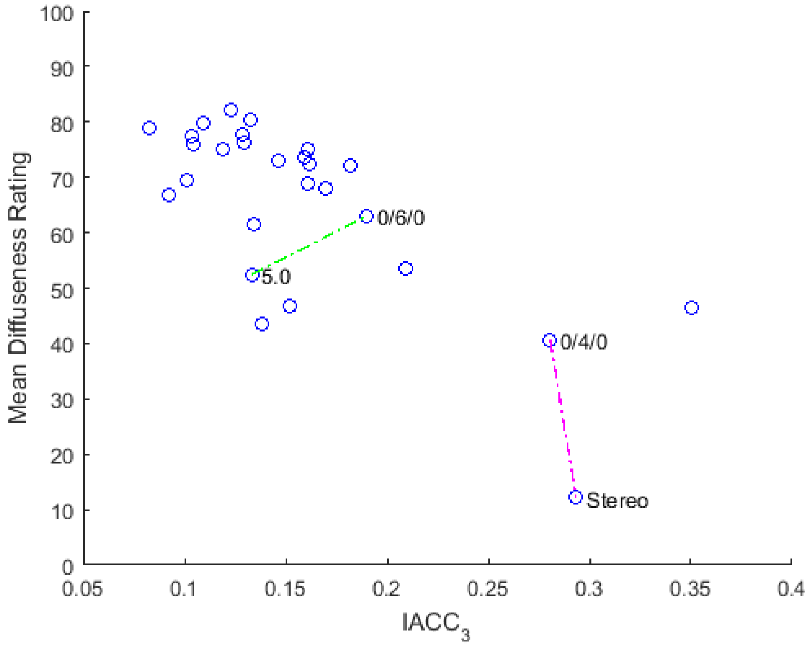


Figure 4: **IACC_3 of Hidaka et al. [6] versus perceived diffuseness**

architectural acoustics where the diffuse field is a result of an effectively continuous distribution of sources rather than a small discrete set of sources. The majority of these stimuli are at the high end of the diffuse scale with 21 of 27 stimuli rated more diffuse than 5.0 so differentiating between these highly diffuse layouts is going to be a difficult task for any metric however, we do find some consistent biases.

The diffuseness estimation from DirAC was used as a test of isotropy. It was found layouts with loudspeakers opposite the measurement position were predicted to be highly diffuse irrespective of the number of sources and that this does not correlate with the subjective data. Similarly, adding elevated sources to a 2D head-height only layout decreased the diffuseness estimation despite always being perceived as more diffuse. The layout 5.0 which has more energy from in front of the measurement position than behind was predicted to be less diffuse than $0/4/0$ despite being perceived more diffuseness.

The SPL uniformity was used as a test of homogeneity. Loudspeakers far from the measurement array have less effect on the SPL than those closer to the array meaning layouts with loudspeakers far from the measurement array were overestimated.

The IACC assumes that loudspeakers placed in the same cone of confusion are equivalent to a single loudspeaker at a higher level. However, in the subjective experiment (where listeners were allowed to rotate their head) they were able to differentiate between these cases. Therefore, arrangements of loudspeakers that have more rotational symmetry about the interaural axis (fewer cones of confusion relative to the number of loudspeakers) are underestimated by the IACC.

Acknowledgements

The author would like to thank Andrew Mason at the BBC for his invaluable assistance throughout.

References

[1] Nélisse, H.; and Nicolas, J. Characterization of a Diffuse Field in a Reverberant Room. The Journal of the Acoustical Society of America, 101(no. 6), 1997, 3517:3524.

[2] Cook, R. K.; et al. Measurement of Correlation Coefficients in Reverberant Sound Fields. The Journal of the Acoustical Society of America, 27(no. 6), 1955, pp 1072-1077.

[3] Jacobsen, F.; and Roisin, T. The Coherence of Reverberant Sound Fields. The Journal of the Acoustical Society of America, 108, 2000, pp 204.

[4] Griesinger, D. Objective Measures of Spaciousness and Envelopment. Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction, Rovaniemi, Finland, 1999.

[5] Ahonen, J.; et al. Directional Analysis of Sound Field with Linear Microphone Array and Applications in Sound Reproduction. Audio Engineering Society: 124th Convention, Amsterdam, Netherlands, 2008. http://www.aes.org/e-lib/online/browse.cfm?elib=14459.

[6] Hidaka, T.; et al. Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls. Journal of the Acoustic Society of America, 98(no. 2), 1995, pp 988-1007

[7] ISO, BS EN ISO 3382-1:2009 Measurement of room acoustic parameters. International Organization for Standardization. 2009.

[8] ITU-R. Rec. ITU-R BS.2051-0(02/2014) Advanced Sound System for Programme Production. Int. Telecommun. Union, Geneva, Switzerland, 2014.

[9] Cousins, M.; et al. Maximising Perceived Diffuseness in Loudspeaker Systems with Height. Institute of Acoustics, Vol. 37 (pt. 4), Moreton-in-Marsh, UK, 2015.

[10] ITU-R. 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.

[11] Cousins, M.; et al. Subjective Diffuseness in Layer-based Loudspeaker Systems with Height. Audio Engineering Society Conference: 139th Convention, New York City, USA, 2015.

[12] Power, P.; et al. Investigation into the Impact of 3D Surround Systems on Envelopment. Audio Engineering Society: 137th Convention. Los Angeles, USA, 2014. http://www.aes.org/e-lib/browse.cfm?conv=137&papernum=9108.