

3-1-2010

Telepresence, soundscapes and technological expectation: putting the observer into the equation

Gary Pettey

Cleveland State University, g.pettey@csuohio.edu

Cheryl C. Bracken

Cleveland State University, c.bracken@csuohio.edu

Bridget Rubenking

Michael Buncher

Follow this and additional works at: https://engagedscholarship.csuohio.edu/clcom_facpub

 Part of the [Communication Commons](#)

[How does access to this work benefit you? Let us know!](#)

Publisher's Statement

"The final publication is available at Springer via <http://dx.doi.org/10.1007/s10055-009-0148-8>

Recommended Citation

Pettey, G., Bracken, C. C., Rubenking, B., Buncher, M., & Gress, E. (2010). Telepresence, soundscapes and technological expectation: putting the observer into the equation. *Virtual Reality*, 14(1), 15–25.
<https://doi.org/10.1007/s10055-009-0148-8>

This Article is brought to you for free and open access by the School of Communication at EngagedScholarship@CSU. It has been accepted for inclusion in Communication Faculty Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

Telepresence, soundscapes and technological expectation: putting the observer into the equation

Gary Pettey Cheryl Campanella Bracken
Bridget Rubenking Michael Buncher
Erika Gress

Abstract In an experiment exploring the impact of sound on sensations of telepresence, 126 participants watched a video clip using either headphones or speakers. The results illustrate that sound is an important factor in stimulating telepresence responses in audiences. Interactions between soundscape and screen size were also revealed. A traverse interaction between aural/visual congruency and soundscapes was evident. A second data set of 102 participants was collected to illuminate the effect of technological expectation that emerged in the first study. Expectations had been mentioned in other studies, and the data support the notion that people have an expectation of the technological quality of a presentation. The results suggest that examining expectations could assist in future conceptualizations of telepresence.

1 Introduction

While birds may be able literally to sleep with one eye (and half a brain) open (Rattenborg et al. 1999), humans appear to monitor their sleeping surroundings using sound. Patients under anesthesia have learned words played to them through headphones, even though they had no explicit memory of hearing them (Deeprouse et al. 2004). For humans, the soldier's notion of sleeping with one eye open may actually be more appropriately stated as keeping one *ear open* (Portas et al. 2000). Meaningful and unexpected sound appears to be actively processed in active parts of the brain that are "awake" even while we sleep. While we often say vision is our most important sense, in fact, sound and how that sound is reflected within its environment may well give us more information—especially as it can tell us something about things that happen outside our visual range and even in the dark (Blessner and Salter 2007a).

Engineers have worked hard over the last century to reproduce sound and then to recreate it more realistically in our living rooms and movie theaters. Stereo split the sound left and right. Surround sound gave the presentation depth. Dolby took the hiss from the tape removing the artificiality of the technology and leaving a cleaner reproduction of the music. Digital recording further recreated sound and a soundscape closer to original sound. Today, Dolby 5.1 and surround sound allow home theaters where listeners feel as though they have been transported into the environment of their movies.

In theaters, Sony Dynamic Digital Sound (SDDS) is an 8 channel reproduction. Digital Theater Systems (DTS) advertise their sound as bringing entertainment to life. Lucas' THX theater system is a balancing of the sound to the environment (soundscape) "At THX we believe that 'sound is fifty percent of the movie experience'," (THX

2009a) and by certifying the theater as THX. “What better way to escape from reality than by going to the cinema? The more immersive the performance, the more you enjoy the feature presentation. This is where a cinema built to THX standards comes into play” (THX 2009b).

Different elements of the aural modality have differential impact on the telepresence experience. This paper will attempt to do two things: Firstly, it will explore the notion of acoustic architecture and its implications both for mediated experiences from a more conceptual, perhaps less technical perspective; and secondly, it addresses the question whether varying elements of the soundscape help us to better understand how the aural modality contributes to dimensions of telepresence.

2 Presence and telepresence

There have been numerous definitions of presence put forth (Kim and Biocca 1997; Lee 2004; Lombard and Ditton 1997; Witmer and Singer 1998). Presence has been defined as a perceptual process where the media user somehow looks past or overlooks the technology to experience the medium (Lombard 2000) or as an illusion of non-mediation (Lombard and Ditton 1997). The concept of presence has been identified as a multidimensional concept (Lombard and Ditton 1997; Freeman 2004) with immersion (a sense of involvement), spatial/physical presence (a sense of being in the media environment), and social/behavioral realism (how plausible is the media environment) being seen in most presence measurement scales (Freeman 2004).

Lombard recently argued for the use of the term *telepresence* instead of presence (Lombard 2008) to stress the role of technology in the perceptual process. While many telepresence studies are conducted in highly immersive technologies such as virtual reality systems, there are a growing number of researchers within the social sciences and specifically the discipline of Communication who are investigating telepresence (Bracken 2005, 2006; Bracken et al. 2008; Ivory and Kalyanaraman 2007; Skalski and Tamborini 2007). Recently, Bracken et al. (2008) offered a definition of telepresence that makes a clearer break from the virtual reality aspects of presence and clearly puts it into a psychological context where the media observer uses technology.

Telepresence requires the use of technology and results in a psychological state in which media users voluntarily suspend the experience of mediation in order to feel a sense of connection with the mediated content they are using (i.e., a part of the action, connected to characters, involved in the story line). This state is often influenced by the expectation of the

technology, the media content, and characteristics of the media user. A sense of telepresence is felt by media users when the technology becomes transparent in the interaction (p. 4).

This definition is unique in that it focuses on the interaction among the user, the content, and the technology. While previous studies have looked at sound and telepresence (Bracken et al. 2008), this study will add the influence of soundscapes on the telepresence experience.

2.1 Telepresence and sound

Recent studies have examined the impacts of the aural modality on elements of telepresence. While the emphasis in presence research has been on the visual, aural effects alone or their interaction with the visual elements have been shown to be important to dimensions of telepresence.

Congruency between aural and visual modalities has been shown to increase levels of presence and enjoyment (Larsson et al. 2007). Using a within-subject design, the authors demonstrated that the congruency between the aural and visual channels reinforced participants experience across a spectrum of telepresence measures (Larsson et al. 2007). While their data show that sound vs. no sound showed the largest differences, in several measures they found that participants perceived differences in acoustical architecture between a conventionally shaped room and a cathedral shaped space about 3× as large.

Nunez (2007) suggests that the expectations of the participant may impact the presence experience. He found that an added sound track (non-diegetic information) congruent with the participant’s expectations had impact on how natural the participant found the experience. This congruency, even when controlling for an emotional response, predicted higher evaluations of the naturalness of the experience. This is noteworthy given that it reflects a learned response of mediated experience because it has a sound track. The lack of a sound track left a void in the experience. In an “ear budded” world, perhaps we shouldn’t find this tendency surprising. Further, Nunez found relationships between Izard’s (1977) DES-II (Differential Emotion Scales; Izard 1972; Izard et al. 1974) and various levels of telepresence.

Based on the findings of these studies concerning congruency, the following hypotheses are proposed:

Hypothesis 1a Participants who are exposed to congruent visual and aural modalities will report higher levels of *immersion* than the participants who are exposed to an incongruent presentation.

Hypothesis 1b Participants who are exposed to congruent visual and aural modalities will report higher levels of

spatial presence than the participants who are exposed to an incongruent presentation.

Hypothesis 1c Participants who are exposed to congruent visual and aural modalities will report higher levels of *social realism* than the participants who are exposed to an incongruent presentation.

3 Sonification

One of the problems in studying the aural modality is that there is no single discipline in which it is studied, and the language used across disciplines is often incompatible (Blesser and Salter 2007a). Blesser's work has been at the junction of audio, acoustics, perception, and cognitive psychology. He and Salter's 2007 work, *spaces speak, are you listening: experiencing aural architecture*, provides a serviceable conceptual framework.

Sonification in this context refers to the aural experience of a particular sonic event in a particular soundscape. A space occupies some physicality and is filled with some objects that reflect/absorb vibrations from the sonic event (soundscape). The position of the sonic event and the observer as well as their relative positions in the soundscape create the aural experience.

The complexity of sonifying a mixed reality space was explored by Le Groux et al. (2007). They argue that sonification is crucial to feelings of immersion. They note the difficulty of sonifying a virtual space in terms of soundscapes, synthetic voices, and dynamic emotive music.

Focusing only on the aural modality, other researchers (Tajadura-Jimenez et al. 2007) found increased accuracy in identifying sonic events emanating from their rear when synchronized with other vibration. Sonic events originating from the front of the participant were significantly less accurately identified. Information coming from the rear must be identified and processed largely independently of the visual.

Kallinen and Ravaja (2007) using a within-subject design compared headphones to speakers while listening to news [28]. Participants reported "liking" the headphones more, even though they reported that "quality" of the sound from the two sources was equal. Participants reported the speaker condition to be more "realistic" while listening to news from a computer. One interpretation is that is how they perceive most people listen to the news, or at least news on a computer—a level of social realism. Using physiological measures and self-report data, they predicted that headphones would elicit a more intense and immersive listening experience. They found that headphones increased attention and, by extension, involvement. They suggested this was due to technologies that tend to

isolate the listener from her surroundings (Witmer and Singer 1998).

Based on the findings of these studies concerning sonification, the following hypotheses are proposed:

Hypothesis 2a Participants who listened through headphones will report higher levels of *immersion* than the participants who listened through speakers.

Hypothesis 2b Participants who listened through headphones will report higher levels of *spatial presence* than the participants who listened through speakers.

Hypothesis 2c Participants who listened through speakers will report higher levels of *social realism* than the participants who listened through headphones.

3.1 An aural language

A *sonic event* is the creation of a sound wave that rises above the background (snapping one's fingers). If we snap our fingers we hear the snap, but we also hear the reflection of that snap from the passive objects in our space (walls, ceiling, and floors). Thus, the same sonic event is perceived quite differently if done in a small room or if it is done in a large one. The space (including sound waves reflected by objects in the space) in which sounds are experienced are *soundscapes*. Just as we need light to appreciate a landscape, we need sound to perceive/appreciate a soundscape.

An *aural architecture*, then, "refers to the properties of a space that can be *experienced* by listening" (Blesser and Salter 2007b, p. 5). We experience space aurally. We can *taste* a space aurally. For example, we *know* the difference between the aural space of a bathroom, relatively small physical space with hard reflective surfaces (tiled walls and floor and porcelain fixtures), especially in comparison with a living room, which is normally a larger physical space with softer absorbing surfaces (drapery, carpeting, furniture). Our experience of a place is directly related to our experience of the place's aural architecture.

We can imagine the aural architecture of a space assumes that we have a *spatial awareness*. Spatial awareness would include more than the complexity of a reflected sonic event or even series of such events, it also includes the individual's emotional and behavioral experiences with the space. Blesser and Salter conceptualize the complexity of spatial awareness as having "spatial attributes, auditory perception, personal history and cultural values" (2007b, p. 11). They argue that spatial awareness influences us in at least four ways:

1. **Social Spatiality:** Influences social behaviors. Some spaces may emphasize aural privacy, which may aggravate loneliness or increase social cohesion.

2. Navigational Spatiality: A sense of orientation that allows us to maneuver. In darkness, spatial awareness actually replaces vision.
3. Esthetic Spatiality: It conveys a sense of the esthetic. Spaces low in acoustic features are aurally barren even boring.
4. Musical Spatiality: Can enhance complex aural experiences such as music and voice giving them depth and texture (or not) (See Pettey et al. 2008; Blesser and Salter 2007a for more explanation and examples).

Blesser and Salter describe the range of auditory awareness as: *sensation, perception, and affect*. They note that generally awareness is problematic because it implies consciousness, but sensation, they argue, is below the conscious level—*mere detection*. We can sense the closeness of the closet without conscious realization of the aural features involved. The lightness of a sunny day when compared with the heaviness of a foggy day is one experience at this level. These may well be the important subconscious cues that would be important in elements of immersion and perceived realism, for example.

Perception or recognition is when conscious processes have been engaged but not emotional ones. Cultural influences and personal experiences give meaning to what may have barely risen above sensation. A musician can tell the difference between a violin string and a cello string. Most westerners can tell the difference between a guitar string and a piano (a plucked rather than struck string). Language is a set of sonic events to which we attach cognitions. The more cultural influence and personal experience we have with a language, the more meaning we may get from any set of sonic events. At some point, the awareness moves from the cognitive realm to one of emotional meaning.

Finally, affect is emotionally engaged hearing and carries meaning. Hearing a song that reminds you of a party and brings a smile to your face, or hearing a lawn mower that reminds you of long unhappy days mowing as a child constitutes such awareness. Such sounds can create high emotional—even visceral—responses (Thayer 1989). Sometimes, the listener may not even consciously understand the relationship between the sound and his/her reaction. In some instances, music and sound has been used to create dream states (Besmer 1983; Rouget 1985). Blesser and Salter note that working with affective extremes of awareness is probably too complex to be adequately studied without brain scans. As the effect is a complex biological, psychological, social one, and individuals are unlikely to be able to adequately describe their emotional state (describe their feelings), such overt affect effects are probably less fruitful at this point. They suggest that awareness at the other end of the spectrum (they call

them subliminal) correspond to more subtle levels of awareness or moods. The task then is to use/find sonic events and acoustical spaces that (1) have some meaningful content for listener and (2) adequately engages the listener. “To use a food metaphor, sonic events are raw ingredients, aural architecture is the cooking style, and, as an inseparable blend, a soundscape is the resulting dish” (Blesser and Salter 2007b, p. 15).

3.2 Experiencing a soundscape

Each culture controls what we expect for social distance. Hall divided these spaces into four: the intimate, the personal, the conversational, and the public. Soundscapes interact with these social spaces (Hall 1966). The intimate space is highly restricted to very close friends and lovers. Acoustically, this space is very private. Words and sounds are for the intimates only. Silence will greet intruders. The personal space is for friends and acquaintances. Think of a group of students or colleagues talking in the hallway. The conversational space is the comfort space reserved for stranger interactions. Finally, the public space has many varying rules usually based on some social rules for the space. There are few restrictions on a public beach or park, but more restrictions may exist on public indoor spaces such as an airport concourse. How one reacts to a soundscape should differ depending upon the interplay of the soundscape and the social distance involved. Headphones would be the most intimate of sound, while listening to a concert in an outdoor venue would be a very public one. Watching television in your own living room is pretty private almost intimate, while watching it at someone else’s living room would be much more public.

Based on this explanation of soundscape and social distance and considering how this might impact the sense of telepresence when combined with aural and visual congruency of the presentation, we offer the following research questions:

RQ 1a: If headphones create more intimacy (thus more immersion), and congruent presentations are also more immersive, will they interact to produce effects over and above their individual main effects?

RQ 1b: If headphones create more intimacy and a soundscape that is separate from the viewing environment and congruent presentations allow for better spatial awareness, will they interact to produce effects over and above their individual main effects?

RQ 1c: If speakers are reported as being more real and congruent presentations allow for higher levels of social realism, will they interact to produce effects over and above their individual main effects?

4 Methods

This overall design employed a between-subject 2 (soundscape) \times 2 (congruence) experiment. The 126 participants watched a video clip where the audio was delivered either through standard headphones or stereo television speakers. The audio was either the one from the motion picture or one where the audio had alternative non-diegetic sound.

4.1 Stimulus

The video shown to the participants was from the movie *Quest for Fire* (Annaud 1981). This film was made in 1981 and was used because there is no character dialog. It was also believed that few participants would have previous exposure to the film. The video clip was approximately 15 min in length, of which the first minute and a half was from another source (the earth rising in the blackness of space with soft string music) and provided a baseline for the physiological measures (not used in this paper). The movie clip began immediately following the introduction. The natural sound track of the film contained sparse classical music that had a primitive sound, which allowed it to match the prehistoric period and theme of the film. The non-diegetic music material was taken from the song Boulevard (Parisian Mix) by Cinemascope from the CD Saint Germain Lounge: Rendez Vous. The non-diegetic music was inserted in the approximately 5-min final scene and contradicted the mood of the scene, which showed one prehistoric tribesmen observing another make fire. The incongruent music that was added was electronic and modern in nature. The non-diegetic music was also more upbeat and pleasant sounding, which tended to mitigate the drama of the fire creation, which was the climax of the film and the intensity could be seen on one of the men's faces.

4.2 Independent variables

4.2.1 Soundscape

The first independent variable was how the participant experienced the sound. Soundscape was manipulated by having one condition group listen to the video clip through the television speakers (more public) and the other condition group listening through headphones (more intimate).

4.2.2 Congruence

Congruence had two conditions. The first contained the film's natural congruent audio content. It contained grunts and diegetic sound elements with occasional classically based music sound track. The other condition (incongruent)

contained the same sound track until the final, climactic scene where the music was overlaid. This music was not jarring to viewers in a pretest, but left them confused about what the scene was "trying to say."

4.3 Dependent variables

The amount of telepresence experienced by the participants was measured using a multidimensional presence scale (Lombard and Ditton 1997; Lombard et al. 2009; Bracken 2005, 2006). The telepresence dimensions adapted from the Temple Presence Inventory (TPI) including immersion, spatial presence, and social realism are detailed in the following paragraphs.

4.3.1 Immersion

Immersion was measured by asking participants to respond to six statements using a scale from not at all (1) to very much (7). The immersion scale was developed to assess the extent to which audience members feel they are absorbed in the media programming. The items included were "How involving was the video?" "How completely were your senses engaged?" "To what extent did you feel mentally immersed in the video?" and "I was so involved in the video that I lost track of time." The scale had a Cronbach's alpha of .90.

4.3.2 Spatial presence

Participants responded from very strongly disagree (1) to very strongly agree (7) for six Likert-type statements designed to measure the extent to which television viewers feel a sense of sharing a physical space with the mediated environment (Lombard and Ditton 1997). The items were: "How often when an object seemed to be headed toward you did you want to move to get out of its way?" "How much did it seem as if the objects and people you saw/heard had come to the place you were?" "How much did it seem as if you and the people you saw/heard were together in the same place?" "To what extent did you experience a sense of 'being there' inside the video you saw/heard?" "How much did you feel as if you were inside the video you saw/heard *observing* the events?" and "How much did it seem as if you left the places where you were and went to a new place?" The scale had a Cronbach's alpha of .92.

4.3.3 Social realism

Participants responded from strongly disagree (1) to strongly agree (7) for three statements intended to measure the extent to which television viewers feel a sense of realism when viewing television. The three statements

were “The events I saw/heard would occur in the real world.” “The events I saw/heard could occur in the real world.” “The way the events I saw/heard occurred is a lot like the way they occur in the real world.” The scale had a Cronbach’s alpha of .82.

4.4 Participants

Undergraduate students were recruited from several social science courses to participate. One hundred twenty-six undergraduate students from introductory classes were given credit from their instructors for participation. The group was 57% female (balanced across treatments) with an average age of 24.01 and a standard deviation of 7.2.

4.5 Procedure

Each participant was met by the experimenter and escorted into a carpeted, 8 × 10 foot room that contained a 46-inch LCD television, DVD player, and a comfortable chair that faced the television screen. Various other amenities, such as a decorative table lamp and pictures on the wall, provided the verisimilitude of a living room.

The experimenter explained that the participant would be watching a 15-min video clip and that after the clip they would exit the viewing room. After the participant exited the experiment room, the experimenter escorted the participant to a second room that contained tables and chairs. The experimenter instructed the participant to sit at one of the table and to answer a paper-and-pencil questionnaire.

The experimenter emphasized that there were no wrong answers and that the participant should follow the directions in the questionnaire. The entire procedure took about 35 min.

5 Results

As the design contained multiple independent variables and three dependent variables that previous studies had shown to be moderately correlated, a MANOVA was run. The independent variables were both included (soundscape and congruence). The dependent variables were all three of the telepresence variables (immersion, spatial presence, and social realism). None of the main effects reached significance, but the one 2-way interaction component was significant (Pillai’s Trace, $F = 2.86$, $p < .05$). Table 1 presents the ANOVA sets from the MANOVA.

Hypotheses 1a–c predicted higher levels of immersion, spatial presence, and social realism in the condition where the sound was congruent with the visual. There was no support for these hypotheses.

Hypotheses 2a–b predicted higher levels of immersion and spatial presence in the condition where the sound was presented via headphones. There was no support for these hypotheses.

Hypotheses 2c predicted higher levels of social realism in the condition where the sound was presented via speakers. There was no support for this hypothesis.

Research questions 1a–c asked whether congruence would interact with the soundscape for each of the

Table 1 Analysis of variance table: congruence and soundscape for presence dimensions

| Source of variation | Sum of squares | <i>df</i> | Mean squares | <i>F</i> | η^2 |
|-------------------------|----------------|-----------|--------------|----------|----------|
| <i>Immersion</i> | | | | | |
| Main effects | | | | | |
| Congruence | 2.36 | 1 | 2.36 | 0.03 | .00 |
| Soundscape | 6.39 | 1 | 6.39 | 0.09 | .00 |
| Two-way interactions | | | | | |
| Congruence × soundscape | 548.25 | 1 | 548.25 | 7.39** | .07 |
| <i>Spatial presence</i> | | | | | |
| Main effects | | | | | |
| Congruence | 25.58 | 1 | 25.58 | 0.41 | .00 |
| Soundscape | 17.91 | 1 | 17.91 | 0.28 | .00 |
| Two-way interactions | | | | | |
| Congruence × soundscape | 8.00 | 1 | 8.00 | 0.13 | .00 |
| <i>Social realism</i> | | | | | |
| Main effects | | | | | |
| Congruence | 6.50 | 1 | 6.50 | 0.93 | .01 |
| Soundscape | 2.06 | 1 | 2.06 | 0.09 | .00 |
| Two-way interactions | | | | | |
| Congruence × soundscape | 99.33 | 1 | 96.33 | 4.33* | .04 |

* $p < .05$; ** $p < .01$

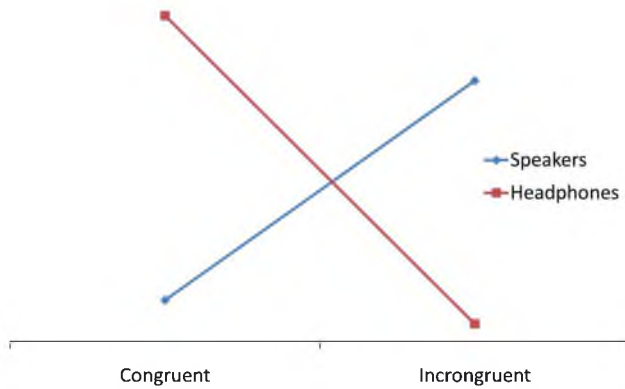


Fig. 1 Interaction of soundscape and congruency on immersion

variables. Previous studies had suggested that soundscape may interact with other independent variables (Bracken et al. 2008).

Here for the interaction between congruence and soundscape, we see that the MANOVA was significant, while (as seen in Table 1) there were no main effects for soundscape or the congruency of the aural and visual modalities.

RQ1a asked about the effect for soundscape on immersion. The interaction effect for congruence and soundscape for immersion was significant ($F = 7.39$; $p < .01$; $\eta^2 = .07$). The interaction was transverse with Headphones and Congruent ($M = 23.46$; $SD = 8.56$) and Speakers and Incongruent ($M = 22.69$; $SD = 9.23$) in the high cells, and Speakers and Congruent ($M = 18.54$; $SD = 6.86$) and Headphones and Incongruent ($M = 18.72$; $SD = 9.57$) in the low cells. These data are presented in Fig. 1.

For RQ1b, the soundscape did not interact with the congruency of the presentation for spatial presence.

RQ1c asked about the effect for soundscape on social reality. The interaction effect for congruence and soundscape for social reality was significant ($F = 4.33$; $p < .05$; $\eta^2 = .04$). The interaction was also transverse and follows the same pattern as with immersion. Headphones and Congruent ($M = 11.61$; $SD = 4.80$) and Speakers and Incongruent ($M = 10.62$; $SD = 4.79$) in the high cells, and Speakers and Congruent ($M = 9.31$; $SD = 4.74$) and Headphones and Incongruent ($M = 9.31$; $SD = 4.71$) in the low cells. These data are presented in Fig. 2.

6 Discussion

This study tried to further explore the conceptualization, the role of sound in creating sensations of telepresence. We attempted to provide a theoretical basis by using Blesser and Salter (and others) notion of soundscape. Other studies

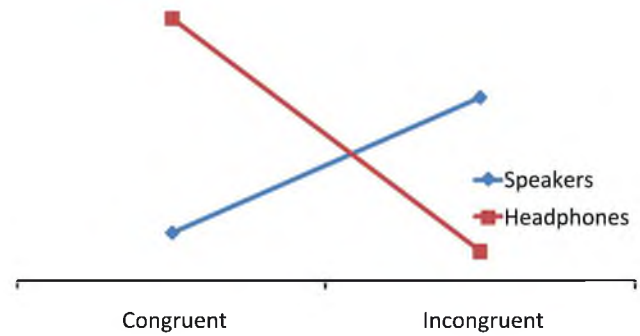


Fig. 2 Interaction of soundscape and congruency on social realism

have noted, there are differences between perception among listeners according to whether they experienced a presentation through headphones or television speakers (Kallinen and Ravaja 2007). Given that, we considered this the best place to begin. Even though other studies found main effect differences between aural and visual congruency in measures of telepresence, we found none. It should be noted that many of those previous studies were working primarily in a virtual reality setting maximizing differences. We worked within a mediated environment with considerably subtler manipulations. Given that we did not seek to maximize effects but rather were more concerned with external forms of validity, it may be possible that increasing the number of participants would be beneficial in finding main effects for congruency and soundscapes. But, the significant interactions between congruency and soundscapes are noteworthy. That the interactions were transverse suggests that increasing the number of participants may not lead to significant main effects. Further, since the relationships are the same for both immersion and social realism (but not spatial awareness) suggests we have encountered something about soundscapes that crosses the two dimensions of telepresence.

Bracken et al. (2008) conducted an experiment manipulating sound presentation (headphone vs. speakers), pace (fast vs. slow), and screen size (large screen [32 inches] vs. iPod [2.5 inches]). Participants saw one of two different presentations—both with congruent aural/visual modalities. They found that participants who wore headphones reported higher levels of immersion than those who listened to speakers. This is consistent with our prediction and understanding of the social distance of a soundscape—with headphones being more intimate and thus more immersive. It is also consistent with Kallinen and Ravaja (2007) findings suggesting that headphones are more immersive. Bracken et al.'s non-significant interaction shows that those watching on a personal medium (iPod) had higher levels of immersion than those who watched on more public one (standard television). So when the aural/visual are congruent, headphones show higher levels of immersion. This

is also the case in this study. Looking only at the congruent, headphones were more immersive than speakers. The question then is what is going on within the incongruent condition. We will further consider this in the following paragraphs where social realism shows a similar congruent/sonification relationship.

Given the previous work within this area, finding a non-significant result for spatial presence was surprising. But much of that literature deals with presence and virtual reality. Spatial presence is rarely found in popular media which is less immersive than virtual reality systems or even videoconferencing. We believe that our technology may have also been insufficient to produce the effects we expected. Our television sets were new, high quality, and stereo, while our headphones were more pedestrian—midrange stereo headphones that people commonly buy for use with their personal media. We believe that the study should be conducted with more advanced stereo headphones which may be a fairer comparison and worth doing in the future.

Social realism shows an interaction effect with congruency and sonification. We first consider the congruent side. Again, here we see that headphones showed higher levels of social realism than television speakers. This is consistent with the findings for immersion and with the intimacy hypothesis. Kallinen and Ravaja (2007) reported that participants rated listening to news via speakers more real than headphones, but their design did not include a visual modality. Perhaps, when there is discrepancy in the two modalities, people trust (find more real) the aural. Bracken et al. (2008) may shed some light on this. They report that for the television condition, headphones showed higher levels of social realism than did the speaker cells which is consistent with the findings here. Looking only at the congruent cells, headphones were rated as more socially realistic than speakers. Again, it is the incongruent cells that give us pause (as they did with immersion mentioned earlier).

One testable explanation is that when the technology leads to expectations of intimacy (and higher levels of telepresence), higher levels are experienced. Congruence between the two modalities would resonate with the intimate expectation and leads to higher levels of telepresence. However, when there is incongruence between the modalities, it breaks the intimate expectation lowering the participants' expectations and they find them more consistent with a more public soundscape, which is also less telepresent. The missing variable then is user (participant) expectation of the experience both in terms of content and the technological reproduction of that content. Should this be borne out, it would be an important element as it is not commonly part of the telepresence explication. Therefore, we conducted a second experiment that measured the

technological expectation of the participants. Such expectations on the part of a person may well summarize both the delivery quality of the technology and the experience she or he may have. The study is presented in Sect. 7.

7 Expectation of technology and experience

Heeter makes the argument that expectation can have an impact on sensations of telepresence. She states “Expectations, lack of familiarity, limited prior experience, and limited cognitive schemas dampened my sense of presence” (Heeter 2003, p. 336). More recently, Nunez (2007) also suggested that expectations mattered. He reported (as we noted earlier) that an added sound track (non-diegetic information) congruent with the participant's expectations had impact on how natural the participant found the experience.

We have seen that as quality of the technology increases (e.g., high definition television Bracken 2005), levels of telepresence increase. We envision that when individuals encounter a technological experience, they make certain judgments about what that experience will be like. Further, when the technological quality exceeds their initial judgments, one would expect that their levels of telepresence would also increase. To that end, we offer one additional hypothesis:

Hypothesis 3a Individuals whose expectations of a mediated experience were exceeded by the technology will report higher levels of *immersion*.

Hypothesis 3b Individuals whose expectations of a mediated experience were exceeded by the technology will report higher levels of *spatial presence*.¹

7.1 Methods

This overall design employed a between-subject 3 (audio expectation) × 3 (visual expectation) experiment. As part of another study, the 102 participants watched a video clip where the audio was delivered either through standard headphones or stereo television speakers. For this analysis, the soundscapes were collapsed to examine only how the participants' expectations of the technology they were to use affected their reports of telepresence.

7.1.1 Stimulus

Upon being seated in the experimental room, students were asked to rate the levels of the quality expected from the

¹ Please note that Study 2 only examined technological expectations (not content), so social realism was not tested.

television, its speakers, or the headphones. They were then asked after the presentation to rate their experiences as better than they expected, about the same, or worse than they expected.

7.2 Independent variables

7.2.1 Audio expectations

Participants were sensitized to think about the equipment they saw in the room and to rate on a 0–10 scale (Very low quality to Very high quality). After the presentation, they were then asked to consider and rate the quality of the audio in the presentation. The response options were “Was it better than they expected”, “About the Same,” or “Worse than they expected.”

7.2.2 Visual expectations

Just as with the audio grouping, participants had been sensitized to think about the equipment they saw in the room and to rate on a 0–10 scale (Very low quality to Very high quality). After the presentation, they were then asked to consider and rate the quality of the video in the presentation. The response options were “Was it better than they expected”, “About the Same,” or “Worse than they expected.”

7.3 Dependent variables

The dependent variables were measured in the same way as in study 1. The Cronbach’s alpha for immersion was .86 and .80 for spatial presence.

7.4 Participants

Undergraduate students were recruited from several social science courses to participate. One hundred and two undergraduate students from introductory classes were given credit from their instructors for participation. The group was 48% female with an average age of 24.75 and a standard deviation of 7.9.

7.5 Procedure

The procedure is the same as in study one.

7.6 Results

The design contained two independent variables and two dependent variables. Even though no previous studies had examined this relationship, a MANOVA was run as a baseline. Two independent variables (video expectation

and audio expectation) and the two relevant telepresence variables (immersion, spatial presence) were included in the model. Audio expectation showed a significant multivariate main effect (Pillai’s Trace, $F = 2.60$, $df 4,194$, $p < .05$). As only four participants reported the television image was worse than expected, these were merged with the “about what was expected” group leaving only two conditions of the visual modality.

The hypothesis was partially supported with significant differences being found for immersion ($t(79) = 2.27$, $p < .03$, $\eta^2 = .05$). The participants reported higher levels of immersion when the visual modality exceeded their expectation ($M = 41.0$, $SD = 9.96$) than when the media experience met their expectations ($M = 36.8$, $SD = 10.06$). Significant differences were not found for spatial presence but the means are in the predicted direction. The participants reported higher levels of spatial presence when their expectations were exceeded ($M = 7.98$, $SD = 4.75$) than the participants who reported the experience was as expected ($M = 7.16$, $SD = 3.70$).

The same pattern was found for the auditory modality. Specifically, participants reported higher levels of immersion ($F(2,99) = 5.73$, $p < .01$, $\eta^2 = .10$)² when the experience exceeded their expectation ($M = 43.41$, $SD = 9.47$) than when the experience met their expectation ($M = 37.15$, $SD = 9.7$) or was reported as being “worse than expected” ($M = 35.32$, $SD = 9.97$). Significant differences were not reported for not spatial presence, though the means, while not significantly different, were in the direction predicted.

7.7 Discussion

The findings of this second experiment support the definition of telepresence proposed by Pettey et al. (2008) and their claim that expectation of technology will impact the level of telepresence experienced. A consistent pattern of higher levels of telepresence was reported when expectation was exceeded, followed by lower levels when expectation was met was found for both immersion and spatial presence. These results provide additional support for the claims made by Heeter (2003) and Nunez (2007).

It appears that media users today may be sophisticated enough about technologies that they adjust their anticipation of the presentation accordingly. Additionally, the number of presentation choices (i.e., IMAX, portable devices) available to consumers is higher than ever, and their ability to make choices between these presentation forms influences their media experience. This has not been

² Scheffe post hoc shows all mean pairs with Better than Expected are different.

generally a concern in the study of telepresence. If the technology provides a “better” experience than observers expect, it can affect their feelings of telepresence. Expectation of technology should be examined when studying telepresence, and if that’s the case, then there may be *other observer variables* that ought to be included in the telepresence model. It has not escaped the authors’ attention that this could substantially expand the types of variables that should be conceptualized and measured (user/observer) and may allow telepresence to integrate and include other traditional mass media concepts such as attention, cognitive capacity, and motivations to use specific media and content.

8 Conclusion

Clearly, the most important implication of this study is that soundscapes matter. Soundscapes can be explicated and do impact on elements of telepresence. Further, that the findings are consistent with other studies demonstrating that sound has interaction effects, leads us to conclude that telepresence may need further elaboration possibly including the audience expectations of their mediated experience—both in terms of content and soundscape, and possibly the interaction between them. The results demonstrating the role of technological expectation and the implications of observer variables suggest further investigation by telepresence researchers is warranted.

Future studies might consider more explicitly the separation of and relationship between technology form and content. Much in the presence literature is appropriately concentrated on the technological impact on individuals. Telepresence includes both content and observer psychological variables. Further, future studies should consider content across types of technologies. As content and observer variables appear to have impact on levels of telepresence, future research might try to include such concepts in their explications.

References

Annaud JJ (1981) Quest for fire [Motion Picture]. International Cinema Corporation (ICC), France
 Besmer F (1983) Horses, musicians, and gods: the Hausa cult of possession-trance. Bergen & Garvey, South Hadley, MA
 Blesser B, Salter LR (2007a) Spaces speak, are you listening? Experiencing aural architecture. The MIT Press, Cambridge, MA
 Blesser B, Salter LR (2007b) Working vocabulary for the experience of aural architecture. Retrieved February 17, 2008, from <http://www.blesser.net/downloads/Glossary.pdf>
 Bracken CC (2005) Presence and image quality: the case of high definition television. *Media Psychol* 7:191–205

Bracken CC (2006) Perceived source credibility of local television news: the impact of image quality and presence. *J Broadcast Electron* 50:723–741
 Bracken CC, Pettey G, Rubenking B, Guha T (2008) Sounding out small screens and telepresence: the impact of screen size, pace & sound. Presented to the Information Processing division of the International Communication Association annual convention in Montreal, Quebec, Canada
 Deeprose C, Andrade J, Varma S, Edwards N (2004) Unconscious learning during surgery with propofol anaesthesia. *Br J Anaesth* 92:171–177
 Freeman J (2004) Implications for the measurement of presence from convergent evidence on the structure of presence. Paper presented to the Information Systems Division at the annual meeting of the International Communication Association, New Orleans, LA
 Hall E (1966) *The hidden dimension*. Doubleday, New York
 Heeter C (1992) Being there: the subjective experience of presence. *Presence Teleop Virt Environ* 1(2):262–271
 Heeter C (2003) Reflections on real presence by a virtual person. *Presence Teleop Virt Environ* 12(4):335–345
 Ivory JD, Kalyanaraman S (2007) The effects of technological advancement and violent content in video games on players’ feelings of presence, involvement, physiological arousal, and aggression. *J Commun* 57:532–555
 Izard CE (1972) *Patterns of emotions: a new analysis of anxiety and depression*. Academic Press, New York
 Izard CE (1977) *Human emotions*. Plenum Press, New York
 Izard CE, Dougherty FE, Bloxom BM, Kotsch NE (1974) The differential emotions scale: a method of measuring the subjective experience of discrete emotions. Vanderbilt University Press, Nashville, TN
 Kallinen K, Ravaja N (2007) Comparing speakers versus headphones in listening to news from a computer—individual difference and psychophysiological responses. *Comput Hum Behav* 23:303–317
 Kim T, Biocca F (1997) Telepresence via television: two dimensions of telepresence may have different connections to memory and persuasion. *J Comput Med Commun* 3(2)
 Larsson P, Vastfjall D, Olsson P, Kleiner M (2007) When what you hear is what you see: presence and auditory-visual integration in virtual environments. Paper presented at the 10th annual international workshop on presence, Barcelona, Spain
 Le Groux S, Jonatas Manzolli J, Verschure P (2007) Interactive sonification of the spatial behavior of human and synthetic characters in a mixed-reality environment. P10 Ann Int Worksh
 Lee KM (2004) Presence, explicated. *Commun Theory* 14:27–50
 Lombard M (2000) Resources for the study of presence: presence explication. Retrieved February 28, from http://www.temple.edu/ispr/frame_explicat.htm
 Lombard M (2008) Presence and telepresence scholarship: challenges ahead. Paper presented to the “When Media Becomes Reality Workshop”, Berne, Switzerland
 Lombard M, Ditton T (1997) At the heart of it all: the concept of presence. *J Comput Med Commun* 13:573–590
 Lombard M, Ditton T, Weinstein L (2009) Measuring telepresence: the temple presence inventory. Paper presented to the twelfth annual international conference of the presence workshop, Los Angeles, CA, USA
 Nunez D (2007) Effects of non-diegetic information on presence: a content manipulation experiment. P10 Ann Int Worksh
 Pettey G, Bracken CC, Rubenking B, Appley E, Brown M, Butler D, Cline K, Coralic Z, Hargiti J, Liou H (2008) No vision is complete without sound*: the sonification of telepresence. In: *Proceedings of the eleventh annual international meeting of the presence workshop*, Padova, Italy, 38–45

- Portas C, Krakow K, Allen P, Josephs O, Armony J, Frith C (2000) Auditory processing across the sleep-wake cycle simultaneous EEG and fMRI monitoring in humans. *Neuron* 28:991–999
- Rattenborg NC, Lima SL, Amlaner CJ (1999) Half-awake to the risk of predation. *Nature* 397:397–398
- Rouget G (1985) *Music and trance: a theory of the relations between music and possession* (trans: Biebuyck B). University of Chicago Press, Chicago
- Skalski P, Tamborini R (2007) The role of social presence in interactive agent-based persuasion. *Media Psychol* 10:385–413
- Tajadura-Jimenez A, Vajamae A, Kitagawa N, Ho HN (2007) Whole-body vibration influences sound localization in the median plane. P10 *Ann Int Worksh*
- Thayer R (1989) *The biopsychology of mood and arousal*. Oxford University Press, New York
- THX (2009a) HDTV buying tips from THX. Retrieved April 1, 2009 from <http://www.thx.com/products/home/displays/shopping.html>
- THX (2009b) THX certified cinema. Retrieved April 1, 2009 from (<http://www.thx.com/cinema/index.html>)
- Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. *Presence Teleop Virtual Environ* 7:225–240