Multi-Modal Stimulation, Response Time, and Presence

David Hecht[1,2], Miriam Reiner[1] & Gad Halevy[1]
[1] Technion – Israel Institute of Technology
Department of Education in Technology and Science
Haifa, 32000, Israel
[2] The University of Haifa
Haifa, 31905, Israel

davidh@techunix.technion.ac.il, miriamr@techunix.technion.ac.il, gadh@techunix.technion.ac.il

Abstract

Multi-modal virtual environments succeed better, than single-channel technologies, in creating the 'sense of presence'. We hypothesize that the underlying cognitive mechanism is related to a faster mental processing of multi-modal events. Comparing reaction time in uni-modal, bi-modal and tri-modal events, we show that the processing speed is: uni-modal > bi-modal > tri-modal. Given this advantage, users of multi-modal VEs start their cognitive process faster, thus, in a similar exposure time they can pay attention to much more informative cues. This results in a more rich, complete and coherent experience and a greater sense of presence.

***Keywords**: Multi-Modal, Visual, Audio, Haptic, Processing-Speed, Presence.*

1. Introduction

Multi-modal virtual environment systems able to combine efficiently sensory information from two or three channels (vision, audio, haptic) have an advantage in generating the 'sense of presence'. This "multi-sensory" experience differentiates them from older technologies, communicating only via a single sensory channel, which can generate only a limited degree of immersion and presence. Therefore, it is assumed that the more *multi*-modal a virtual environment is designed, the greater the sense of presence it generates [1] [4] [6] [7] [8] [11]. However, the *underlying cognitive mechanisms* in which multi-modal environments succeed to create 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 possible mechanism.

1) Environmental richness results in a complete and coherent experience

A rather intuitive idea suggests that a *single channel* media is relatively *sensory-poor* and conveys *limited and insufficient* information to the senses, thus it engenders only a lower sense of presence. Conversely, *multi*-modal environments provide a greater extent of sensory information to the observer. This *sensual richness* translates into a *more complete and coherent experience*. And therefore, the sense of "being there", in the virtual realm, is felt stronger [4] [8] [11].

2) Multi-modal VEs mimic "reality" better

Another way in which *multi*-modal environments succeed in creating a stronger sense of presence is by better mimicking "reality" [6]. An elaboration of this idea argues as follows: many of our *natural daily experiences* in the real world are fundamentally *multi*-modal by their nature, for instance, reaching to grasp an object or even simple posture and movement control are a co-production of visual, haptic and vestibular systems [5]. Communicating with another person through speech is a fine combination of producing and receiving audio and visual cues – sound, lip movements and gestures [10]. Our gastronomic pleasures result from a fine integration of taste, smell and vision [2] [3].

Therefore, *multi*-modal VEs have a clear advantage, in mimicking a multi-modal phenomenon, since they stimulate not only the user's *auditory* and *visual* sensory systems, but they do it with a *realistic 3D depth* perception. In addition, as a result of capturing the entire perceptual field (via head mounted display or 360^0 presentation) they stimulate also the *proprioceptive* and *vestibular* systems, as evidenced by the simulators sickness phenomenon and user's 'natural body movements' in virtual environments. The experience is especially felt as "real" if it includes also haptic sensations.

3) "Filling in" of missing information

Biocca et. al [1] proposed another mechanism which may help *multi*-modal virtual environments gain an edge in creating the sense of presence. They argue that the sensation of presence in virtual environments is related to the mind's attempt to integration. Since synthetic virtual environments 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, 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 and help ambiguous information from *another* sensory channel.

Thus, the process of *inter-modal integration* enables an inter-sensory *"filling in" of missing information*. This is a rather active and creative process, depending on the user abilities, and this *active* cognitive process results in an enhanced immersion into the virtual scene and a greater sense of presence.

4) Faster processing enables deeper and richer experience

While the above explanations - coherent experience, mimicking reality and filling in missing information - focus mainly on *higher cognitive functions*, occurring at the *end* of the *cognitive processing stream*, we suggest another possible mechanism which occurs *earlier* in *beginning* of the processing stream, at the *initial perception level*, which gives an advantage to multi-modal environments over single channel systems in creating the sense of presence.

Using a *simple reaction time* paradigm we compared the brain processing speed of *uni*-modal events (audio, visual or haptic) with the processing speed of *bi*-modal combinations of these signals and a *tri*-modal combination of these signals. Our hypothesis suggested an advantage, in processing speed, for bi-modal signals over uni-modal signals. Furthermore, we hypothesized that *tri*-modal signals will be processed *even faster than all bi-modal combinations*.

The rational for this study is that a processing speed advantage in multi-modal events may indicate a *greater focus of attention*, which may affect the entire event to be experienced as richer, more complete and coherent. In addition, a faster processing speed in the initial perceptual stage (at the first 300-400 msc.) allow users *more time in the consequent cognitive stages* enabling them to 'fill in' missing information and thus create a richer experience.

2. Experimental design

Materials

We used a touch-enabled computer interface which can generate for the users visual, auditory and haptic sensation. The haptic device (shown in figure 1) is based on a force-feedback mechanism which can generate haptic sensations felt by the user as a resisting force. Full technical descriptions of this system are available at: http://www.reachin.se and http://www.sensable.com.

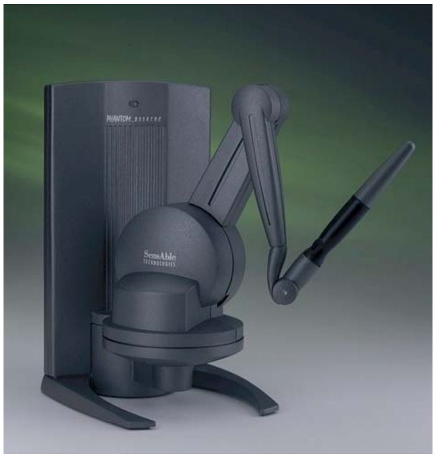


Figure 1: While users held the pen-like stylus (on the right) performing writing-like movements, the attached force-feedback mechanism generated a resisting force – haptic stimulation. Users responded by pressing a button on the stylus.

Participants

Sixteen students, 11 males and 5 females, (mean age - 25.5 years) participated in this study. They were recruited at the Technion, thus having had a minimum of 12 years education. All had normal hearing and normal or corrected to normal vision. They were paid for their participation but were *not unaware* of the purpose of the experiment, except that it has to do with eye-hand coordination. The experiment was carried out under the guidelines of the ethical committee and with its approval.

Stimuli

Seating in front of the computer system, participants were presented visually with 2 parallel green lines. Their task was to hold the stylus in their hand and move it by crossing these lines as if they are writing (see figure 2). On each trial the computer generated a sensory stimulation, either *uni-modal* (visual (V), auditory (A) or haptic (H)), *bi-modal* - a combination of the visual and auditory (VA), the haptic and visual (HV) or the haptic and auditory (HA) stimulations, or *tri-modal* – a combination of the haptic, visual and auditory (HVA) stimulations. The visual stimulus consists of the 2 lines changing their color from *green* to *red*. The auditory stimulus was a compound sound pattern (8 KHz, 560 msc.) emitted from 2 loudspeakers located at both sides of the subject. The haptic stimulus was a *resisting force* (4 Newton) delivered through the stylus.

Procedure

Participants were instructed to react, by pressing a button on the stylus, as soon as they detect *either* one of the three stimuli or *any* of their combinations. Reaction 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 the index finger). The other hand rested freely on the table.

In order to prevent participants from knowing and/or expecting 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 5th and the 13th crossings. (For example, in the 1st trial, the stimulation was delivered immediately after the 5th crossing, in the 2nd trial, the stimulation was delivered only after the 12th crossing. and in the 3rd trial, the stimulation was delivered after the 10th crossing etc.). In this way, although the participants' movements triggered the stimulations, they were not aware of this arrangement so they could not predict the timing of the next stimulation, thus, they continued to cross the lines until they were actually stimulated.

Before the beginning of the experiment, each subject was trained briefly how to perform his task. The experiment consisted of 6 blocks of trials, 3 performed with the dominant hand and 3 with the other hand. Each of these 6 blocks consisted of 105 single trials, in which each of the 7 conditions (V, A, H, VA, HV, HA, HVA) appeared 15 times. All 7 conditions were randomly intermixed in order to prevent participants from expecting a stimulus in a *specific modality* [9] so in each block, every consecutive 7 trials contained one trial of every condition, but their *internal* arrangement - within the 7 - differed randomly (For instance, the initial seven were: A, HV, H, VA, HVA, V, HA, the next seven were: H, V, VA, HA, A, HVA, HV etc.). Total number of trials for each subject was 630 (105 (trials) x 3 (blocks) x 2 (both hands).

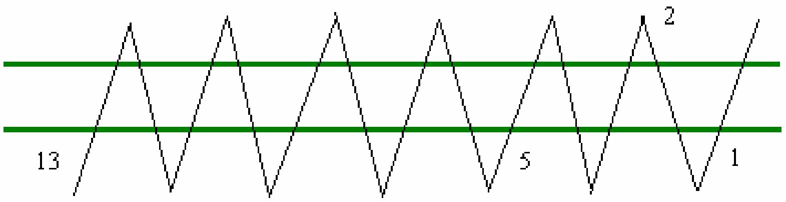


Figure 2: Participants performed writing-like movements with the stylus crossing the parallel green lines. Between the 5th and the 13th crossings, the computer generated, randomly, a sensory stimulation, either *uni*-modal, *bi*-modal or *tri*-modal.

3. Results

A repeated measures ANOVA (GLM) indicated a significant main effect for condition, in both, the dominant

[F(6,10) = 32.71 , p < 0.000] and non-dominant hands [F(6,10) = 29.54 , p < 0.000].

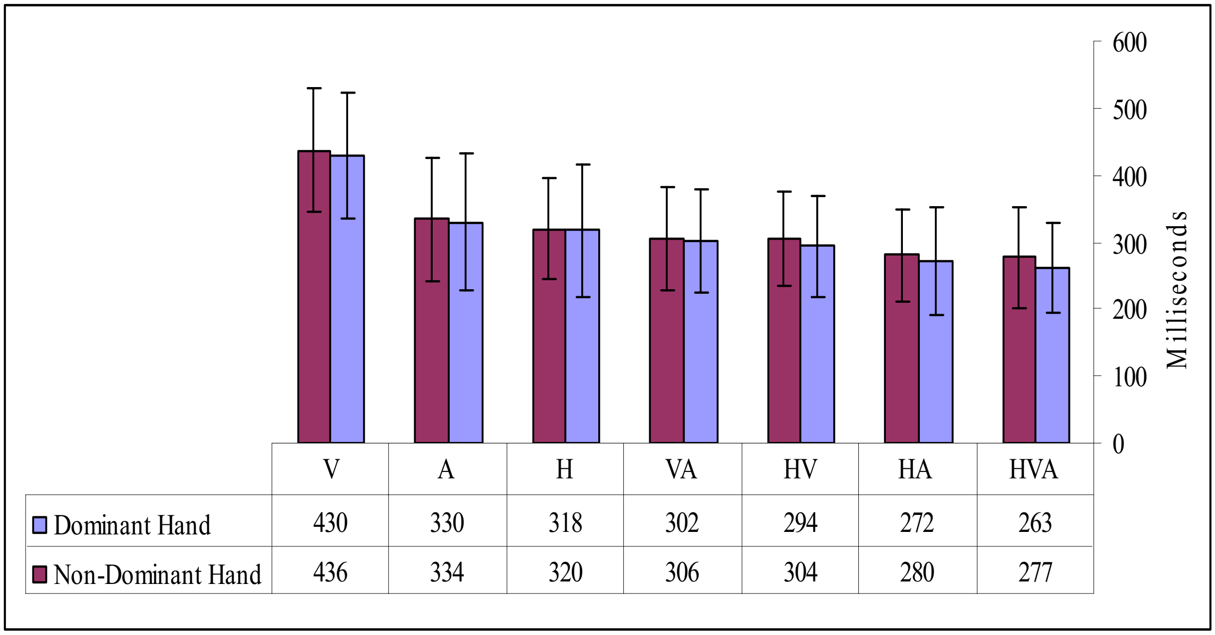
Dominant hand

Mean RT in the *uni*-modal conditions were the longest. 430 ms. for the visual stimulus, 330 ms. for the auditory stimulus and 318 ms. for the haptic stimulus. All three *bi*-modal conditions were *shorter than any uni-modal condition*, 302 ms. for the audio-visual combination, 294 ms. for the haptic-visual combination and 272 ms. for the haptic-audio combination. RT in the *tri*-modal combination was the *shortest* – 263 ms. See figure 3 for a summary of the results.

Paired comparisons analysis revealed that: a) When participants received a bi-modal combination of auditory and visual cues simultaneously, their RT [mean = 302, SD = 78] was faster than the *shortest* of their *uni*-modal component – auditory - [mean = 330, SD = 103]. The difference between these two conditions was highly significant [paired-t(15) = 3.60, p = 0.001]. b) When participants received a bi-modal combination of haptic and visual cues simultaneously, their RT [mean = 294, SD = 75] was faster than the *shortest* of their *uni*-modal component – haptic - [mean = 318, SD = 99]. The difference between these two conditions was also highly significant [paired-t(15) = 3.05, p = 0.004]. c) When participants received a bi-modal combination of haptic and auditory cues simultaneously, their RT [mean = 272, SD = 81] was faster than the *shortest* of their *uni*-modal component – haptic - [mean = 318, SD = 99]. The difference between these two conditions was also highly significant [paired-t(15) = 5.64, p < 0.000]. d) When participants received a tri-modal combination of haptic, visual and auditory cues simultaneously, their RT [mean = 263, SD = 69] was faster than the *shortest* of their *bi*-modal component – haptic and auditory - [mean = 272, SD = 81]. The difference between these two conditions was also significant [paired-t(15) = 2.2, p = 0.02].

Non-dominant hand

In the non-dominant hand, mean RT in the *uni*-modal conditions were also the longest. 436 ms. for the visual stimulus, 334 ms. for the haptic stimulus and 320 ms. for the auditory stimulus. All three *bi*-modal conditions were *shorter than any uni-modal condition*, 306 ms. for the haptic-visual combination, 304 ms. for the visual-auditory combination and 280 ms. for the haptic-auditory combination. RT in the *tri*-modal combination was the *shortest* – 277 ms. See figure 3.



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

Figure 3: Reaction times in the uni- bi- and tri-modal conditions.

Paired comparisons analysis revealed that: a) When participants received a bi-modal combination of haptic and visual cues simultaneously, their RT [mean = 306, SD = 77] was faster than the *shortest* of their *uni*-modal component – haptic - [mean = 334, SD = 91]. The difference between these two conditions was highly significant [paired-t(15) = 3.4, p = 0.001]. b) When participants received a bi-modal combination of visual and auditory cues simultaneously, their RT [mean = 304, SD = 70] was faster than the *shortest* of their *uni*-modal component – auditory - [mean = 320, SD = 76]. The difference between these two conditions was also highly significant [paired-t(15) = 3.72, p = 0.001]. c) When participants received a bi-modal combination of haptic and auditory cues simultaneously, their RT [mean = 280, SD = 69] was faster than the *shortest* of their *uni*-modal component – auditory - [mean = 320, SD = 76]. The difference between these two conditions was also highly significant [paired-t(15) = 5.27, p < 0.000]. d) When participants received a tri-modal combination of haptic, visual and auditory cues simultaneously, their RT [mean = 277, SD = 76] was faster than the *shortest* of their *bi*-modal component – haptic and auditory - [mean = 280, SD = 69]. However, the difference between these two conditions was *not* significant [paired-t(15) = 0.51, p = 0.30].

Comparison of RT *between hands* in each condition, revealed *insignificant* differences (P values well above 0.05) between the dominant and the non-dominant hand in all *uni*- and *bi*-modal conditions, except for the *tri*-modal condition, in which there was a clear difference between the hands [paired-t(15) = 2.49, p < 0.01]. These *preliminary* results are still under analysis, especially the apparent difference between the dominant/non-dominant hands. Nevertheless, the results indicate a clear enhancement in all three *bi*-modal conditions, as compared to the *uni*-modal conditions, in *both* hands.

4. Discussion

These results provide evidence for a clear processing-speed advantage in all three *bi*-modal stimulations (VA, HV, HA) over *any* uni-modal stimulation (V, A, H). This advantage appeared in *both* hands. Furthermore, the results suggest a *special tri*-modal (HVA) processing-speed advantage over all three *bi*-modal conditions, at least in the dominant hand.

From a neuro-cognitive perspective, a possible explanation of these phenomena may be that our brain allocates *greater attention* to events activating *several neural systems simultaneously*, in comparison to events activating fewer neural systems. This *enhanced attention* may be the factor beyond the faster processing of these multi-modal events.

Although, reaction-time measurements do not *directly* indicate presence, we suggest the possibility that *they both share a common factor - enhanced attention*. That is to say, *multi-modal virtual environments may achieve a greater sense of presence, since they employ their users' attention and receptiveness to its maximum.* This greater attentional focus enables them to absorb more details and subtle cues from the display and integrate them creatively. An enhanced attention leads at the end of the cognitive process to a richer, more complete and coherent experience, and possibly, a greater sense of presence.

In addition, the advantage of multi-modal events at the initial perceptual stage (at the first 260-300 msc.) allow users *more time in the consequent cognitive stages* to creatively 'fill in' missing information and form a richer experience. For instance, in processing an event which lasts a *similar* time period, person A, stimulated by a single channel environment, is processing the incoming information *slower* than person B, stimulated by a multi-modal environment. Thus, *in a similar exposure time person B finishes the initial perception stage faster and can advance much further in the cognitive stream by paying attention to much more details and subtle cues in the graphic/auditory/haptic display*.

Therefore, *multi*-modal virtual environments *provide their users with a 'cutting edge' already early in the perceptual stage*, in the beginning of the cognitive stream, since multi-modal informative cues are perceived *faster*. Their clear advantage over users of single channel technologies, in the *starting point*, allow them *more time at the consequent stages 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. c) Integrate these informative cues from different sensory systems in an active and creating manner. As a result, the end product of this *longer, detailed and active cognitive effort* is a robustly richer, more 'colorful' and coherent experience, and possibly, a greater sense of being present in the virtual scene.

Implications for VE 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 proprioception) simultaneously. Since, in these simulators, one of the most important parameters for assessing driving and flying skills is the time it takes users to detect a car, a traffic sign, an object or a topographic view, creating multi-modal environments in which information is presented via multiple channels may significantly shorten reaction time.

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

References

[1] 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 & Virtual Environments.*** 10(3); 247-266.

[2] Dalton, P., Doolittle, N., Nagata, H. & Breslin, P.A. (2000). *The merging of the senses: integration of subthreshold taste and smell.* ***Nature Neuroscience.*** 3(5): 431-2.

[3] Gottfried, J.A. & Dolan, R.J. (2003). *The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception.* ***Neuron.*** 39(2): 375-86.

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

[5] 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):118-35.

[6] Romano, D.M., & Brna, P. (2001). *Presence and Reflection in Training: Support for Learning to Improve Quality Decision-Making Skills under Time Limitations.* ***CyberPsychology & Behavior.*** 4(2); 265-278.

[7] Sanchez-Vives, M.V. & Slater, M. (2005). *From presence to consciousness through virtual reality.* ***Nature Reviews: Neuroscience***. 6: 332-9.

[8] Sheridan, T. B. (1992). *Musings on Telepresence and Virtual Presence.* ***Presence: Teleoperators and Virtual Environments***. 1(1); 120–125.

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

[10] Bernstein, L. E., Auer, E. T. Jr., Moore, J. K. (2004). *Audiovisual speech Binding: Convergence or Association?* In Calvert. G., Spence, C. & Stein B. E. (Eds.) ***The Handbook of Multisensory Processes.*** Pp. 203-223. (MIT Press).

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