HERE, THERE AND EVERYWHERE:
THE EFFECTS OF MULTICHANNEL AUDIO ON PRESENCE

J. Freeman & J. Lessiter

Psychology Department
Goldsmiths College
University of London
New Cross
London SE14 6NW
UK
J.Freeman@gold.ac.uk
J.Lessiter@gold.ac.uk

ABSTRACT

In this paper 'presence' (a sense of 'being there' in a mediated environment) is proposed as a global metric with which to evaluate audio displays that are part of advanced multi-modal media systems. An evaluation of different audio mixes using a self-report measure of presence, the ITC-Sense of Presence Inventory, and standard audio/visual quality evaluations is described. The results indicate that ratings of presence and audio/visual quality are enhanced by either the addition of bass or increase in volume. However, none of the ratings (with the exception of audio-related enjoyment) were elevated by increasing the number of audio channels. Our study demonstrates that presence is a useful quality metric that can be used in conjunction with more conventional audio measures to evaluate and optimize audio displays.

1. INTRODUCTION

Sound accurately reproduced to provide a faithful spatial representation of sounds in the real world increases the naturalness, realism, and richness of a mediated environment. Further, spatial audio in advanced media systems is essential for achieving consistency with high quality surrounding visual stimuli.

According to Reeves and Nass [1] degraded mediated *visual* imagery is tolerated to some extent, partly because in the real world most of our visual field is peripheral, which is of less high fidelity that our foveal vision. On the other hand, they argue that degraded audio fidelity is a far less familiar experience in the real world (e.g., consider the 'hiss' on an audio tape).

Ratings of specific audio properties (e.g., fullness, clarity, spaciousness) of a mediated presentation clearly have utility in the optimisation of the audio elements of a display system. However, a global measure of a user's quality of experience is likely to be useful for multi-modal display system optimisation. A candidate measure in this regard is 'presence', a user's sense of 'being there' in a scene depicted by a medium [2]. Presence involves feeling physically surrounded by a mediated, but seemingly natural and believable, space to the exclusion of 'real world' sensations. Theoretically, maximal presence requires a fully multi-modal experience as presence increases with the extent of sensory information presented [3]. Manipulations of isolated physical components (such as extent of surround audio) can affect presence. Furthermore, manipulations of audio attributes have been demonstrated to influence perceived visual properties [1], thus affecting the more global experience.

Presence appears to offer particular utility as a global media quality metric because people report finding displays that elicit high presence enjoyable and entertaining [4,5]. Our research team is investigating psychological aspects of immersive television – TV that makes the viewer feel as though they are present at a live event. Our team's goal is to optimise cost-effective novel broadcast services. Given that there is limited bandwidth in which to transmit multi-modal information, a measure that helps identify system requirements affords great utility. To identify optimal configurations of the immersive TV system we evaluate different system components using our questionnaire, the ITC-Sense of Presence Inventory [6].

1.1. Research goals

This paper describes one of our audio evaluation studies that aimed to explore the influence of: (a) number of discrete audio channels, (b) bass, and (c) volume on presence as measured by the ITC-Sense of Presence Inventory. In addition, the influence of audio manipulations on ratings of visual qualities were also examined.

1.2. Hypotheses

1. Five channel presentations will provide a more accurate spatial representation than two channel presentations and will receive higher presence ratings and more favourable audio quality evaluations.
2. Audio presentations that include a bass output provide an additional source of sensory stimulation, offering some vibration to the (rally car) stimulus. Presentations with bass will therefore receive higher ratings than those that do not include a bass output on presence and audio quality evaluations.
3. Audio manipulations will enhance evaluations of the visual properties of the presentations.
4. The contribution of bass to the experience and evaluation of the audio/visual properties of the displayed environment will not be solely attributable to an overall increase in volume.

2. METHOD

Thirty participants were exposed to each of five audio mixes: stereo (2.0), stereo with bass (2.1), stereo 'control' matched to the volume of 2.1 ($2.0^{control}$), five channel (5.0) and five channel with bass (5.1) as part of a complete audio/visual mediated experience. The 5.1 channel mix was generated based on the techniques in Movie Production (www.dolby.com). Since the reproduction area was very small (and indeed realistic in replicating the perceived car interior) the mix-down listening arrangements, although following the standard pattern, were set up in the smallest practical area. Constituent items for the mixes were post-sync and recorded in 2 channel stereo form. These included engine effects, gear noise, and the noise of "stones" hitting the base of the car. Additional electronic samples were used for "bumps", as the car drove over dips in the road. From this collection of stereo recordings, the 5.1 channel mix was performed manually with time code assistance. Much of the "stones" and "bump" strains were fed to the rear speakers with 150msec delay, representing the time between front and rear wheels passing a static road object at a speed of 50mph. The 2 channel (stereo) mix was created with a fixed gain Matrix from the six tracks produced for the 5.1 mix. The gain coefficients were chosen manually largely to reproduce the same loudness and balance as the 5.1 mix. The mixes were delivered via speakers positioned at front left and right [for stereo mixes] and, additionally, front centre, rear left and right [for five channel mixes]). The bass output for 2.1 and 5.1 mixes was presented via a sub woofer located behind the seat in the enclosed testing platform. The presentation comprised a rally car video sequence, presented on a 28'' colour TV, with accompanying synchronised audio. Viewing distance was 120cm, rendering a 29 degree visual angle video display. The two 'without bass' audio conditions were matched at 70dB (pink noise) sound pressure level (SPL). The two 'with bass' conditions were matched at 83/84dB SPL. Note that the SPL of the 'with bass' presentations was larger than those without bass. Therefore a $2.0^{control}$ condition was adjusted to match the stereo 'with bass' dB SPL. Trial orders were not *fully* counterbalanced; with five trials there are 120 possible combinations. However, trial orders were counterbalanced insofar as each condition (e.g., 2.0, 5.1) was represented six times at each trial time (i.e., 1, 2, 3, 4, 5) and each condition was followed by each other condition (e.g., 2.0-2.1; 2.0-$2.0^{control}$, 2.0-5.0; 2.0-5.1) approximately equal number of times across the sample.

Following each presentation, participants were required to complete (a) The ITC-Sense of Presence Inventory and (b) The Media Experience Questionnaire.

The ITC-SOPI is a 44-item (rated 1 'strongly disagree' to 5 'strongly agree') presence questionnaire developed by our research group [6]. The questionnaire yields scores on four scales: (i) a sense of being located in a physical space depicted by the media system ('Sense of Physical Space': 19 items), (ii) a sense of involvement with the narrative/content of the mediated environment ('Engagement': 13 items), (iii) a sense of naturalness and believability of the depiction of the environment itself and events within the environment ('Ecological Validity': 5 items), and (iv) 'Negative Effects' from viewing immersive media, such as eye-strain, headache, sickness etc. (6 items).

The Media Experience Questionnaire (MEQ) was constructed by our research group and comprises 18 items. It was developed from a number of dimensions of perceived sound quality identified by Gabrielsson and Lindstrom [7]. Participants rate nine dimensions of their audio experience (excitement, spaciousness/surrounding, full/completeness, clarity, loudness, uncomfortableness of volume, audibility of extraneous sounds, fidelity/quality, and enjoyableness), five dimensions of their visual experience (uncomfortableness, depth/3Dness, excitement, fidelity/quality, and enjoyableness), and the synchronicity of the audio and visuals. Finally, participants are asked to provide an overall rating for the: (i) audio, (ii) visuals, and (iii) presentation as a whole. In general, for each item, a rating of 1 corresponds to 'not at all' or 'very poor' and 7 indicates 'extremely' or 'excellent'.

3. RESULTS

A series of two factor repeated measures ANOVAs were run for each of the dependent measures. There were two within group factors: bass (on or off) and number of channels (2 or 5), corresponding to the conditions 2.0, 2.1, 5.0 and 5.1.

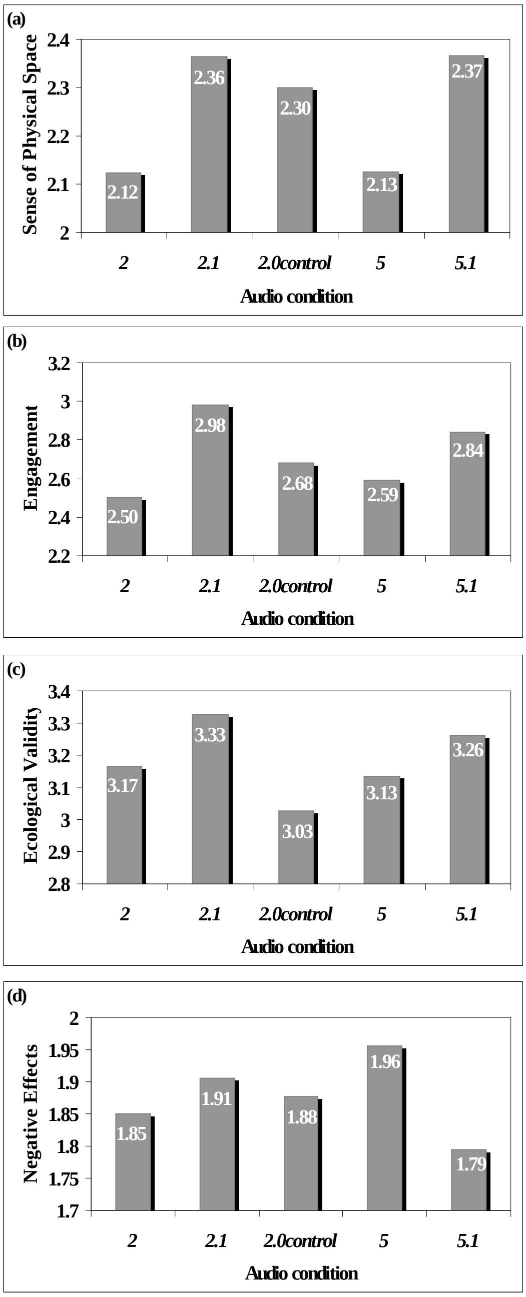

Figure 1. *The effects of bass and channel manipulations on (a) Sense of Physical Space, (b) Engagement, (c) Ecological Validity, and (d) Negative Effects*

In terms of the ITC-SOPI (see Figure 1a-d), overall the results indicated that the inclusion of bass to a presentation significantly enhanced presence-related variables, specifically, Sense of Physical Space ($F_{(1,29)}$ = 11.12, $p < 0.01$) and Engagement ($F_{(1,29)}$ = 16.26, $p < 0.001$). While the naturalness and believability (i.e., Ecological Validity) of the presentation was also enhanced with bass, this just failed to reach significance ($F_{(1,29)}$ = 2.98, p = 0.095). However, as this was predicted the one-tailed probability can be accepted, which is significant (i.e., $p < 0.05$). In terms of the number of audio channels, the results were less positive, and none of these results reached significance. Five channel presentations were only superior to two channel presentations in eliciting higher Sense of Physical Space ratings. For all other presence-related variables, two channel presentations received higher ratings. Finally, there were no bass x channel interaction effects for any of the dependent variables.

A series of paired samples t-tests were run to examine whether the significant effects of bass, documented above, were attributable to an increase in volume or whether bass made a unique contribution to ratings of presence-related variables.

T-test comparisons between the three stereo conditions suggested that bass offered a unique contribution to the media experience in addition to the increased volume that it afforded. For every ITC-SOPI factor, the 2.1 audio condition was rated more highly than the volume-matched 2.0 condition (i.e., 2.0^{control}). This difference was significant for Engagement (t_{29} = 2.11; $p < 0.05$) and Ecological Validity (t_{29} = 2.47; $p < 0.05$). Interestingly, for Sense of Physical Space there was no significant difference between 2.1 and 2.0^{control}, suggesting that increased volume, rather than the inclusion of bass enhanced ratings on this factor. Nevertheless, the 2.1 condition was rated more highly on Sense of Physical Space than 2.0^{control}, suggesting that bass might have offered some (non-significant) enhancement. Further, it is noteworthy that the louder stereo control condition was rated as the least natural and believable of the five audio mixes (see Figure 1c).

In terms of the MEQ ratings the repeated measures ANOVA results suggest that irrespective of the number of audio channels (2 or 5), the inclusion of bass to a presentation significantly enhanced audio-related ratings of excitement ($F_{(1,29)}$ = 17.92, $p < 0.001$), spaciousness ($F_{(1,29)}$ = 9.20, $p < 0.01$), fullness ($F_{(1,29)}$ = 22.62, $p < 0.001$), clarity ($F_{(1,29)}$ = 10.05, $p < 0.005$), loudness ($F_{(1,29)}$ = 28.37, $p < 0.001$), volume-related discomfort ($F_{(1,29)}$ = 11.69, $p < 0.005$), fidelity ($F_{(1,29)}$ = 14.30, $p < 0.005$), enjoyment ($F_{(1,29)}$ = 11.68, $p < 0.005$), and the overall audio rating ($F_{(1,29)}$ = 12.01, $p < 0.005$). Interestingly, presentations with, rather than without bass, also enhanced ratings of the *visual* properties of the presentation, namely, perceived audio/visual synchronicity ($F_{(1,29)}$ = 16.26, $p < 0.001$), excitement ($F_{(1,29)}$ = 4.31, $p < 0.05$), fidelity ($F_{(1,29)}$ = 4.57, $p < 0.05$), and enjoyment ($F_{(1,29)}$ = 13.70, $p < 0.005$). In contrast, five channel presentations (with or without bass) only significantly enhanced audio-related ratings of ‘enjoyment’ compared with two channel presentations ($F_{(1,29)}$ = 11.68, $p < 0.005$). Finally, ratings of volume-related discomfort were significantly more pronounced when bass was added to two channel, rather than five channel presentations ($F_{(1,29)}$ = 4.22, $p < 0.05$).

On the whole, the 2.1 presentation was rated more favourably than either the 2.0 or the (2.1 volume matched) $2.0^{control}$ condition in terms of the majority of the audio and visual evaluations (i.e., audio-visual synchronicity, excitement [audio and visual], spaciousness [audio], depth [visuals], fullness [audio], clarity [audio], fidelity [audio], audibility of extraneous sounds, enjoyment [audio and visual], overall [separate audio and visual and cummulatively]). However, there were no significance differences between the volume matched, 2.1 and $2.0^{control}$, conditions on these variables. These results suggest that the increase in volume that the 2.1 condition affords primarily accounted for the enhancement of audio and visual ratings. However, the means suggest that bass might have offered some (non-significant) enhancement to MEQ ratings which the ITC-SOPI detects.

4. DISCUSSION

Overall, the majority of measures – presence and audio quality evaluations – were rated more highly when the presentation included bass, supporting Hypothesis 2. Furthermore, the results indicate that *irrespective of the increase in volume*, adding bass to the presentation enhanced ITC-SOPI ratings of Engagement and Ecological Validity. Thus, the vibration that bass affords increased the perceived naturalness and enjoyment and interest in the presentation. However, the bass-related increase in ITC-SOPI Sense of Physical Space and MEQ ratings for the 2.1 mix could be largely attributed to the increase in volume. Thus Hypothesis 4 was only partly supported.

Overall, five channel presentations were not rated more highly on presence and audio quality evaluations, contrary to Hypothesis 1. One interpretation is that there was no perceived advantage of 5 channel mixes over 2 channel mixes. In terms of our Immersive TV system, this suggests that it would be more cost-effective to send stereo audio signals only. It is more likely, however, that this result was content dependent. The rally car audio stimulus did not fully capitalise on the subtleties of surround sound. A further study is planned to examine the effects of multi-channel sound on presence using a more appropriate stimulus. A second explanation relates to the fact that the high quality five channel audio was not consistent with the visuals, which did not surround the participant. Mis-matches between audio and visual fidelities might be distracting and reduce the sense of presence. However, in support of Hypothesis 3, manipulation of the audio properties, bass and volume, were demonstrated to have a positive cross-over effect on ratings of the visual properties.

This paper has demonstrated how a combination of measures, both global (presence) and specific (e.g., audio clarity, fullness) can be used in conjunction with one another in the evaluation of different audio configurations. Results from both measures were relatively consistent, although the global presence results were more easily and quickly interpretable. Most importantly, presence provides a subjective index of the ultimate goal of immersive audio systems and for this reason we advocate the use of the ITC-SOPI.

5. REFERENCES

[1] Reeves, B. & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press.

[2] Barfield, W., Zeltzer, D., Sheridan, T.B., & Slater, M. (1995). Presence and performance within virtual environments. In W. Barfield & Furness, T.A. (Eds). Virtual environments and advanced interface design. Oxford: Oxford University Press.

[3] Sheridan, T.B. (1992). Musings on telepresence and virtual presence. Presence: Teleoperators and Virtual Environments, 1, 120-125.

[4] Freeman, J. & Avons, S.E. (2000). Focus Groups Exploration of Presence through Advanced Broadcast Services. Proceedings of the SPIE, Human Vision and Electronic Imaging V, 3959-76, presented at Photonics West - Human Vision and Electronic Imaging, San Jose, CA, 23-28 January 2000.

[5] Lodge, N. (1999). Being part of the fun – Immersive television! Keynote address at the Conference of the Broadcast Engineering Society (India), New Delhi, 2-4 February 1999.

[6] Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (in press). A cross media presence questionnaire: The ITC-Sense of Presence Inventory. Presence: Teleoperators and Virtual Environments (Special Issue).

[7] Gabrielson, A., & Lindstrom, B. (1985). Perceived sound quality of high-fidelity loudspeakers. Journal of the Audio Engineering Society, 33, 33-52.