David Hecht*
Technion — Israel Institute
of Technology
Department of Education in
Technology and Science
Haifa, Israel
and
The University of Haifa
Haifa, Israel

Miriam Reiner
Gad Halevy
Technion — Israel Institute
of Technology
Department of Education in
Technology and Science
Haifa, Israel

Multimodal Virtual Environments: Response Times, Attention, and Presence

Abstract

Multimodal virtual environments (VE) succeed better than single-sensory technologies in creating a sense of presence. We hypothesize that the underlying cognitive mechanism is related to a faster mental processing of multimodal events. Comparing simple detection times of unimodal (auditory, visual, and haptic) events, with bimodal and trimodal combinations, we show that mental processing times are in the following order: unimodal > bimodal > trimodal. Given this processing-speed advantage, multimodal VE users start their cognitive process faster, thus, in a similar exposure time they can pay attention to more informative cues and subtle details in the environment and integrate them creatively. This richer, more complete and coherent experience may contribute to an enhanced sense of presence.

1 Introduction

Multimodal virtual environment (VE) systems, able to efficiently combine sensory information from two or three channels (vision, audio, haptic), have an advantage in generating a sense of presence. This *multi*-sensory experience differentiates them from older technologies, communicating only via a *single*-sensory channel, which were usually less immersive and created only a limited degree of presence. Therefore, there is an agreement in the presence literature, that the more multimodal a virtual environment is designed, the greater the sense of presence it will be able to generate (Held & Durlach, 1992; Sheridan, 1992; Witmer & Singer, 1998; Biocca, Kim, & Choi, 2001; Romano & Brna, 2001; Sanchez-Vives & Slater, 2005). However, the underlying cognitive mechanisms, in which multi-modal VE succeed in creating an enhanced sense of presence, are still elusive and unknown. In the following sections, we present the ideas introduced by researchers, and then suggest another complementary mechanism.

1.1 Environmental Richness Results in a Complete and Coherent Experience

A rather intuitive idea suggests that single channel media is relatively sensory-poor and conveys limited and insufficient information to the senses;

Presence, *Vol. 15, No. 5, October 2006,* 515–523

---

*Correspondence to davidh@tx.technion.ac.il thus it engenders a lower sense of presence. Conversely, multimodal environments provide a greater extent of sensory information to the observer. This sensorial richness translates into a more complete and coherent experience. And therefore, the sense of being present in the virtual realm is felt stronger (Held & Durlach, 1992; Sheridan, 1992; Witmer & Singer, 1998).

1.2 Multimodal VE are Mimicking Reality Better

Another way in which multimodal VE succeed in creating a stronger sense of presence is by better mimicking reality (Romano & Brna, 2001). An elaboration of this idea would be as follows: many of our natural daily experiences in the real world are fundamentally multimodal by their nature. For instance, reaching to grasp an object, or even simple posture and movement control, are a coproduction of the visual, haptic, and vestibular systems (Mergner & Rosemeier, 1998). Communicating with another person through speech is a fine combination of producing and receiving audio and visual cues—sound, lip movements, and gestures (Bernstein, Auer, & Moore, 2004). Our gastronomic pleasures, too, result from a fine integration of taste, smell, and vision (Dalton, Doolittle, Nagata, & Breslin, 2000; Gottfried & Dolan, 2003).

Therefore, multimodal VE have a clear advantage, in mimicking a multimodal phenomenon, since they stimulate not only the user's auditory and visual sensory systems (with a realistic 3D depth perception), but in addition, as a result of capturing the entire perceptual field, via head mounted display or a wide field of view (up to 360° presentation), even when the user is stationary and completely passive, movements in the screen stimulate the vestibular system, as evidenced by illusory self-motion or vection (Palmisano, Gillam, & Blackburn, 2000; Bonato & Bubka, 2004) and the simulator sickness phenomenon (Draper, Viire, Furness, & Gawron, 2001; Duh, Parker, Philips, & Furness, 2004), and by users' production of natural body movements in virtual environments (Slater & Steed, 2000). The experience is especially felt as real if it includes also haptic—tactile and kinesthetic—sensations (Reiner, 2004; Basdogan, Ho, Srinivasan, & Slater, 2000).

1.3 Intersensory Experience Enables Better Integration and Filling in of Missing Information

Another proposed mechanism, in which multimodal VE gain an edge in creating a sense of presence, is related to our mind's attempt for sensory integration. Since synthetic VEs provide fewer sensory cues than most physical environments in which we act, the user needs to interpolate sensory stimuli to create a functional mental model and use these cues to walk towards, reach out to, and manipulate objects in the environment. During the process of integrating and augmenting impoverished sensory cues, information from one sensory channel may be used to augment ambiguous information from another sensory channel (Biocca et al., 2001).

Hence, since in multimodal VE the cognitive process of integration induces an intersensory filling in of missing information, in a rather active and creative manner (depending on the user abilities), this active cognitive-integration process result in an enhanced sense of presence.

1.4 Faster Mental Processing Enables Deeper and Richer Experience

While the aforementioned three explanations focus mainly on higher cognitive functions, occurring at the end of the cognitive processing-stream we suggest another complementary mechanism that occurs already earlier, at the initial perception level, at the beginning of the processing-stream, which gives an advantage to multimodal VE, over single-channel systems, in creating a sense of presence.

Using a simple reaction time (RT) paradigm we compared the brain processing speed of unimodal signals (audio, visual, or haptic) with its processing speed of bimodal combinations of these signals and a trimodal combination of these same signals. Our hypothesis suggests an advantage, in processing speed, for bimodal signals over unimodal signals. Furthermore, we hypothesized that trimodal signals will be processed even faster than all bimodal combinations.

The rationale for this study is that if indeed multimodal events are processed faster than unimodal events, it

Figure 1. *While users held the pen-like stylus (on the right) performing writing-like movements, the attached force-feedback mechanism generated a resisting force—haptic stimulation. Users responded by pressing a button on the stylus.*

can affect the entire event to be experienced as richer, more complete, and coherent. Faster processing at the initial perceptual stage (at the first 250–400 ms) allows users more time in the consequent cognitive stages, enabling them better integration and filling in of missing information. Thus, at the end of the cognitive process, this richer experience may contribute to the greater sense of presence.

2 Experimental Design

2.1 Materials

In this study, we used a touch-enabled computer interface that can generate visual, auditory, and haptic stimulations. The haptic device (shown in Figure 1) is based on a force-feedback mechanism. Full technical descriptions of this system are available at www.reachin.se and www.sensable.com.

2.2 Participants

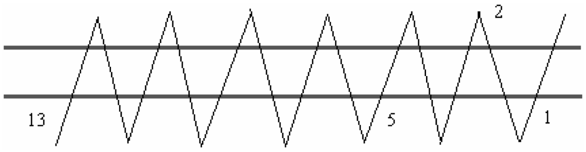


Figure 2. *Participants performed writing-like movements with the stylus crossing the parallel horizontal lines. Between the 5th and the 13th crossings, the computer generated, randomly, a sensory stimulation, unimodal, bimodal, or trimodal.*

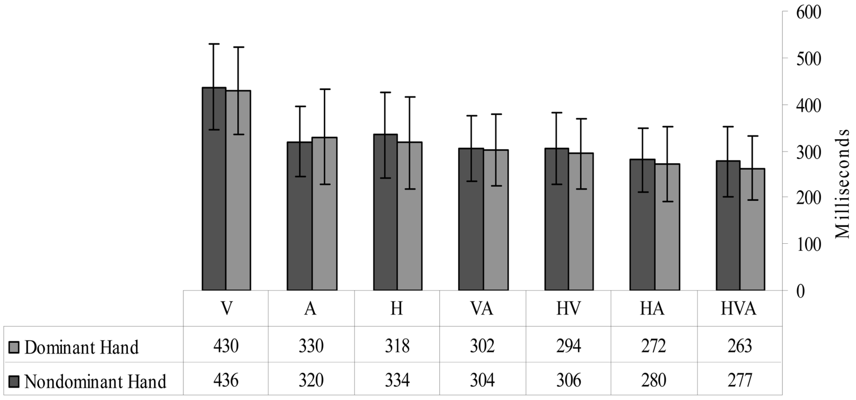
Sixteen students (11 males, 5 females, mean age 25.5 years) with a minimum of 12 years education participated in this study. All had normal hearing and normal or corrected to normal vision. They were paid for their participation, and were unaware of the purpose of the experiment, except that it was related to a study on hand-eye coordination. The experiment was carried out under the guidelines of the ethical committee and with its approval.

2.3 Stimuli

Seating in front of the computer system, participants were presented visually with two parallel green lines. Their task was to hold the stylus in their hand and move it by crossing these lines as if they were writing (see Figure 2). On each trial the computer generated a sensory stimulation, either *unimodal* [visual (V), auditory (A), or haptic (H)]; *bimodal*—a simultaneous combination of the visual and auditory (VA), the haptic and visual (HV), or the haptic and auditory (HA) stimulations; or *trimodal*—a simultaneous combination of the haptic, visual, and auditory (HVA) stimulations. The visual stimulus consisted of the two lines changing color from green to red. The auditory stimulus was a compound sound pattern (8 KHz, 560 ms) emitted from two loudspeakers located at both sides of the subject. The haptic stimulus was a resisting force (4 N) delivered through the stylus.

2.4 Procedure

Participants were instructed to react, by pressing a button on the stylus, as soon as they detected either one of the three stimuli or any of their combinations. Reac-



	V	A	H	VA	HV	HA	HVA
▨ Dominant Hand	430	330	318	302	294	272	263
■ Nondominant Hand	436	320	334	304	306	280	277

Figure 3. *Detection times of the unimodal, bimodal, and trimodal signals.*

tion time was measured from the beginning of the stimulation until the subject's reaction, and recorded by the computer. Participants used the same hand to move the stylus and to react (by pressing the button with their index finger). The other hand rested freely on the table.

In order to prevent participants from knowing and/or anticipating the exact timing of the stimulation, they were delivered randomly in the following manner. The computer counted each crossing (of both upper and lower lines) made by the subject and generated the stimulation, randomly, between the fifth and the thirteenth crossings. (For example, in the first trial, the stimulation was delivered immediately after the participant made his fifth crossing, in the second trial, the stimulation was delivered only after the twelfth crossing, and in the third trial, the stimulation was delivered after the tenth crossing, etc.). In this way, although the participants' movements triggered the stimulations, they were unaware of this arrangement so they could not predict the timing of the next stimulation and they continued to cross the lines until they were actually stimulated.

Before the beginning of the experiment, each participant was briefly trained how to perform his task. The experiment consisted of six blocks of trials, three performed with the dominant hand and three with the other hand. Each of these six blocks consisted of 105 single trials, in which each of the seven conditions (V, A, H, VA, HV, HA, HVA) appeared 15 times. All seven conditions were randomly intermixed in order to prevent participants from expecting a stimulus in a specific modality (Spence, Nicholls, & Driver, 2001), so in each block, every consecutive seven trials contained one trial of every condition, but their internal arrangement—within the seven trials—differed randomly (e.g., the initial seven were: A, HV, H, VA, HVA, V, HA, the next seven were: H, V, VA, HA, A, HVA, HV, etc.). The total number of trials for each subject was 630; 105 (trials) × 3 (blocks) × 2 (both hands).

3 Results

A repeated-measures ANOVA (GLM) indicated a significant main effect for modality combination in both the dominant [$F(6,10) = 32.71$, $p < .000$] and nondominant [$F(6,10) = 29.54$, $p < .000$] hands.

3.1 Unimodal Stimulation

Mean detection time of the unimodal stimuli were the longest. Response to the visual stimulus was at 430 ms post-stimulus ($SD = 94$) in the dominant hand and 436 ms ($SD = 92$) in the nondominant hand. Response to the auditory stimulus was at 330 ms ($SD = 103$) in the dominant hand and 320 ms ($SD = 76$) in the nondominant hand. Response to the haptic stimulus was at 318 ms ($SD = 99$) in the dominant hand and 334 ms ($SD = 91$) in the nondominant hand. See Figure 3 for a summary of the results.

3.2 Bimodal Stimulation

All three bimodal conditions were detected faster than any unimodal condition. Response to the audio-visual combination was at 302 ms post-stimulus ($SD = 78$) in the dominant hand and 304 ms ($SD = 70$) in the nondominant hand. Response to the haptic-visual combination was at 294 ms ($SD = 75$) in the dominant hand and 306 ms ($SD = 77$) in the nondominant hand. Response to the haptic-audio combination was at 272 ms ($SD = 81$) in the dominant hand and 280 ms ($SD = 69$) in the nondominant hand (see Figure 3).

Paired comparisons analysis between the unimodal and bimodal conditions revealed that: a) when participants received a bimodal combination of auditory and visual cues simultaneously, their RT was faster than the shortest of their unimodal component—auditory. The difference between these two conditions was highly significant, in both hands, [paired-$t_{(\text{va-a})}(15) = 3.60$, $p = .001$ in the dominant hand, and paired-$t_{(\text{va-a})}(15) = 3.72$, $p = .001$ in the nondominant hand]. b) When participants received a bimodal combination of haptic and visual cues simultaneously, their RT was faster than the shortest of their unimodal component—haptic. The difference between these two conditions was also highly significant in both hands [paired-$t_{(\text{hv-h})}(15) = 3.05$, $p = .004$ in the dominant hand, and paired-$t_{(\text{hv-h})}(15) = 3.4$, $p = .001$ in the nondominant hand]. c) When participants received a bimodal combination of haptic and auditory cues simultaneously, their RT was faster than the shortest of their unimodal component—haptic in the dominant hand, and auditory in the nondominant hand. The difference between these two conditions was also highly significant, in both hands, [paired-$t_{(\text{ha-h})}(15) = 5.64$, $p < .000$ in the dominant hand, and paired-$t_{(\text{ha-a})}(15) = 5.27$, $p < .000$ in the nondominant hand].

3.3 Trimodal Stimulation

RT in the trimodal combination was the shortest of all seven conditions—263 ms ($SD = 69$) in the dominant hand, and 277 ms ($SD = 76$) in the nondominant hand (see Figure 3). Paired comparisons analysis between bimodal and tri modal conditions revealed that when participants received a trimodal combination of haptic, visual, and auditory cues simultaneously, their RT was faster than the shortest of their bimodal component—haptic and auditory. The difference between these two conditions was significant in the dominant hand [paired-$t_{(\text{hva-ha})}(15) = 2.2$, $p = .02$] but not significant in the nondominant hand [paired-$t_{(\text{hva-ha})}(15) = 0.51$, $p = .30$].

3.4 Between-Hands Comparisons

Paired t-tests comparing RT between hands in all three unimodal conditions revealed insignificant differences (p values well above .1) between the dominant and the nondominant hands. Similarly, between hands comparisons in the bimodal conditions revealed insignificant differences in the audio-visual and haptic-audio conditions, and only a marginal difference ($p = .05$) in the haptic-visual condition. However, in the trimodal condition, there was a clear difference between the hands [paired-$t_{(\text{Dominant-Nondominant})}(15) = 2.49$, $p < .01$].

4 Discussion

These results provide evidence for a clear mental processing speed advantage (shorter RT) in all three bimodal stimulations (VA, HV, HA) over any unimodal stimulation (V, A, H). This advantage appeared in both hands. Furthermore, these results suggest a special trimodal (HVA) processing speed advantage over all three bimodal conditions, at least in the dominant hand.

Faster responses to bimodal signals, compared to unimodal signals, had been previously reported in visual-auditory combinations (Fort, Delpuech, Pernier, & Giard, 2002; Doyle & Snowden, 2001) and in haptic-visual combinations (Honore, Bourdeaud'hui, & Sparrow, 1989; Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002). This study extends the phenomena to haptic-auditory combinations and also to a trimodal (haptic-visual-auditory) combination. Furthermore, whereas in each of the aforementioned studied different types and intensities of visual, auditory, and haptic stimuli were tested, making comparisons and generalizations across studies difficult, in all seven conditions of the present study, the same stimuli (characteristics and intensities) were tested uni-, bi-, or trimodally. This extends the validity of our results beyond the stimuli specifications as a general principle that simultaneous trimodal signals are processed faster than simultaneous bimodal signals, which in turn, are processed faster than unimodal signals.

Electroencephalography (EEG) recordings, comparing unimodal and bimodal simple detections, revealed that bimodal audio-visual interactions start within 45–85 ms post-stimulus in the occipitoparietal visual cortex (Molholm et al., 2002; Fort et al., 2002), and that haptic-visual interactions start within 80 ms post-stimulus in the somatosensory cortex (Taylor-Clarke, Kennett, & Haggard, 2002). Psychophysical studies on signals intensity ratings showed that observers tend to rate a near-threshold light as brighter when presented with a low-intensity burst of white noise than when presented alone (Stein, London, Wilkinson, & Price, 1996) and that noise presented with light is rated as louder than the same noise presented without the light (Odgaard, Arieh, & Marks, 2004).

This psychophysical evidences for perceptual enhancement of bimodal signals and the neurophysiological evidence for intersensory interactions at the very first perceptual stages (long before the elicitation of any motor component), together with the fact that participants' motor response—pressing the (same) button on the stylus—was kept constant across all seven conditions, while the only manipulated variable was the perceived stimulus (or combined stimuli), suggests that this study's RT advantage reflects not merely a faster motor response, but also (if not only) a faster perceptual process of bimodal and trimodal signals.

This faster processing of multimodal events, starting at the very first initial perceptual stages of simply detecting the signals (in this study, the first 250–450 ms) allow, users of multimodal VE, more time at the consequent cognitive stages, enabling them to creatively fill in missing information and form a richer experience. For instance, in processing an event that lasts a similar time period, person A, stimulated by a single-sensory technology, is processing the incoming information slower than person B, stimulated by a multimodal VE. Thus, in a similar exposure time, person B finishes the initial perception processing-stage faster and can advance much further in the cognitive stream by paying attention to many more details and subtle cues in the graphic-auditory-haptic display and creatively integrate all these cues, by filling in missing cues, from one sensory modality with cues from another modality (Biocca et al., 2001). This longer and detailed process may be behind the greater sense of presence of multimodal VE.

In addition to the extended time that multimodal VE users have to absorb and integrate all sensory cues to a unified experience compared to unimodal technology users, the entire cognitive process in multimodal VE may be qualitatively richer since users' level of attention and awareness is greater. We raise here the hypothesis that the observed differences in RT between uni-, bi-, and trimodal signals result from different levels of attention that our brain allocates for processing uni-, bi-, and trimodal signals. Namely, the attention resources allocated by the brain to incoming multimodal stimuli are greater. This enhanced attention improves awareness and consciousness in general, and behaviorally this is expressed in faster responses.

4.1 Multimodal Stimulation and Attention

Several studies suggest that multisensory enhancement is modulated by attention. For instance, orienting attention involuntarily to the location of a sudden sound improves perception of subsequent visual stimuli that appear nearby (McDonald, Teder-Sälejärvi, & Hillyard, 2000). This effect had been reported even for subthreshold masked visual stimuli (Frassinetti, Bolognini, & Làdavas, 2002). The opposite effect—improved detection of an auditory signal if it is accompanied with a concurrent light—is also documented, even when the visual stimulus was completely task-irrelevant (Lovelace, Stein, & Wallace, 2003). Similarly, a sudden touch on the hand shifts spatial attention to the hand, and vision near that location is improved (Macaluso, Frith, & Driver, 2000). Reports on infant perception showed that information presented redundantly and in temporal synchrony across two sensory modalities (audio-visually) attracts infant attention and facilitates perceptual learning more effectively, so that young infants can detect a change in the tempo and the rhythm of an event when they experience the event bimodally, but not when they experience it unimodally (Bahrick, Flom, & Lickliter, 2002). This enhanced attention for the amodal properties when stimulated multimodally may be retained in adulthood in a novel or particularly difficult situation, and it is likely to have significant implications on perception, learning, and memory (Bahrick, Lickliter, & Flom, 2004). An fMRI study even suggested specific brain regions and neural networks that are involved in directing attention to multimodal (visual, auditory, and haptic) changing events in the sensory environment (Downar, Crawley, Mikulis, & Davis, 2000).

Based on this evidence, we hypothesize that the brain allocates a greater amount of attention to multimodal stimuli activating several (two or more) sensory modalities simultaneously and employing a larger neural network compared to the relatively lower level of attention that a single sensory stimulation, activating a limited neural system, draws for processing. This enhanced attention differentiates the entire experience in multimodal VE as qualitatively richer than single-modal technologies.

4.2 Attention and Presence

The link between attention and presence is already documented in the literature. When attention is maximal, there is a greater sense of presence (Witmer & Singer, 1998; Darken, Bernatovich, Lawson, & Peterson, 1999; Waterworth & Waterworth, 2001). Hence, the sense of presence may be greater in multimodal VE, since by activating several neural networks simultaneously they capture a maximum of their users' attention and receptiveness. This improved awareness and consciousness enables users to absorb much more detail and subtle cues from the display, integrate these stimuli more creatively, and, in interactive VE, respond more quickly. The results of this enhanced cognitive process are that the entire experience is felt as richer, more complete, and more coherent, and that may contribute to the greater sense of presence.

To conclude, multimodal VE provide users with a cutting edge information processing, even at the initial perceptual stage, at the beginning of the cognitive stream, since: a) multimodal information is processed faster; and b) multimodal VE activate larger neural networks, increasing users' attention. These clear advantages over single-sensory technology users, at the starting point, allow them, at the consequent cognitive stages, more time to: a) acquire a wider range of details and subtle cues from the display; b) fill in missing information from one sensory channel with cues from another sensory channel; and c) integrate all these informative cues, from the different senses, in an active and creative manner. As a result, the end product of this longer, detailed, and more active cognitive effort is a richer and more coherent experience, which may contribute to a greater sense of presence.

4.3 Implications for Designing Virtual Simulators

Designers of virtual driving and flying simulators may find special interest in this study, as these simulators can be upgraded by using *multiple* signals (visual, auditory, haptic, and proprioceptive) simultaneously. This is because in these simulators, one of the most important parameters for assessing driving and flying skills is the *time* that users need for detecting a car, a traffic sign, an object or a certain topographic view. Creating *multi*modal environments in which information is presented via multiple channels may significantly shorten response times.

These *multi*modal simulations may be especially useful to teach and assess driving and flying during limited-vision conditions such as twilight, night, around sharp curves and turns on the road, and so forth, as users can amplify the weak visual data and fill it in with appropriate auditory, proprioceptive, and haptic cues.

Acknowledgment

This research was funded by the European Union's FET initiative: *PRESENCIA*—Presence: Research Encompassing Sen-

sory Enhancement, Neuroscience, and Cognition, with Interactive Applications.

References

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology, 41*(4), 352–363.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science, 13*(3), 99–102.

Basdogan, C., Ho, C., Srinivasan, M. A., & Slater, M. (2000). An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction, 7*(4), 443–460.

Bernstein, L. E., Auer, E. T., Jr., & Moore, J. K. (2004). Audiovisual speech binding: Convergence or association? In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 203–223). Cambridge, MA: MIT Press.

Biocca, F., Kim, J., & Choi, Y. (2001). Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence: Teleoperators and Virtual Environments, 10*(3), 247–266.

Bonato, F., & Bubka, A. (2004). Visual/vestibular conflict, illusory self-motion, and motion sickness. *Journal of Vision, 4*(8), 798a.

Dalton, P., Doolittle, N., Nagata, H., & Breslin, P. A. (2000). The merging of the senses: Integration of sub-threshold taste and smell. *Nature Neuroscience, 3*(5), 431–432.

Darken, R. P., Bernatovich, D., Lawson, J., & Peterson, B. (1999). Quantitative measures of presence in virtual environments: The roles of attention and spatial comprehension. *Cyber-Psychology and Behavior, 2*(4), 337–347.

Doyle, M. C., & Snowden, R. J. (2001). Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location. *Perception, 30*(7), 795–810.

Downer, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience, 3*(3), 277–283.

Draper, M. H., Viire, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors, 43*(1), 129–146.

Duh, H. B. L., Parker, D. E., Philips, J. O., & Furness, T. A. (2004). 'Conflicting' motion cues to the visual and vestibular self-motion systems around 0.06 Hz evoke simulator sickness. *Human Factors, 46*(1), 142–153.

Forster, B., Cavina-Pratesi, C., Aglioti, S., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research, 143*(4), 480–487.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex, 12*(10), 1031–1039.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147*(3), 332–343.

Gottfried, J. A., & Dolan, R. J. (2003). The nose smells what the eye sees: Crossmodal visual facilitation of human olfactory perception. *Neuron, 39*(2), 375–386.

Held, R., & Durlach, N. (1992). Telepresence. *Presence: Teleoperators and Virtual Environments, 1*(1), 109–112.

Honore, J., Bourdeaud'hui, M., & Sparrow, L. (1989). Reduction of cutaneous reaction time by directing eyes towards the source of stimulation. *Neuropsychologia, 27*(3), 367–371.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. *Brain Research: Cognitive Brain Research, 17*(2), 447–453.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science, 289*(5482), 1206–1208.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*(6806), 906–908.

Mergner, T., & Rosemeier, T. (1998). Interaction of vestibular, somatosensory and visual signals for postural control and motion perception under terrestrial and microgravity conditions: A conceptual model. *Brain Research: Brain Research Review, 28*(1–2), 118–135.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Brain Research: Cognitive Brain Research, 14*(1) 115–128.

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2004). Brighter noise: Sensory enhancement of perceived loudness by concurrent visual stimulation. *Cognitive, Affective, and Behavioral Neuroscience, 4*(2), 127–132.

Palmisano, S., Gillam, B. J., & Blackburn, S. G. (2000). Global-perspective jitter improves vection in central vision. *Perception, 29*(1), 57–67.

Reiner, M. (2004). The role of haptics in immersive telecommunication environments. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(3), 392–401.

Romano, D. M., & Brna, P. (2001). Presence and reflection in training: Support for learning to improve quality decision-making skills under time limitations. *CyberPsychology and Behavior, 4*(2), 265–278.

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews: Neuroscience, 6,* 332–339.

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

Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments, 9*(5), 413–434.

Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception and Psychophysics, 63*(2), 330–336.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 8,* 497–506.

Taylor-Clarke, M., Kennett, S., & Haggard, P. (2002).Vision modulates somatosensory cortical processing. *Current Biology, 12*(3) 233–236.

Waterworth, E. L., & Waterworth, J. A. (2001). Focus, locus, and sensus: The three dimensions of virtual experience. *CyberPsychology and Behavior, 4*(2), 203–214.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.