Investigation into the Impact of 3D Surround Systems on Spatial Audio Quality Focussing On Envelopment

P Power The University of Salford, Salford, Greater Manchester, M5 4WT
W.J. Davies The University of Salford, Salford, Greater Manchester, M5 4WT
J Hirst The University of Salford, Salford, Greater Manchester, M5 4WT
C Dunn BBC R&D, Dock House, Media City, Salford, Greater Manchester, M50 2LH

1 ABSTRACT

Current 2D surround systems have benefitted from significant research that has helped identify attributes related to listener preference. Now with high definition visuals, including 3D cinema, there is a need to further develop spatial audio systems. One method would be to add height channels, however it is unknown how this will impact on the listening experience and the spatial audio quality. This paper investigates the spatial capabilities of a number of different loudspeaker configurations including stereo, 2D surround systems, and 3D systems focussing on the subjective attribute of envelopment. An initial computer simulation of a number of 2D and 3D surround systems utilising de-correlated white noise signal and using a measurement closely related to the human perception of spaciousness, IACC (inter-aural cross correlation) was used identifying variations in IACC between systems. The 8 systems tested in the simulation were then used in a subjective experiment again driving each speaker in the 2 and 3D arrays with independent de-correlated noise, where it was found that there were no significant differences between the 2D and 3D surround systems under test in terms of envelopment. A strong negative correlation was found between the objective measure and the subjective scores.

2 INTRODUCTION

Surround sound systems have been common place now for several years, however there is a need to move spatial audio systems forward and look for new ways in which to enhance the listening experience further. Certain surround systems only provide sound in the horizontal dimension at the listener's ears, one way of adding to this experience would be to extend the listening experience using height channels, however it is uncertain how this will impact on spatial audio quality.

One of the objectives of surround sound systems is to portray the sense of being present within a room or environment/space like a concert hall in a musical setting, or room in terms of film postproduction. One of the most important factors in concert hall acoustics is spatial impression [1] and this has two further sub attributes, with these being apparent source width and envelopment [2].

Extensive research over the past few years into spatial attributes in reproduction systems have shown that spatial impression is a highly rated attribute [3], [2], [4], [5], [6], [7], [8] and is strongly related to listener preference therefore it is necessary to investigate the effect future spatial audio systems have on envelopment and if they enhance it significantly better than current standards?

A small number of recent studies have been carried out in this area with a view to the creation of a continuous sound-field to enhance envelopment using height channels. One such study was conducted by Oode et al [9] using de-correlated white noise as the stimulus and varying the vertical spacing of the speakers in regular and irregular intervals and asking subjects to compare the resultant sound-field with that of a reference of 13 loudspeakers. The authors found that the difference grade worsened when the number of loudspeakers in the vertical were reduced to three with larger than 90° spacing in the vertical, So it was concluded that at least four loudspeakers of less than 45° vertical intervals were needed to produce what the authors term spatially uniform sound (a continuous soundfield with no gaps). In terms of an irregular layout, which would probably be found in most consumers homes, the results showed that when no loudspeakers were located directly above the listeners head e.g. +90° a minimum of five speakers were required at an elevation of 60°.

A similar study has been carried out by the national broadcaster in Japan[10] (NHK), who evaluated several different speaker configurations utilising de-correlated noise signals. This was to determine the minimum number of speakers using a height configuration that was needed to give significant improvements over horizontal systems. It was found that a configuration of two rings of 8 speakers, one at ear height and another above the listener at 45° was enough to provide a significant improvement in terms of envelopment. Further, they found that an increase in speakers beyond 12 in the horizontal plane saturated the sensation of envelopment providing no further improvement. Conversely a further increase in speakers in the vertical plane continued to give further improvements. These results differ from the recommendations of the previous author, since the speakers were placed at an elevation of 45° whereas Oode *et al* [9] recommends that when no speaker is placed above the listeners head the elevation of the height channels should be 60° to produce spatially uniform sound.

3 ENVELOPMENT

Parallels can be drawn between concert hall acoustics and spatial audio systems like the common 5.1 system. One of the earliest studies of surround perception was that conducted by Nakayama [3] into the subjective effects of multichannel reproduction using recordings from a large concert hall. One of their findings is that there is evidence of a relationship between fullness of the sound, and a small value of the objective measure the IACC (inter aural cross correlation) In terms of listener preference, this finding could be interpreted as referring to the attribute envelopment. Rumsey states that "*spatial* *factors account for around a third of sound quality ratings in listening tests, and therefore must be considered seriously when determining the sound quality of a system*" [11].

It has also been described by George [12] that envelopment is one of the important attributes that separates multichannel sound from mono and that envelopment is one of the main attributes driving surround sound development [8].Spatial impression was first identified as an important attribute in early studies of concert hall acoustics by Barron [1]. Spatial impression was firstly related to a broadening of the frontal sound scene caused by early reflections, termed apparent source width. However following this listener envelopment was identified by Beranek [13] and was related to late arriving reflections from walls which envelopes the listener, these two terms are collectively referred to as 'spatial impression' and can be applied to multichannel reproduction in a slightly different context where a listener can be enveloped by reproducing direct sound source through all channels surrounding the listener, or by reproducing reverberation from the surround channels [14].

4 SPEAKER ARRAYS

Current surround systems have standardised speaker layouts which are expected to provide optimum spatial performance. However with the development of new 3D systems, there has been only a small amount of work carried out into the perception of these systems. A number of commercial 3D systems have emerged although there has been no scientific evidence of their performance or in what way they enhance the listening experience.

Since there was an infinite number of configurations that could have been subjectively tested and limited resources, it was decided in the first instance that the best way forward would be to simulate a number of different configurations first, to evaluate which would provide the optimum envelopment.

5 INTER-AURAL CROSS CORRELATION

In order to objectively measure each of the soundfields, a measure which has been shown to be closely related to subjective preference [3], [15] called IACC (Inter-aural Cross Correlation) was used [16]. IACC is defined as the maximum value of the normalised cross correlation function for the time interval of ± 1ms. *See Equation1* and *2* where Pl and Pr are the impulse responses at the left and right ears, τ is the offset between ±1ms and is set to account for the maximum inter-aural time difference. t1 and t2 are the time limits of the integration, different variants of IACC exist depending on the time limits. The most general version is IACC for which t1 would be 0 and t2 = inf. This IACC is based on the degree of correlation between the impulse responses at the left and right entrance of the ear. A well defined source has a sharp maximum of 1, whilst subjective diffuseness has a low value <0.15 [17].

$$IACF_{t1t2}(\tau) = \frac{\left[\int_{t1}^{t2} Pl(t) \cdot Pr(t+\tau)\, dt\right]}{\left[\int_{t1}^{t2} Pl^2(t)\, dt \int_{t1}^{t2} Pr^2(t)\, dt\right]^{\frac{1}{2}}}$$ Equation 1

$$IACC_{t1t2} = max\left|IACF_{t1t2}(\tau)\right|, for -1ms < \tau < +1ms$$ Equation 2

A large amount of research has been carried out into the IACC and the measurement is not a perfect solution since it does not take into account the effect of low frequencies due to the wavelength of the signal and the distance between the ears [18] and as a consequence, it has been explained by Blauert and Lindemann [19] that in terms of concert hall acoustics early lateral reflections which contain mainly frequencies up to 3 kHz contribute solely to the expansion in the front back direction, referring to the attribute envelopment. In addition it has also been found by Hidaka *et al* [20] that in terms of the IACC measurement mainly frequencies between 500Hz-2kHz predicted (ASW) apparent source width, and frequencies above 3kHz were not correlated with (ASW).

6 SIMULATION

A computer simulation of a number of different speaker arrays was carried out. In order to simulate diffuse soundfield's a centrally positioned listener was assumed. Independent white noise sources of 6 seconds duration were convolved with HRTF's [21] (Head Related Transfer Functions) of the speaker positions of the different arrays of the surround 2D and 3D configurations, The HRTF database elevations increased in inclinations of 10° therefore the elevation of the height channels was 40° relative to the listeners gaze. To measure the resultant soundfield, the output of each system was summed for the left and right ears and a cross correlation was applied for the duration of the audio signal (no windowing of the signal was applied).

6.1 IACC OF REAL SOUNDFIELD

From the initial simulation 8 systems were identified which provided large differences in IACC measures.

Further to the simulated measurements of IACC, measurements of the real soundfield were carried out for each of the systems. A total of twenty four loudspeakers split between twelve at ear level and twelve above the listening position at an elevation of +35° since the stands used would not permit a

40° elevation used in the previous simulation. This would allow the coverage of the 8 systems. *See Table1*

Array	Number of Channels Horizontal	Number of Channels Vertical
Mono	1	0
Stereo	2	0
ITU 5.1	5	0
Six	6	0
2x6	6	6
9.1	5	4
Twelve	12	0
2xTwelve	12	12

Table 1 Speaker Arrays Used

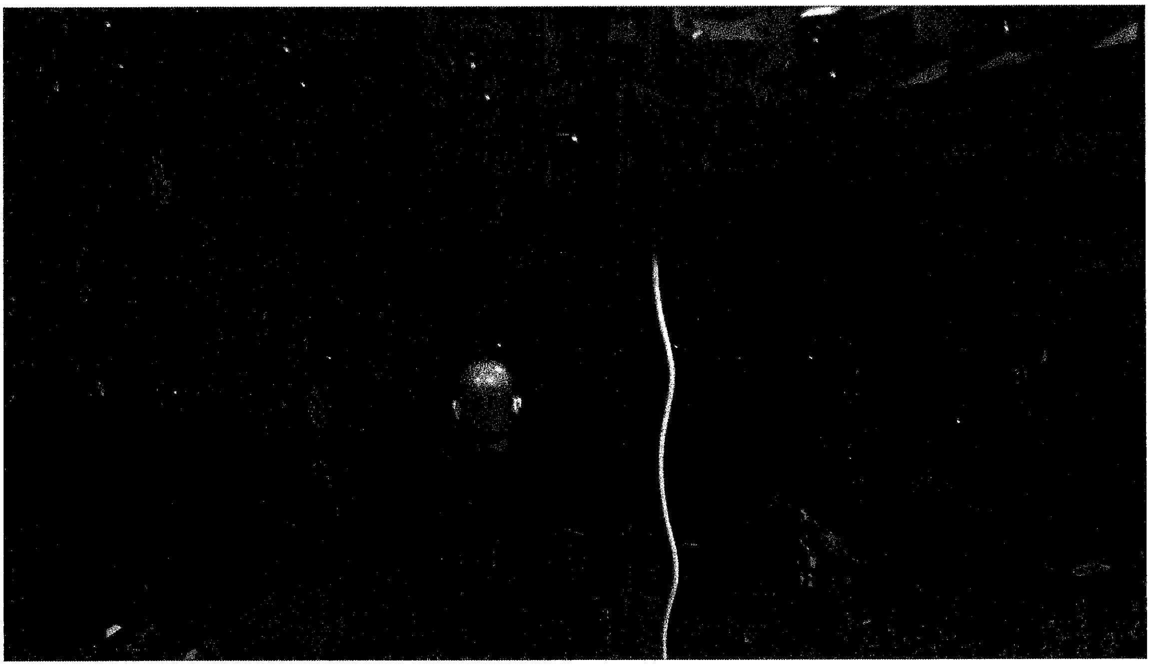

Figure 1 Speaker array used in objective measurement process

The speaker array was set up in a semi anechoic chamber *See Figure 1*, the anechoic chamber conforms to the ISO 3744, ISO 3745, and BS 4196 standards.

Firstly each of the individual speakers were calibrated by driving them with pink noise of -20dB so that they produced 70dB(A) at the listening position. Then the 8 individual systems were level aligned by measuring the output of each system using a sound level meter at the listening position so that they produced 70.1dB(A). Each of the speakers on the horizontal plane were located with a radius of 1.3m from the listening position. However the elevated channels were at a larger distance of 1.7m, resulting in a difference of .4m, therefore this would cause a difference in time of arrival at the listening position when the two layers were used. This was corrected by introducing a delay in the lower channels of 1.2 ms when the two layers played.

Each of the speakers belonging to the arrays were then driven with de-correlated pink noise of 30 seconds duration which was subsequently recorded using a HATS (Head and Torso Simulator) *See Figure 1* this is different than the recommendations of the ISO document[16] where it is recommended for room acoustics to use impulse responses for the measurement, however the IACC has been used by others to measure noise type signals reproduced over loudspeakers [18]. Further, it was intended to use the same signal for the measurement and the subjective test. Finally, the resultant binaural wav was then analysed using the previously detailed IACC measurement.

6.2 RESULTS

A large amount of results were gathered, therefore the focus was directed to 8 systems which featured current commercial offerings so that the expected improvement utilising height systems could be observed. *See Table 2* where it can be seen that as expected the mono system is correlated whilst the 2xTwelve system provides a highly de-correlated soundfield.

System	Linear IACC (Sim)	Linear IACC (Real)
Mono	1	0.9
Stereo	0.39	0.27
ITU-5.1	0.17	0.48
Six	0.37	0.29
9.1	0.12	0.2
Twelve	0.1	0.17
2xSix	0.36	0.3
2xTwelve	0.10	0.19

Table 2 IACC of Simulated and Real Soundfield

The differences between systems are larger than the JND (just noticeable differences) for listeners which is 0.075 ±0.008 as measured by Cox *et al* [22]. This would suggest that utilising these systems in a subjective test should show differences in listener's subjective scores in terms of envelopment. Further, it can also be seen that the JND between the simulated measures and the real measures are slightly different, in the first instance this could be because of the fact that the HATS was used for the real soundfield whilst the KEMAR (Knowles Electronic Mannequin for Acoustic Research) was used for the simulated measures. Further the simulation used was not a highly complex one, therefore only approximating a real listening scenario.

7 SUBJECTIVE TEST

To investigate how listeners will perceive the envelopment between systems a subjective test was carried out. Again this test featured the previously detailed 8 systems and used the 24 speaker array shown in *Figure 1.*

An acoustically transparent curtain was hung between the listener and the speaker array to negate any visual cues, however on entering the chamber participants could see the rear speakers.

Each of the speakers in the arrays were driven with de-correlated pink noise of 30 seconds duration and listeners were asked to rate the perceived envelopment of the various systems using a test interface with eight sliders one for each of the systems, with the extremes of each slider being least enveloping and most enveloping. No numerical markers were provided on each of the test interface scales. Participants scored each system three times and the test was fully randomized for each participant.

7.1 TEST PROCEDURE

In evaluating the attribute envelopment it can be difficult to ensure that each participant is judging the soundfield in the same way and it has been detailed by Berg [23] that numerous researchers have provided different definitions of envelopment, it is therefore necessary to ensure that the listeners understand what they are assessing.

Before commencing the test some training was provided for each participant, this started by asking each participant for their understanding of the attribute envelopment.

Next each participant was provided with a statement describing the attribute:

"Envelopment is a subjective attribute of audio quality that accounts for the enveloping nature of the sound, a sound is said to be enveloping if it wraps around the listener"

As a further stage each listener was given two aural examples of a non enveloping source and an enveloping source. A similar method has been used by George et al [8] who provided listeners with a small training session which features an aural example of an enveloping source and a non enveloping source along with a definition of the attribute. A similar method to George *et al* was followed, A 10 sec section of applause was edited into five 2sec sections then looped, each edit was fed to one of the speakers in the ITU 5.1 array, this provided an enveloping example, with the non enveloping example being one instance fed to the mono centre channel. It was decided to use the ITU 5.1 system over the 3D system in order to observe how the listeners reacted to the 3D system with no prior awareness.

Fourteen listeners from a pool of professional acoustic and audio researchers took part that could be considered 'selected assessors', these are participants who have received training, (as previously described) and have taken subjective tests before [24].

7.2 RESULTS

Participants commented that the majority of the samples had very small differences making it hard to separate them. This was also found by Bech et al [25] in their test listeners could not tell the difference in envelopment between the 5 and 4 channel surround systems. Interestingly a common statement was that for some of the samples it felt like you were standing under a waterfall, referring to the systems with elevated channels, this was probably due to the fact that the listeners had no prior awareness of the 3D system. Comments were also made regarding large spectral differences between the systems, Martin [26] states that an increase in speakers around a listener can cause detrimental effects on timbral characteristics compared to stereo, because of constructive and destructive interference caused by multiple coherent sources with similar amplitudes and the reduced head shadowing. Further, Silzle [27] also points out that comb filtering is much more audible in the vertical plane than the horizontal plane due to the binaural processing.

An intial inspection of the data indicated that as the speaker count increased so did the subjective score, however it can be seen in *Figure 2* that the scores did start to plateau above the 5.1 surround system. Looking closer at the spread of the confidence intervals this does show that the systems with the greatest number of channels e.g 2x12, 9.1 and the twelve on the horizontal did receive consistent scores because of the smaller confidence intervals.

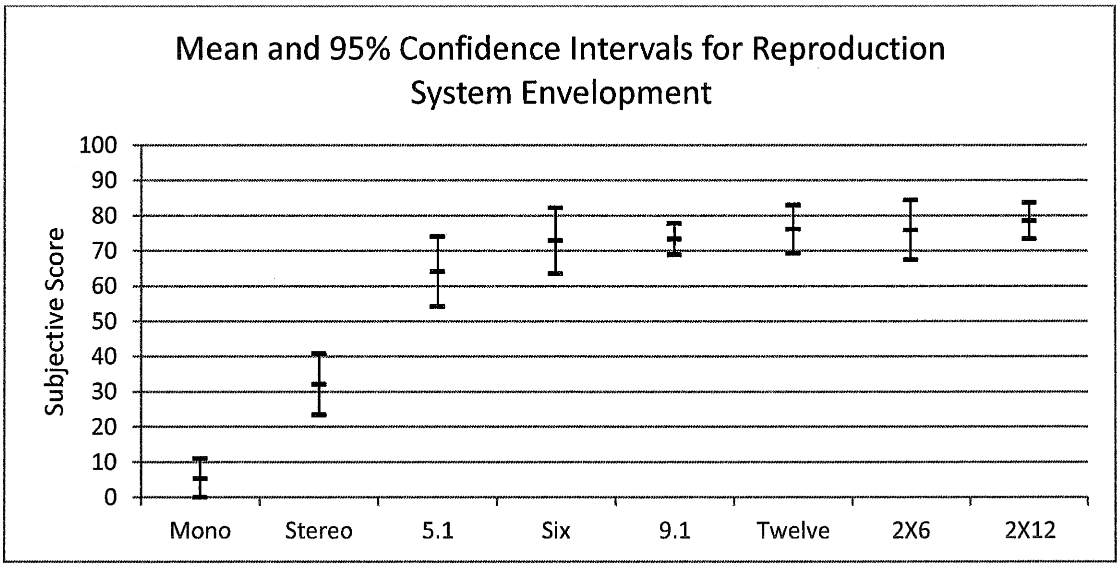

Figure 2 Mean and 95% Confidence Intervals from Subjective Test

Before carrying out an ANOVA to check for significance it was first necessary to check that the data meet the assumptions of the ANOVA, A Mauchly's test of sphericity [28] was carried out and this was found to be not significant indicating the data meet the assumptions of the ANOVA.

A Type III sum of squares was carried out between systems [28], This found significant differences between systems (F=90.8, p=0.000). From the plot of confidence intervals this is suspected to be between the mono and stereo, and surround systems with and without height.

A Bonferoni post hoc test [28] showed there were significant differences between the mono and stereo systems <0.05 significance and also the mono system and the surround systems 2D and 3D <0.05 significance level. However looking exclusively between the 2D and 3D surround systems there were no significant differences >0.05 significance level.

To determine how well the IACC predicted the subjective perception of envelopment the linear IACC was plotted against the averaged subjective scores for each system. It can be observed that the as the degree of correlation goes down the subjective score increases, which is also related to the number and positioning of the speakers in each. Further, it can be observed that the trend displayed plateaus around the systems with the greatest number of channels in the horizontal and with height, which are highlighted *See Figure 3.*

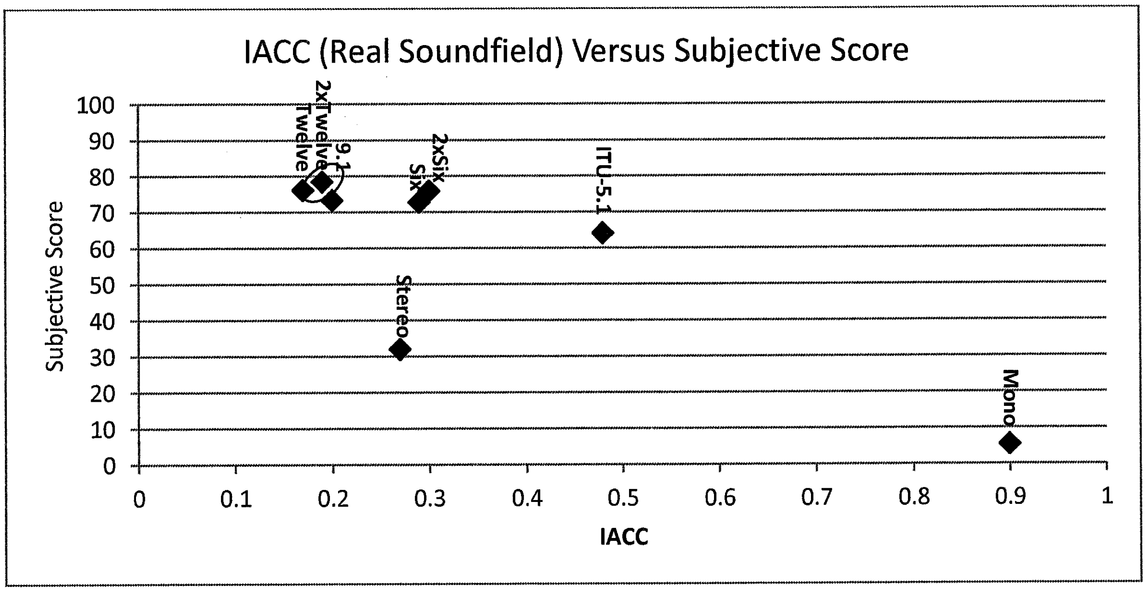

Figure 3 Scatter Plot of Subjective Score Vs IACC Measurement of Real Soundfield

However there does seem to be the outlier which was the stereo system, this was judged as not being enveloping, however the low IACC could be pointing to the fact that it had a broad soundstage or ASW (apparent source width). This also highlights the fact that participants were focussing on envelopment and not confusing ASW with envelopment.

In order to establish the strength of the relationship between the IACC and the subjective scores a Pearsons correlation [28] was carried out for the real and simulated measures, As expected this showed a strong negative correlation *See Table 2.* However it can be seen that the simulated soundfield provided a stronger correlation.

Subjective Score Vs Real Soundfield (IACC)	Correlation Coeff -0.81
Subjective Score Vs Simulated Soundfield (IACC)	Correlation Coeff-0.87

Table 3 Result of Pearsons Correlation Between Objective Measures and Subjective Measure for Simulated and Real Soundfields

8 DISCUSSION

From the outset the objective measures of the simulated soundfield and of the real soundfield showed that three systems, the 9.1, the twelve speakers on the horizontal and the system with twelve speakers on the horizontal and twelve above the listener produced the lowest IACC. Linking this to the subjective test it was observed that these systems received the most consistent scores observing the smaller confidence intervals *See Figure3*. However looking at the mean score for all of the height systems this was similar.

In comparing the 2D surround systems with the 3D systems no significant differences were found and it was interesting to see that in the case of the system with six on the horizontal plane and then adding a further ring of six above the listener the scores were identical, this could also be seen for the twelve speaker array. Contrary to this the 9.1 system which utilises the 5.1 array on the horizontal and an additional four channels above the listener showed a bigger improvement than the system with six on the horizontal and six above the listener pointing to the fact that the actual positioning of the elevated speakers is important to reproduce low correlation between channels, as found by Hiyama *et al* [15] for horizontal layouts.

When considering the training given to listeners the decision was made to use the ITU 5.1 system instead of the 3D system. In retrospect it may have been better to have given listeners training using the 3D system instead, since this is a system which they would not have been familiar with. On the other hand it was interesting to get the listeners reaction to this reproduction format without their prior knowledge, Had training been provided using the 3D system this might have influenced the results.

Taking into account the limited aspects of this investigation, which include only the assessment of envelopment, the limited elevation of the height channels and the small number of systems. Overall in this investigation the improvements height brings over horizontal in terms of the current 5.1 system the improvements are small and more importantly not significantly different.

Some of the systems tested in this evaluation feature large amounts of speakers, therefore it is necessary to appreciate that if wider adoption of 3D systems is to happen a smaller number of speakers is more desirable, especially in the case of domestic use. Therefore in light of this investigation it may be possible when using diffuse sound fields to use a small amount of speakers above a listener. Although, because the minimum number of elevated channels in this test was four it's hard to say where the threshold is before the reproduction is affected.

9 Conclusion

A number of horizontal and with height surround systems have been objectively and subjectively measured focussing on the attribute envelopment. The objective measures of the simulated and real soundfield found large variations in IACC and also identified three systems with similar measures, these being the systems with twelve on the horizontal plane, the system with twelve on the horizontal plane and twelve above the listener and the 9.1system. The subjective test found that these systems were scored the most consistently, however overall no significant differences were found between the 2D surround systems and the 3D systems.

10 Further Work

It is open to question how the use of more realistic sound source material would be perceived and also the evaluation of other spatial sound quality attributes in with height reproduction, this will be the subject of further work.

11 REFERENCES

1. Barron, M., *The subjective effects of first reflections in concert halls—the need for lateral reflections.* Journal of Sound and Vibration, 1971. **15**(4): p. 475-494.

2. Morimoto, M., *The role of rear loudspeakers in spatial impression*, in *Audio Engineering Society Convention 103*. 1997.

3. Nakayama, T., et al., *Subjective assessment of multichannel reproduction.* Journal of the Audio Engineering Society, 1971. **19**(9): p. 744-751.

4. Rumsey, F., *Subjective assessment of the spatial attributes of reproduced sound*, in *Proceedings of the Audio Engineering Society 15th International Conference: Audio, Acoustics and Small Space*. 1998.

5. Hamasaki, K. and K. Hiyama, *Reproducing spatial impression with multichannel audio*, in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. 2003.

6. Soulodre, G.A., M.C. Lavoie, and S.G. Norcross, *Objective measures of listener envelopment in multichannel surround systems.* Journal of the Audio Engineering Society, 2003. **51**(9): p. 826-840.

7. Hirst, J., *Spatial Impression in Multichannel Surround Sound Systems*, in *Acoustic Research Centre*. 2006, University of Salford

8. George, S., et al., *Development and Validation of an Unintrusive Model for Predicting the Sensation of Envelopment Arising from Surround Sound Recordings.* Journal of the Audio Engineering Society, 2010. **58**(12): p. 1013-1031.

9. Oode, S., et al., *Vertical Loudspeaker Arrangement for Reproducing Spatially Uniform Sound*, in *Audio Engineering Society 131st Convention*. 2011.

10. ITU Document 6C/85-E, Sensation of listener's envelopment and localization of phantom sound images on loudspeaker configurations with height channels, contribution from Japan 2012 Oct.

11. Rumsey, F., *Spatial audio and sensory evaluation techniques-context, history and aims*, in *IOSR Workshop* 2006.

12. George, S., *Objective models for predicting selected multichannel audio quality attributes*, in *Department of Music and Sound Recording School of Arts, Communication and Humanities* 2009, University of Surrey

13. Beranek, L.L., A.S.o. America, and A.I.o. Physics, *Concert and opera halls: how they sound*. 1996: Published for the Acoustical Society of America through the American Institute of Physics.

14. Conetta, R., *Towards the automatic assessment of spatial quality in the reproduced sound environment*, in *Institute of Sound recording* 2011, University of Surrey.

15. Hiyama, K., S. Komiyama, and K. Hamasaki, *The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field*, in *113th Audio Engineering Society Convention* 2002.

16. ISO, *Acoustics:Measurement of the reverberation time of rooms with reference to other acoustical parameters*, in *BS EN ISO 3382:2000*. 2000.

17. Ando, Y., *Concert Hall Acoustics* 1985, Berlin: Springer-Verlag.

18. Tohyama, M. and A. Suzuki, *Interaural cross-correlation coefficients in stereo-reproduced sound fields*. The Journal of the Acoustical Society of America, 1989. **85**: p. 780.

19. Blauert, J. and W. Lindemann, *Auditory spaciousness: Some further psychoacoustic analyses*. The Journal of the Acoustical Society of America, 1986. **80**: p. 533.

20. Hidaka, T., L.L. Beranek, and T. Okano, *Interaural cross-correlation, lateral fraction, and low-and high-frequency sound levels as measures of acoustical quality in concert halls*. The Journal of the Acoustical Society of America, 1995. **98**: p. 988.

21. Gardener, B. and K. Martin. *HRTF Measurement of a KEMAR Dummy-Head Microphone*. 1997; Available from: http://sound.media.mit.edu/ resources/ KEMAR.

22. Cox, T., W. Davies, and Y. Lam, *The sensitivity of listeners to early sound field changes in auditoria*. Acta Acustica united with Acustica, 1993. **79**(1): p. 27-41.

23. Berg, J., *The contrasting and conflicting definitions of envelopment*, in *126th Audio Engineering Society Convention* 2009.

24. Bech, S. and N. Zacharov, *Perceptual Audio Evaluation*. 2006, London: John Wiley&Sons Ltd.

25. Bech, S., et al. *Evaluating the Sensation of Envelopment Arising from 5-Channel Surround Sound Recordings*. in *124th Audio Engineering Society Convention* 2008.

26. Martin, G. *Interchannel interference at the listening position in a five-channel loudspeaker configuration*. in *113th Audio Engineering Society Convention* 2002.

27. Silzle, A., H. Strauss, and P. Novo. *IKA-SIM: A system to generate auditory virtual environments.* in *116th Audio Engineering Society Convention* 2004: Audio Engineering Society.

28. Field, A.P., *Discovering statistics using SPSS*. 2009: Sage Publications Limited.