Really hear? The effects of audio quality on presence

Jane Lessiter & Jonathan Freeman

Psychology Department

Goldsmiths College

University of London

J.Lessiter@gold.ac.uk

4th International Workshop on Presence

21st-23rd May 2001, Philadelphia, USA

1. Abstract

The extent of sensory information presented to a user within a mediated environment has been proposed as a determinant of presence. (e.g., Sheridan, 1992). While a large proportion of presence research has focused on visual manipulations, research on manipulations of auditory characteristics is currently limited. In this paper the effects of several audio manipulations on the sense of presence and ratings of specific audio/visual dimensions were explored. In the first study, a 5.1 'rally car' audio mix was rated significantly more highly than either a mono or stereo audio mix on some presence measures and specific audio dimensions. A second experiment was designed to investigate potential contributory factors. Overall, the increase in number of discrete audio channels from two to five did not significantly enhance presence or audio/visual quality ratings, with the exception of audio-related enjoyment. Overall, the inclusion of bass to a presentation, whether two channels or five, did significantly enhance presence and a number of audio/visual quality ratings. An overall increase in volume in the stereo condition could account for the observed bass-related enhancement for Sense of Physical Space and the audio/visual quality ratings. However, for ratings on Engagement, Ecological Validity and a general presence measure (SUS3), bass offered a unique contribution to the experience irrespective of the increased volume that it afforded.

2. Introduction

According to Sheridan (1992) one of the principal external determinants of presence is the extent of sensory information presented to a user within a mediated environment; that is, the greater the number of sensory inputs provided to different modalities, the greater the sense of presence. To date, the majority of presence research has focused on the visual modality, for instance, the impact of geometric field of view (e.g., Hendrix & Barfield, 1996a), stereoscopic depth cues (e.g. IJsselsteijn, de Ridder, Hamberg, Bouwhuis, & Freeman, 1998; Hendrix & Barfield, 1996a; Freeman *et al.*, 1999; Freeman *et al.*, 2000), display update rate (e.g., Barfield, Baird, & Bjornseth, 1998) and delay of visual feedback (e.g., Welch *et al.*, 1996).

By contrast, the audio modality has received relatively little research attention despite evidence suggesting that audio fidelity has a more pronounced impact than visual fidelity on attention and memory, and that audio has a cross-over effect on visual evaluations (Reeves & Nass, 1996). According to Reeves & Nass (1996) we are less tolerant of degraded audio than we are of visual imagery. We are familiar with viewing the world as blurred (e.g., in fog and rain), with objects obscured, and in partial darkness. However, recorded audio that is of poor fidelity (e.g., audio cassette 'hiss') has few comparable examples in the real world. Good quality audio can enhance the perceived fidelity of the visuals (Reeves & Nass, 1996) and requires comparatively less bandwidth than good quality visual information. However, Beerends and de Caluwe (1999) found that video quality contributes more to perceived audio quality than vice versa. Audio is also important in terms of audio-visual synchronicity. Asynchrony detracts from the plausibility of a mediated experience; it appears less natural and realistic. It distracts the viewer from the content of the experience and directs his/her attention more towards the audio-visual mismatch (Reeves & Nass, 1996), in turn, presumably reducing presence.

Increases in the number of discrete audio channels offers increased accuracy and richness in the spatial representation of an environment rendering it more natural and realistic. These attributes are associated with increases in the sense of presence.

Hendrix and Barfield (1996b) conducted two studies to explore the effect of spatialised sound on the sense of presence. In both, participants navigated through a virtual room replete with tables and chairs. The first study found that spatialised sound (delivered via headphones), compared with no sound, significantly enhanced the sense of presence (as measured by two separate questions using different response scales) but not the sense of realism (as measured by one question). In the second study, spatialised sound was compared with non-spatialised sound. Spatialised sound produced significantly higher presence but, again, not realism ratings. They suggested that the non-significant realism results might be attributed to the fact that the sound in the environment was relatively meaningless (it consisted of a continuous radio broadcast and the sound of monetary exchange in a drinks machine which was repeated every 10 seconds). Further, the visuals were not synchronous with the audio in that a drinks can did not appear in the dispenser, and in fact, no-one appeared to be using the drinks dispenser! They also suggest that participants may have interpreted the ‘realism’ question in visual terms.

It is the aim of the present paper to explore whether multi-channel sound enhances presence and audio/visual quality evaluations. In contrast to the Hendrix and Barfield study, this study will use (a) a photorealistic video stimulus with synchronous accompanying audio, and (b) more detailed presence/realism questionnaires.

3. Study A

Eighteen students and college staff (9 male, 9 female; aged 20-57 years, $\bar{x}$ = 30.8 years, SD = 10.3) were exposed to each of three audio mixes (mono: single channel; stereo: two channels; and 5.1: five channels plus bass) as part of a complete audio/visual ‘rally car’ experience. The mono output was played through both the front left and front right speakers: thus, the only difference between the mono and stereo conditions was the phase differences. The 5.1 mix was delivered via five speakers surrounding the participant at positions: front left, right and centre, rear left and right. The bass speaker was located behind the seat. The video was presented on a 100Hz Phillips 28 inch colour TV. Viewing distance was 120cm, rendering a 29 degree visual angle video display. Trials were fully counterbalanced across the sample. This design was chosen to enable a rigorous evaluation because an earlier investigation of the three audio mixes using an independent groups design (which also included within groups factors) found no significant difference in ratings of presence and specific audio dimensions. The sound levels for the mono and stereo mixes were matched by adjusting the pink noise to 70dB sound pressure level (SPL). For the 5.1 mix, the bass channel was added after the sound level had been adjusted, elevating the loudness by approximately 10dB.

Following each presentation, participants completed a battery of questionnaires: ITC-Sense of Presence Inventory (‘ITC-SOPI’: Lessiter, Freeman, Keogh, & Davidoff, 2001), Slater-Usoh-Steed

(‘SUS6’ rating scale: e.g., Usoh, Catena, Arman, & Slater, 2000) and an Audio Experience Questionnaire (‘AEQ’). The ITC-SOPI is a 44-item presence-related questionnaire measuring four scales (representing the mean of constituent items [scored 1-5]): Sense of Physical Space, Engagement, Ecological Validity, and Negative Effects. The SUS6 is a 6-item scale (scored 1-7) which was summed for the present study. The AEQ is an 11-item audio rating scale (scored 1-7) based on a number of dimensions of perceived sound quality identified by Gabrielsson and Lindstrom (1985) and included some of our own items (audio/visual synchronisation, excitement, spaciousness/surrounding, full/completeness, clarity, loudness, uncomfortableness of volume, audibility of extraneous sounds, fidelity/quality, and enjoyableness, plus one overall rating). Each AEQ item was analysed separately. Finally, at the end of all three presentations, participants were required to state which was their favourite.

3.1 Results

A series of one-factor repeated measures ANOVAs were run. In terms of the ITC-SOPI, there was a significant main effect of audio mix on Engagement ($F_{(2,34)} = 4.71$; $p < 0.05$), but not on Sense of Physical Space, Ecological Validity or Negative Effects (see Figures 1a-d).

There were also significant main effects of audio condition on SUS6 ($F_{(2,34)} = 4.69$; $p < 0.05$: see Figure 1e) and several AEQ items: spaciousness/surrounding ($F_{(2,34)} = 5.11$; $p < 0.05$), loudness ($F_{(2,34)} = 15.73$; $p < 0.001$), discomfort associated with loudness ($F_{(2,34)} = 5.64$; $p < 0.01$), enjoyment ($F_{(2,34)} = 6.83$; $p < 0.01$), and overall audio rating ($F_{(2,34)} = 5.46$; $p < 0.01$).

In each instance where there was a significant main effect of audio, the 5.1 mix was rated significantly more highly than the stereo mix, which in turn did not differ significantly from the mono mix. Furthermore, as predicted the 5.1 mix was nominated as favourite significantly more frequently than the other two mixes (df = 2; $\chi^2 = 5.77$; $p < 0.05$ [alpha adjusted for predicted order]).

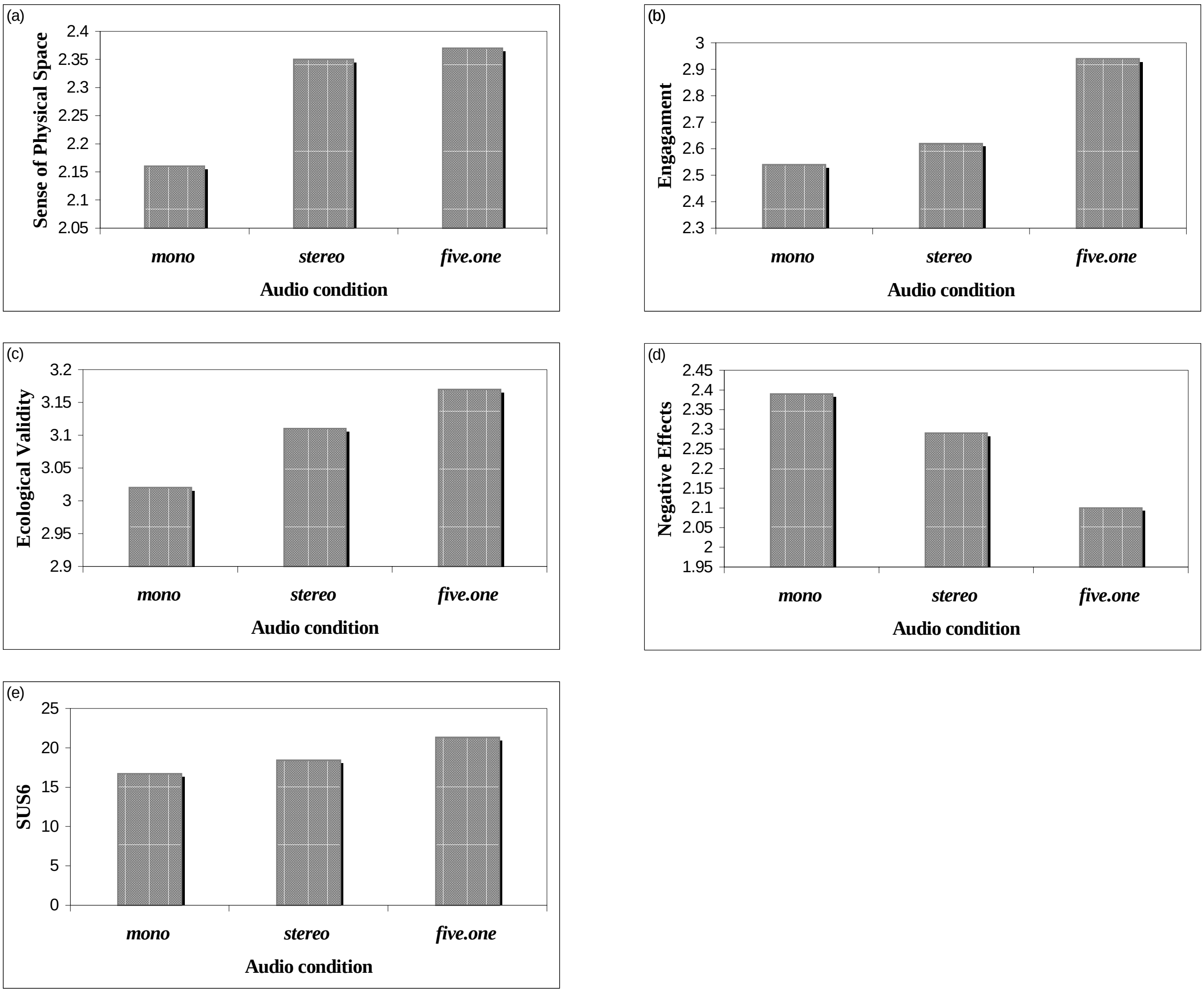


Figure 1. Differential effects of the audio condition (mono, stereo, 5.1) on ratings of (a) ITC-SOPI Sense of Physical Space; (b) ITC-SOPI Engagement; (c) ITC-SOPI Ecological Validity; (d) ITC-SOPI Negative Effects; and (e) SUS6

4. Study B

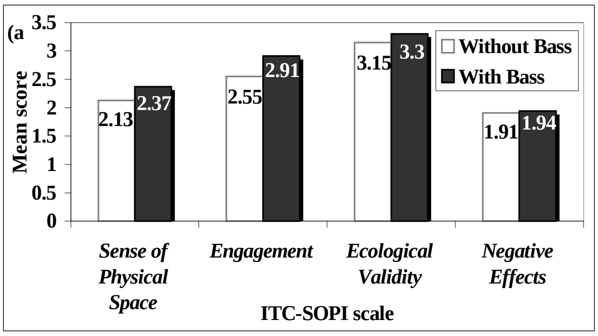
Thirty students (n = 17) or respondents to an advert placed in the local press (n = 13) took part in this study. There were 15 males and 15 females aged between 18 and 44 years ($\bar{x}$ = 28 years, SD = 8.42). The stimulus in this study (i.e. rally car audio and visuals) was identical to that used in Study A but had five conditions/trials for the within groups factor, audio mix. There were two levels of channel: two (i.e., stereo) and five; and two levels of bass (on: 'x.1' and off: 'x.0') yielding four trials (2.0, 2.1, 5.0, 5.1). In addition, there was one 'control' trial for the 2.1 audio mix ('$2.0^{control}$'). This controlled for the effects of volume and consisted of a stereo output (no bass) that matched the volume of the stereo plus bass condition. The 2.1, 5.1, and $2.0^{control}$ mixes were matched at 83/84dB

SPL, and the 2.0 and 5.0 mixes were matched at 70dB SPL. The order of the five presentations across the sample were partially counterbalanced.

Following each presentation, participants were required to complete the ITC-SOPI (as described above), a 3-item SUS rating scale (‘SUS3’) and the Media Experience Questionnaire (‘MEQ’). The MEQ is an 18-item (score 1-7) modified version of the AEQ and includes questions about the visual as well as audio properties of the presentation. New items comprise five visual-related questions (uncomfortableness, depth/3Dness, excitement, fidelity/quality, and enjoyableness), one overall visual rating and one overall audio/visual rating.

4.1 Results

A series of repeated measures ANOVAs with two factors, bass (on/off) and channel (2/5), were run. There were significant main effects of bass on ITC-SOPI ratings of Sense of Physical Space ($F_{(1,29)} = 11.12$, $p < 0.01$: see Figure 2a) and Engagement ($F_{(1,29)} = 16.26$, $p < 0.001$; see Figure 2a), and SUS3 ($F_{(1,29)} = 7.68$, $p < 0.05$: see Figure 2b). The main effect of bass just failed to reach significance for Ecological Validity ($F_{(1,29)} = 2.98$, $p = 0.095$: see Figure 2a), but as this was predicted the one-tailed probability can be accepted, which is significant (i.e., $p < 0.05$). There were no significant main effects of channel, and no significant interactions.



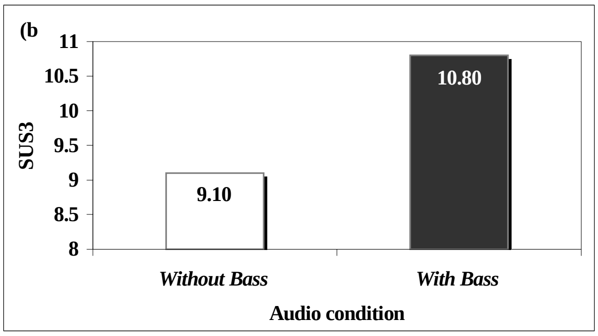


Figure 2. Differential effects of audio bass for two and five channel presentations combined on ratings of (a) ITC-SOPI scales (b) SUS6

To explore the extent to which this bass-related enhancement of presence ratings was attributable to increased volume, paired samples t-tests compared the two volume-matched stereo conditions ($2.0^{control}$ and 2.1). Whilst the 2.1 mix was rated more highly than $2.0^{control}$ on all of the presence measures, this difference was only significant for ITC-SOPI Engagement (see Figure 3a) and Ecological Validity (see Figure 3a), and SUS3 (see Figure 3b). The extra bass provides increased vibration and motion thereby enhancing the naturalness of the rally car experience and enjoyment.

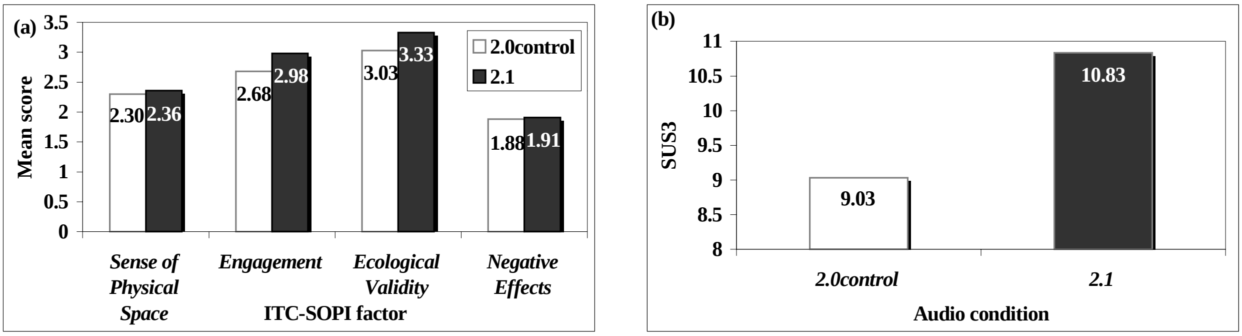


Figure 3. Differential effects of audio bass on stereo ratings of (a) ITC-SOPI scales (b) SUS6

In terms of the MEQ, repeated measures ANOVAs revealed that bass significantly enhanced audio quality ratings of excitement, spaciousness, fullness, clarity, loudness, volume-related discomfort, fidelity, enjoyment, and the overall audio rating. Interestingly, presentations with, rather than without bass, also significantly enhanced ratings of the visual properties of the presentation, namely, perceived audio/visual synchronicity, excitement, fidelity, and enjoyment. In contrast, five channel presentations (with or without bass) only significantly enhanced audio-related ratings of ‘enjoyment’ compared with two channel presentations. Finally, ratings of volume-related discomfort was more pronounced when bass was added to two channel, rather than five channel presentations. Paired samples t-tests revealed that while the 2.1 mix was rated higher than $2.0^{control}$ on the majority of MEQ variables, none of these differences were significant. This suggest that the increase in volume that the 2.1 condition affords primarily accounts for the enhancement of audio and visual ratings.

5. Discussion

Overall the results suggest that the 5.1 audio mix was rated more favourably than either the mono or stereo mix in terms of presence and audio/visual quality ratings. Potential contributory factors: number of channels, bass and volume, were explored. The inclusion of bass (irrespective of the increase in volume) appeared to be responsible for the increase in presence ratings, namely Engagement, Ecological Validity and SUS3. Interestingly, of all the audio conditions, $2.0^{control}$ received the lowest Ecological Validity ratings. However, the increase in ITC-SOPI Sense of Physical Space and MEQ ratings for the 2.1 mix could be attributed to the increase in volume. It is of note that the ITC-SOPI factors were differentially sensitive to the bass/volume manipulation. Sense of Physical Space, which contains items that typify presence, did not yield the same results as did the more simple SUS3 presence rating scale. SUS3, in fact, produced findings comparable with Engagement and Ecological Validity. This raises the issue of whether SUS3 provides an aggregate measure of presence across the three components of a media experience that the ITC-SOPI taps. In effect, a benefit of using the ITC-SOPI is that it provides a more in-depth evaluation of a media experience.

It was not practical to have included a five channel volume control condition, adding a further condition to the already lengthy experimental session. Future studies should seek to explore whether the present stereo findings can be replicated using three five channel conditions (5.0, 5.1 and $5.1^{control}$). However, the appropriateness of the stereo volume control is a further issue. Sounds of different frequencies, but equivalent dB SPLs are experienced as perceptually different in terms of loudness. Thus, although the stereo volume control offered some comparison, it may not be entirely valid, but it is not easy to envisage a suitable control.

Contrary to prediction, the increased accuracy of the spatial representation (i.e. 5 vs. 2 channels) did not significantly affect presence and MEQ ratings (other than audio-related ‘enjoyment’). This is inconsistent with the findings of Hendrix and Barfield (1996b) who found increases in presence (but not realism) with spatialised (vs. non-spatialised) sound. There are several possible explanations for this finding. First, the rally car stimulus may not represent the most appropriate stimulus to capitalise on discrete audio channels surrounding the listener. It consists of engine noises, sounds of stones being flicked against the car, occasional bumps as the car lunges over dips in the road, and sporadic difficult-to-comprehend driver conversation. Overall, this produces a general noisy environment, whether two channels or five. It is planned to further evaluate the effect of channel on presence using a more suitable stimulus. Second, the audio and visual qualities in terms of spaciousness were not congruous. The multi-channel audio was more surrounding than the visual imagery, which was presented on a 28 inch colour TV. This mis-match may detract from presence and reduce direct audio/visual evaluations. This suggests that visual quality may have a stronger impact on audio quality than vice versa, consistent with the findings of Beerands and de Caluwe (1999). Nevertheless, manipulation of the audio properties, bass and volume, were demonstrated to have a positive cross-over effect on ratings of the visual properties.

6. References

Barfield, W., Baird, K.M., & Bjorneseth, O.J. (1998). Presence in virtual environments as a function of type of input device and display update rate. Displays, 19, 91-98.

Beerends, J.G. & de Caluwe, F.E. (1999). The influence of video quality on perceived audio quality and vice versa. Journal of the Audio Engineering Society, 47 (5), 355-362.

Freeman, J., Avons, S.E., Pearson, D., & IJsselsteijn, W. (1999). Effects of sensory information and prior experience on direct subjective ratings of presence. Presence: Teleoperators and Virtual Environments, Vol. 8, 1-13.

Freeman, J., Avons, S.E., Meddis, R., Pearson, D.E., & IJsselsteijn, W. (2000). Using behavioural realism to estimate presence: A study of the utility of postural responses to motion stimuli. Presence: Teleoperators and Virtual Environments, Vol. 9 (2), 149-164.

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

Hendrix, C. & Barfield, W. (1996a). Presence within virtual environments as a function of visual display parameters. Presence: Teleoperators and Virtual Environments, 5, 274-289.

Hendrix, C. & Barfield, W. (1996b). The sense of presence within auditory virtual environments. Presence: Teleoperators and Virtual Environments, 5, 290-301.

IJsselsteijn, W., de Ridder, H., Hamberg, R., Bouwhuis, D., & Freeman, J. (1998). Perceived depth and the feeling of presence in 3DTV. Displays, 18, 207-214.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross media presence questionnaire: The ITC-Sense of Presence Inventory. Presence: Teleoperators and Virtual Environments (Special Issue), 10 (3), 282-297

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

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

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. Presence: Teleoperators and Virtual Environments, 9 (5), 497-503.

Welch, R.B., Blackmon, T.T., Liu, A., Mellers, B.A., & Stark, L.W. (1996). The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. Presence: Teleoperators and Virtual Environments, 5, 262-273.