The Influence of Spatial De-location on Perceptual Integration of Vision and Touch

Marco Congedo[a], Anatole Lécuyer[a], Edouard Gentaz[b]

a. National Institute for Research in Informatics and Random Systems (IRISA), Rennes, France.

b. National Centre of Scientific Research (CNRS), University "Pierre Mendès-France", Grenoble, France.

Correspondence and requests for materials should be addressed to Anatole Lécuyer, IRISA, Campus de Beaulieu, F-35042 Rennes Cedex, (e-mail: anatole.lecuyer@irisa.fr).

Acknowledgements. This work was supported by grants from the ROBEA French Program and French DGA. The authors would like to thank all the participants who took part in the experiment for their kindness and their patience and to Dr. Joe McIntyre for suggestions on a preliminary draft of the manuscript.

ABSTRACT

How do we perceive objects when what we see and what we touch is not at the same place? In a virtual environment, we observed that spatial de-location promotes visual dominance instead of visuo-haptic integration when judging the rotation angle of a hand-operated crank. Thus, the de-location of perceptual information appears to increase considerably the weight of the dominant sense at the expenses of the other. We relate this result to the design of teleoperation and virtual reality systems, in which, typically, the visual and haptic sensory information originates in spatially distinct devices.

INTRODUCTION

The interaction with the external world requires the combination and integration of more or less redundant information simultaneously acquired via our senses (Ernst & Bülthoff, 2004; Stein & Meredith, 1993; Welch, 1978). The process appears to be unceasing and effortless and can be conceived as the extraction of commonalities in the sensorial patterns (Ghahramani, Wolpert and Jordan, 1997). Typically, when exploring the properties of an object in *nature*, what we touch and what we see is a unique percept. Exceptions exist. For example, the hand that we watch through the mirror while combing our hair is just an *artificial* representation of the physical hand. Similarly, in many technological settings, visual and haptic (i.e., tactile and/or kinesthetic) information is provided by objects physically segregated. While we manipulate a computer mouse on the desk, we watch the cursor moving on the screen. In the teleoperation of a robot arm, the operator generally looks at a control monitor while manipulating an input device located outside the vision field. In these and many other artificial situations the brain is forced to associate spatially separated signals and unifies their percepts in order to ensure a coherent interpretation of the whole sensory experience. Therefore, for the design of teleoperation and virtual reality systems it is important to know how and to what extent the *spatial de-location* prevents the sensory integration process and/or affects the perceptual performance.

METHOD

To manipulate the spatial location of the visual and haptic percept we used a virtual reality apparatus known as “fish-tank” (Ware, Arthur & Booth, 1993). The participants manipulated an object while looking at its virtual representation reflected by a mirror located in between the eyes and the hand (Fig. 1, screen 2). Using stereovision we created the illusion of spatial collocation of the visual and haptic percepts. We compared this collocation situation to a de-location situation in which the visual information was provided by a computer monitor located on the left side of the hand workspace (Fig. 1, screen 1). Six participants were asked to judge the rotations of a handle grasped with the dominant hand and watched on the computer screen or in the mirror. We used a typical 2 forced-choice (2FC) psychophysical discrimination task (Wichmann and Hill, 2001), where the participants performed repeated trials to select the wider rotation angle between a standard stimulus and a comparison stimulus. The standard stimulus was equal to 36° and was randomly paired at each trial with one of five comparison stimuli. The smallest comparison stimulus was 4% wider than the standard stimulus and the remaining four were also in 4% increment (37.44°, 38.88°, 40.32°, 41.76°, and 43.2°). There were 24 repetitions of each standard-comparison pair in each of nine conditions. We investigated the bimodal sensory condition (vision and haptic) in collocation setting (VHc) and de-location setting (VHd), the haptic-only condition (subjects did not see the handle), and the visual-only conditions (subjects did not touch the handle) using either the computer screen or the mirror. There were also two additional conditions of *sensory conflict* for both VHc and VHd. Situations of sensory conflicts between two senses allow the estimation of the degree of visuo-haptic integration and the relative weight of each modality (Ernst and Banks, 2002). In the first conflict condition (visual bias) the visual rotation of the standard stimulus (α_V) was equal to 39.6° while the rotation of the real handle (α_H) was equal to 32.4°. In the second conflict condition (haptic bias), it was the opposite. In both cases, $(\alpha_V + \alpha_H) / 2 = 36°$. Given $\alpha_H \neq \alpha_V$ (perceptual conflict), the observed weights are given by (Ernst and Banks, 2002)

$$\begin{cases} w_V = \dfrac{PSE - \alpha_H}{\alpha_V - \alpha_H} \\ w_H = \dfrac{PSE - \alpha_V}{\alpha_H - \alpha_V} \end{cases}, \qquad (1)$$

where *PSE* is the observed *Point of Subjective Equality*. In a psychophysical study, the *PSE* is the value for which the comparison stimulus is perceived greater or lower than the standard stimulus with maximal uncertainty (50% of the times). For each value of comparison stimulus, the 24 repetitions are used to estimate the proportion of times the comparison angle is perceived wider than the standard angle. Fitting the observed proportions by a Gaussian cumulative distribution function (CDF), the *PSE* is found as the value corresponding to 0.5 cumulative density.

In total, each subject performed 1080 trials (5 standard-comparison pairs x 24 repetitions x 9 conditions). Within each condition, the sequence of the trials was randomized and so it was the sequence of the conditions. Each subject completed the experiment in five sessions, of which four lasted approximately one hour (allowing the completion of two conditions) and the last lasting approximately half an hour.

RESULTS

First, we compared the weights of the vision and haptic sensory channel in collocation and de-location condition. On the average of haptic and visual conflicts, the weight of the visual channel (Eq. 1) increased of nearly 30% in the de-location condition (98%) as compared to the collocation condition (77%). The increase was stronger for the visual bias conflict ($z=3.44$; $p<0.0006$) as compared to the haptic one ($z=2.76$; $p<0.006$). These results suggest that when the visual and haptic signals are spatially de-located, vision tends to capture touch. Indeed the weight of the visual information became nearly close to 100%, implying that the subjects relied almost exclusively on vision to form their judgments.

Secondly, we compared the performance of the subjects in the unimodal and bimodal de-located and collocated conditions. This analysis concerned the five conditions with no sensory conflict. The slope of the psychophysical CDF is inversely related to the *Just Noticeable Difference* of rotation angles, hence it provides a measure of the *discrimination performance*; the larger the slope, the smaller the detectable difference between two stimuli. We also measured the time needed to discriminate in the 2FC paradigm; its inverse provides a measure of *efficacy performance*. The performance in both discrimination accuracy and speed was significantly lower in the haptic-only modality as compared to all other modalities. Nonetheless, for both performance indexes, there was no significant difference between the VHc and VHd bimodal conditions, or in any other pair-wise comparison (Fig. 2). These results were confirmed further by subjective judgments on the difficulty of the 2FC discrimination task provided by the participants at the end of the experiment (data not shown). On the average, the six participants rated the task using haptic information alone more difficult than in visual-only or bimodal (both VHc and VHd) conditions, whereas there was no significant difference in the difficulty rating between the VHc and VHd condition, nor in any other pair-wise comparison.

CONCLUSIONS

Reviewing the literature, Hatwell, Streri and Gentaz (2003) conclude that in spatial tasks, where vision provides a richer and more economical source of information, visuo-haptic perception is not more efficient than vision-alone. Accordingly, in spatial tasks in which the visual and haptic percepts are spatially *collocated* and conflict is present, vision has been found to partially dominate touch (e.g., Rock & Victor, 1964). In a previous study (Ware & Rose, 1999) it has been reported similar accuracy in the rotation and translation of laterally displaced and co-located virtual objects, although the authors found that in the de-location condition participants needed more time to perform a rotation task. In the present study we employed a spatial task in which both vision and touch were engaged and we enforced spatial de-location of the visual and haptic percept in a virtual environment. In summary, our results confirm that for a spatial task in a collocated setting vision dominates touch so that the visuo-haptic bimodal perception is not more accurate than the vision-alone setting.

Our main interest was the study of the bimodal integration when the visual and haptic percepts are *de-located.* We have found that in this case the visual information becomes even more dominant. In other words, in the integration process the weight of vision becomes stronger. Actually, in our experimental setting and for our task, vision overwhelmed haptic almost completely. We conclude that for spatial tasks, de-location of visual and haptic percepts seems to prevent visuo-haptic integration and favors the dominant sense.

DISCUSSION

On the whole, our results on the judgment of a handle's rotation imply that even if the overall performance may not be significantly affected by the spatial de-location of the visual percept, the natural dominance of vision over touch is greatly increased. The designers of teleoperation systems may relate these findings to the task-specific relevance of the two sensory channels. For instance, if the contribution of touch is important for the task, great efforts should be undertaken to collocate as much as possible the visual and haptic percept. In this way the haptic weight would not be penalized. On the other hand, suppose the designer is making use of a low-quality haptic device. In this case the interest may be to contain the limitations of the haptic feedback. Then the usual spatial de-location of teleoperation systems may be of no concern or could be even exaggerated. As a result, the user will rely more on vision.

REFERENCES

Ernst, M. O., & Banks, M. S. (2002) Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion. *Nature, 415*, 429-433.

Ernst, M. O., & Bülthoff, H. H. (2004) Merging the senses into a robust percept. *Trends in Cognitive Neuroscience, 8*(4), 162-169.

Gharamani, Z., Wolpert, D.M., & Jordan, M.I. (1997) Computational Models of Sensorimotor Integration. In P.G. Morasso and V. Sanguinetti (Ed) *Self-Organization, Computational Maps and Motor Control* (pag. 117-147), Elsevier Press, Amsterdam.

Hatwell, Y., Streri, A., & Gentaz, E., (2003) *Touching for knowing*. Amsterdam: John Benjamins Publishers.

Rock, I., & Victor, J. (1964) Vision and touch: An experimentally created conflict between the two senses. *Science, 143*, 594-596.

Stein, B. E., & Meredith, M. A. (1993) *The merging of the senses*: MIT Press.

Ware, C., Arthur, K., & Booth, K. S. (1993) Fish tank virtual reality. In S. Ashlund, A. Henderson, E. Hollnagel, K. Mullet, & T. White (Ed.) *Proceedings of the ACM Conference on Human Factors in Computing* (pag. 37–42), New York: ACM Press.

Ware, C., Rose, J. (1999) Rotating Virtual Objects with Real Handles. *ACM Transactions on Computer-Human Interactions, 6*(2), 162-180.

Welch, R.B. (1978) *Perceptual Modification: Adapting to Altered Sensory Environments*. Academic Press, New York.

Wichmann, F.A., Hill, J (2001) The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit, *Perception and Psychophysics, 63*(8), 1293-1313.

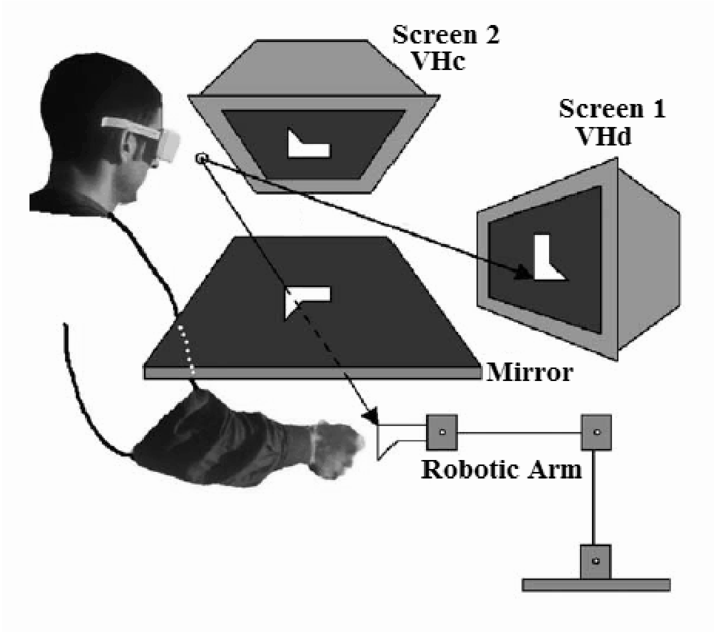

Figure 1: Experimental Materials and Apparatus. Participants grasped with the dominant hand the handle of a force feedback device (VIRTUOSE, Haption) beneath the mirror and judged the handle's rotation angles. They could not see their hand. The handle was set in rotation around its main axis at a constant speed (passive task). The handle first rotated to the final angular position and then rotated back to the neutral position. In the *collocation* setting they watched the reflection of the 3-D textured visual representation of the handle presented binocularly via a cathode ray tube screen positioned upon the mirror (Screen 2). Shutter glasses (CristalEyes, Stereographics) were used to enforce binocular disparity and the perspective was adjusted so that the visual percept was exactly superimposed to the real handle. In the *de-location* situation they also watched the handle with stereovision, but on another screen positioned in front of them and on the left of the hand workspace (Screen 1). VHc: collocation setting. VHd: de-location setting.

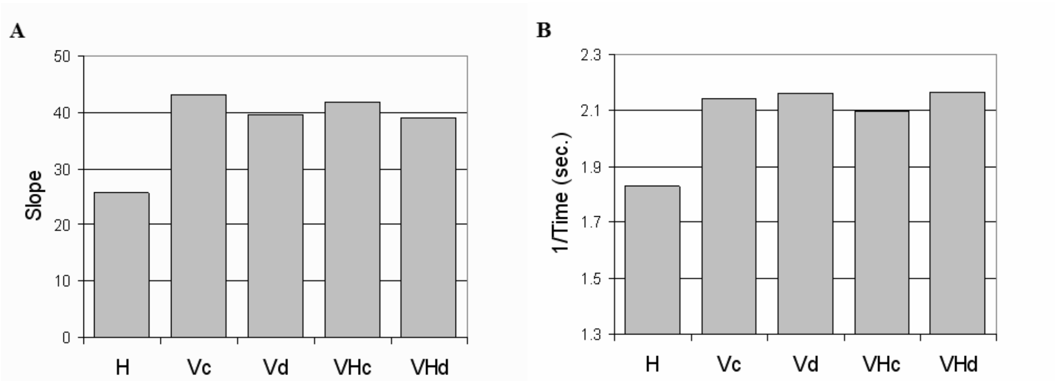

Figure 2: Results for Discrimination and Efficacy Performance. Average (across the six subjects) slope of psychometric functions (A: discrimination performance) and inverse of reaction time in seconds (B: efficacy performance) for the five experimental conditions with no sensory conflict. Psychometric functions were obtained as the best fit of Gaussian cumulative distribution functions to the proportion of responses “the comparison angle is wider” (Wichmann and Hill, 2001). The reaction time refers to the average time needed to select the wider angle in each condition. Visual inspection reveals that the pattern is practically identical for the two measures of performance. For both A and B, all conditions were pair-wise contrasted by means of within-subject student t-tests, all compared to the theoretical Student t-density function with five degrees of freedom. In both cases the only significant t-tests were those comparing the haptic-only condition with all the others. Legend for Experimental Conditions: H=Haptic-only (unimodal); V=Visual-only (unimodal); VH=Visuo-Haptic (bimodal); c=collocated; d=de-located.