

The Binding Problem in Presence Research

Abstract

An important contributor to the feeling of being present is the unity of one's perceptual experience. That is, the constellation of sensorial cues in a virtual environment must be in accord with some basic rules which, in the real world, govern the relationship between sensory events. A similar and long standing problem in neuroscience is how is it that the temporally and spatially segregated activity in neuronal ensembles is reassembled in order to generate a seamless conscious experience. This issue, which was first addressed by the Gestalt psychologists, is commonly referred to as the binding problem. In this paper we will discuss how the problem of binding is related to the problem of presence, and how the study of the neurophysiological substrates subserving this process may lead to an understanding of contextual relationships critical to generating presence in virtual environments.

I Introduction

In an effort to understand the neurological processes that are necessary and sufficient for the induction of presence in virtual environments (VE) we considered whether or not we needed a unique and new definition specifically for VEs, or whether these processes could be legitimately studied in an existing framework using traditional methods. Most definitions of presence echo the common theme that presence is the sense of “being there” (Held & Durlach, 1992; Sheridan, 1992) or the “subjective experience of being in one place or environment, even when one is physically situated in another” (Witmer & Singer, 1998). A critical review of the concept can be found in Draper, Kaber, and Usher (1998) and a more recent review of the literature and experimental evidence considers presence operationally to be

the extent to which people respond and act within a VE as if it were real—where response is taken at every level from physiological to cognitive (Sanchez-Vives & Slater, 2005).

On the face of it, it would seem to be highly unlikely that the brain mechanisms serving such a basic process would be different in or out of a VE. That is, the same processes responsible for generating responses and behaviors corresponding to sensory stimuli in the real world should also be at work in a virtual setting. As people move about their environment—virtual or otherwise—they are continuously bombarded with disparate streams of sensory input that must somehow be reassembled into a coherent whole. The problem of how the brain effortlessly uses spatially and temporally segregated activity in neuronal ensembles to form a unified perception is broadly defined as the *bindingproblem*. Since the goal of presence research is to understand how to achieve a unified perception in the context of a VE, the question arises: To what extent is presence related to binding? This is especially interesting because a naïve belief would be that very high realism in a VE is required for the achievement of high presence. Anecdotally it is known that this is not the case, and there is some experimental evidence indicating that high visual realism is not a necessary condition (Zimmons & Panter, 2003). If presence can be subsumed under the concept of binding, this would also help to explain why high realism is not a requirement.

The study of the binding problem arose from the work of the Gestalt psychologists who were interested in what rules governed object perception. Prior to Gestalt theory it was widely believed that perception was

Michael A. Harvey
Maria V. Sanchez-Vives*

Instituto de Neurociencias de Alicante
 Universidad Miguel Hernández-CSIC
 Campus de San Juan Apartado 18
 03550 San Juan de Alicante, Spain

*Correspondence to mavi.sanchez@umh.es

achieved by summing individual elements (sensations) in order to form a percept. However, structuralist ideas could not explain the perception of non-existent elements such as those that occur during apparent motion (Wertheimer, 1923). This led to the now classic idea that our perception is not solely governed by the summation of discrete areas of neuronal activity, but rather by some emergent property generated by the unique spatio-temporal constellation of that activity. Thus in its broadest sense the binding problem can subsume many cognitive functions, and ultimately may be a question of consciousness (Sanchez-Vives & Slater, 2005). These functions include: perceptual binding, the binding of stimuli within a sensory modality, for example, object form with object color; multimodal binding, the binding of stimuli between modalities, for example, the image and sound of a bat striking a ball; and temporal binding, the binding of events over time, for example, object motion. For instance the proper perception of a virtual baseball game includes all of the aforementioned elements. We use perceptual binding to bind the visual scene, that is, the brown belongs to the bat, the white to the ball, the uniform to the player, and so on; multimodal binding lets us place the sound of the bat striking the ball with the visual image of the event; and temporal binding grants us a perception of the ball moving smoothly through space, and not jumping from one point to another.

Another type of binding, that may in fact be the most important to presence research, is cognitive binding. A key element of cognitive binding is related to the fact that the rules governing the way in which a percept is formed can be learned. In most cases there is a remarkable consistency in the relationships between external and perceptual events, with these relationships being reinforced throughout the entirety of our perceptual lives. Thus when perceptual events are not in agreement, either with each other or with expectations based on prior experience, for example a very small girl speaking in a very deep voice, it produces in us a feeling of surprise or consternation, and in a virtual reality space, a break in presence (Slater & Steed, 2000).

Does this then imply that some breaks in presence are really failures of binding? Anecdotally we can think of

several examples where this would appear to be the case. Here is one: Imagine you are reaching out to touch a wall, and before you see your finger connect with the wall you feel a contact. This task requires the binding together of somatosensory, proprioceptive, and visual information. The early contact disrupts this process and simultaneously would almost surely disrupt our presence. A more extreme form of this happens in virtual reality in nearly every experience. A person reaches out to touch something, or collides with some object—and there is zero contact feeling associated with this, because of a lack of haptic feedback. However, it should hold true that conditions that do not disrupt binding are less likely to impact presence than those that do. For example, incongruous stimuli should be more disruptive than the absence of a stimuli. Why should this be? It is well known that in the absence of stimuli the brain is quite capable of doing a substantial amount of filling in (Ramachandran & Gregory, 1991). A good example is the blind spot generated in the visual field by the optic disc where the ganglion cell axons exit the eye. The brain continuously interpolates the visual scene and fills in this blind spot such that, normally, we are unaware of its presence. Following this logic if one were in a virtual environment, and bent over to smell a virtual rose, it should be less disruptive for the rose to have no smell than the wrong smell. When we see a rose we cognitively bind our visual representation of the rose to what we know about roses. There is plenty of anecdotal evidence to suggest that many people will claim to experience the smell of a rose when there is none actually generated. A wrong smell would interfere with this process whereas no smell would not, or at least not to the same degree. While this seems to be intuitively obvious, it has not been tested, but is one of the predictions we would make based on this model.

If it is true that presence bears some relation to how successfully we can bind the sensory and cognitive elements of a virtual environment, then an understanding of the neurophysiological mechanisms responsible for binding should be important to our understanding, and our ability to control, presence. The ultimate cellular and network mechanisms underlying the binding processes taking place in the brain are still a matter of dis-

cussion, and they will probably not be resolved any time soon. However, we believe that there are some concepts emerging from this basic neuroscience research from which one could generate hypotheses about perception and cognitive processes and how this relates to presence.

2 Hierarchical Processing

One of the original mechanisms proposed for binding came from the work of Hubel and Wiesel (1959), whose studies of visual processing resulted in a hierarchical model. Kuffler (1953) had described that ganglion cells in the retina, the first stage in the visual pathway, had circular receptive fields and therefore they responded best to small circles of light in a particular area of the visual field. Ganglion cells send their axons to the visual thalamus (lateral geniculate nucleus), whose thalamo-cortical neurons project to the next level of visual cortex, the primary visual cortex. Hubel and Wiesel found that while neurons in the visual thalamus had receptive fields similar to the retinal ones, cells in primary visual cortex responded best to bars of light or lines, and therefore were adequate to code for contours. This led them to suggest that the receptive fields of the first visual cortical cells to receive thalamic information, *simple cells* as they called them, were built by combining the inputs of thalamic cells, since a bar can be made by aligning circles together. Simple cells would then converge on *complex* cells, and complex cells would converge on *hypercomplex* (or end-stopped) cells, thereby building cells with receptive fields for increasingly complex stimuli. This vision resulted to a hierarchical model of visual perception, which supposes that the same details of the visual world are reanalyzed at increasing levels of complexity. A problem with this model is that at the top of the hierarchy there should exist cells that would be responsive to very specific stimuli, which has become known as the question of the “grandmother cell.” What it implies is that we should have individual detector cells for every conceivable object we can recognize, exemplified by the neuron that would code for the face of our grandmother. While cells with apparently very complex and specific receptive fields have been

found (Logothetis & Sheinberg, 1996; Tanaka, 1997) the catch is in the combinatorial explosion in the number of different percepts that need to be represented. It is likely that we simply do not have enough cells to account for the near infinite number of perceptual possibilities. This early work provided us with the conceptual building blocks for higher order perceptual processes; but whether the blocks are held together by cardinal cells or by other mechanisms is still an open question. The current view of visual processing is that simultaneous, parallel processing of different attributes of the visual world exists in the visual system, although a hierarchical strategy may characterize each of the parallel systems.

3 Selective Attention

This problem may in part be solved by selective attention. In one study subjects were shown a display consisting of variously colored letters. The subjects task was to correctly identify the letters they had seen and their respective colors. When subjects were allowed to attend to the display the letters and their colors were correctly identified. However when subjects were forced to split their attention between two displays they often made errors in assigning the appropriate color to a given letter. This experiment neatly illustrates both the importance of perceptual binding and a possible role for attention in its actualization (Treisman, 1982). When we look at a colored letter, although the color and the shape of the letter are represented in different visual areas, they share the same physical space. It is possible, therefore, that binding could occur by grouping cells that responded to stimuli in overlapping space. The problem with this idea is that many cells are responsive to stimuli over a large area of the visual field, which would be a large source of error. Therefore there would be frequent mismatches if this were the only mechanism responsible for forming appropriate shape/color conjunctions. This is where some researchers believe attention comes into play.

Attentional mechanisms could be used to select a particular stimulus among many that may lie within a neu-

ron's receptive field. Modulations have been shown in recorded neurons from area V4 (one of the areas specialized in object recognition) of awake behaving monkeys that had been trained to alternatively pay attention to different stimuli within receptive fields. Similar results have been shown in area MT, a visual area specialized in motion detection, where attention has been found to act as a gain control mechanism. Neurons activated by the attended stimulus showed increased gamma frequency synchronization (Fries, Reynolds, Rorie, & Desimone, 2001), therefore enhancing the representation of attended stimuli. One alternative theory of attention then is that it serves to shrink a cell's receptive field such that it only encompasses the attended stimuli (Moran & Desimone, 1985). The idea behind it is that some attentive signal, perhaps from higher cortical areas, impinges upon ensembles in visual cortex biasing responses and imparts greater salience to the attended stimulus. This allows cells with overlapping receptive fields to correctly conjoin stimulus features. This theory is well supported in the literature; however it cannot account for facts like the binding of cross-modal stimuli where the stimuli do not share the same spatial representation. Nevertheless we are well capable of correctly conjoining auditory and visual events, even with a number of distracters present.

4 Temporal Correlation

Perhaps one of the most powerful, and at the same time most controversial, theories of binding is that of temporal correlation. This theory suggests that activity in individual cells, or ensembles of cells, responsive to different elements of a common percept could be grouped together through the synchronization of their responses. Specifically prevalent during sensory stimulation (Gray, Engel, Konig, & Singer, 1990), alertness (Bouyer, Montaron, & Rougeul, 1981; Murthy & Fetz, 1992), and during information processing are fast (20–100 Hz) oscillations in the cortical and thalamic local field potentials (Llinas & Ribary, 1998). Phase locking of this oscillatory activity across cortical space could then provide the glue that would bind together the ele-

ments of a percept. There is evidence that fast, oscillatory activity in local networks could be generated by specialized neuronal oscillators (Nunez, Amzica, & Steriade, 1992; Gray & McCormick, 1996) and the mechanisms to generate fast rhythms already exist in the cortical microcircuit, since they have been characterized even in 400 μm thick *in vitro* cortical slices (Harvey et al., 2003). It has been proposed that these oscillations would depend critically on the time constant of the GABA_A inhibitory synaptic potentials (~ 10 ms) and there is an increasing literature on the importance of GABAergic (inhibitory) interneurons to the synchronization of fast oscillatory activity (Gibson, Beierlein, & Connors, 1999). However little is known about the mechanisms responsible for long range synchronization, axonal conduction delays prohibit a mechanism of synchronization based solely on cortical excitatory connections. The evidence that fast, synchronous activity in neuronal ensembles may play a functional role in binding has been suggested in studies of olfaction (Stopfer, Bhagavan, Smith, & Laurent, 1997), sensorimotor integration (Murthy & Fetz, 1992), and vision (Engel, Konig, Kreiter, Schillen, & Singer, 1992), as well as for attentive processes.

If synchronizing the activity between neuronal ensembles serves as a mechanism to group percepts, synchronizing connections must be susceptible to use-dependent modifications, and new instances of synchrony should arise based on repeated coactivation of these ensembles. It has long been thought that the cortical substrate for these integrative processes exist solely in higher association areas, while the role of the primary sensory cortex was to segregate input arising from different modalities. However there is mounting evidence that areas of sensory cortex, previously believed to be uniquely activated by a particular stimulus modality, are in fact responsive to stimuli from other modalities (for a review see Calvert, Spence, & Stein, 2004). Further, it has been shown that during critical periods of development cells in primary sensory cortices are capable or remapping their representation to include other sensory modalities (Sur, Pallas, & Roe, 1990). Such populations of multimodal neurons possessing experience dependent receptive fields seem likely candidates to form the foun-

dation for higher order integrative processes. We believe that the mechanism for temporal, perceptual, and cognitive grouping may share common features, and that the study of perceptual binding in particular provides a good starting point for the investigation of the neuronal substrates responsible for generating a unified perceptual experience.

5 Conclusions

The integration of the notion of binding with that of presence is useful to both presence research and to the understanding of the brain processing of information. The different aspects of the problem of binding that we have analyzed here provide a theoretical framework for making predictions regarding presence testable in an experimental setting, preferably a virtual environment. Fill-in strategies used by the brain can supplement the lack of rich stimuli, resulting in the sufficiency of minimal cues to provide a high sense of presence. However, incoherent stimuli will alter the binding between top-down and bottom-up mechanisms, resulting in breaks in presence. Alterations in the correct temporal structure of the stimuli, such as delays in the display or too high latencies between stimuli, are bound to decrease the sense of presence, while dramatic alterations in the quality of the display that is not being attended to should not make a noticeable difference in the presence experienced by the subject. Thus we can already see how an understanding of the neurophysiological correlates of binding can facilitate the construction of virtual environments tailored for the necessary and sufficient elements that would generate a perceptual and cognitive coherent experience rather than a highly realistic representation of the world. As we continue to understand more about the physiology of perceptual processes, more of these elements should become apparent, allowing us to construct environments that will evoke higher presence. Similarly, the ability to disassociate perceptual variables (inseparable in the real world) in virtual environments will allow us to unveil these perceptual elements all the more rapidly. We believe that, ultimately, it should be possible to step from the real world into a

virtual environment without a loss of presence and that the actualization of this ideal will benefit most not from the mutually exclusive pursuit of either greater computing power and software, nor through the final understanding of the cellular and network mechanisms of binding, but rather through the fusion of these studies, with each gaining benefit from the other along the way. This is the current goal of the research of the authors.

Acknowledgments

This work has been supported by PRESENCIA (European Union FET IST-2001-37927). We want to thank M. Slater and M. M. Arnold for helpful discussions.

References

- Bouyer, J. J., Montaron, M. F., & Rougeul, A. (1981). Fast fronto-parietal rhythms during combined focused attentive behavior and immobility in cat: Cortical and thalamic localizations. *Electroencephalogr. Clin. Neurophysiol.*, *51*, 244–252.
- Calvert, G., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.
- Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human Factors*, *40*, 354–375.
- Engel, A. K., Konig, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends Neurosci.*, *15*, 218–226.
- Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, *291*, 1560–1563.
- Gibson, J. R., Beierlein, M., & Connors, B. W. (1999). Two networks of electrically coupled inhibitory neurons in neocortex. *Nature*, *402*, 75–79.
- Gray, C. M., & McCormick, D. A. (1996). Chattering cells: Superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science*, *274*, 109–113.
- Gray, C. M., Engel, A. K., Konig, P., & Singer, W. (1990). Stimulus-dependent neuronal oscillations in cat visual cortex: Receptive field properties and feature dependence. *Eur. J. Neurosci.*, *2*, 607–619.

- Harvey, M. A., Puccini, G., Descalzo, V. F., Reig, R., Compte, A., & Sanchez-Vives, M. V. (2003). Fast oscillations during the up states of slow cortical rhythmic activity *in vitro*. In *Computational Neuroscience Society Proceedings*.
- Held, R. M., & Durlach, N. I. (1992). Telepresence. *Presence: Teleoperators and Virtual Environments, 1*, 109–112.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol., 148*, 574–591.
- Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol., 16*, 37–68.
- Llinas, R. R., & Ribary, U. (1998). Temporal conjunction in thalamocortical transactions. *Adv. Neurol., 77*, 95–102; discussion 102–103.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Ann. Rev. Neurosci., 19*, 577–621.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782–784.
- Murthy, V. N., & Fetz, E. E. (1992). Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys. *Proc. Natl. Acad. Sci. USA, 89*, 5670–5674.
- Nunez, A., Amzica, F., & Steriade, M. (1992). Voltage-dependent fast (20–40 Hz) oscillations in long-axoned neocortical neurons. *Neuroscience, 51*, 7–10.
- Ramachandran, V. S., & Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature, 350*, 699–702.
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci., 61*, 332–339.
- Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments, 1*, 120–126.
- Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments, 9*, 413–434.
- Stopfer, M., Bhagavan, S., Smith, B. H., & Laurent, G. (1997). Impaired odor discrimination on desynchronization of odor-encoding neural assemblies. *Nature, 390*, 70–74.
- Sur, M., Pallas, S. L., & Roe, A. W. (1990). Cross-modal plasticity in cortical development: Differentiation and specification of sensory neocortex. *Trends Neurosci., 13*, 227–233.
- Tanaka, K. (1997). Mechanisms of visual object recognition: Monkey and human studies. *Curr. Opin. Neurobiol., 7*, 523–529.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *J. Exp. Psychol. Hum. Percept. Perform., 8*, 194–214.
- Wertheimer, M. (1923). Untersuchungen zur Lehre der Gestalt II. *Psychol Forsch, 4*, 301–350. Translation published as *Laws of Organization in Perceptual Forms*, in W. Ellis, *A Source Book of Gestalt Psychology*. Routledge and Kegan Paul, London, 1938, pp. 1971–1988.
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*, 225–240.
- Zimmons, P., & Panter, A. (2003). The influence of rendering quality on presence and task performance in a virtual environment. *Proceedings of the IEEE Virtual Reality 2003*, p. 293. IEEE Computer Society.