The brain and interaction in a multimodal reality

Daniel Sjölie

UMINF-09.09
ISSN-0348-0542

Umeå University
Department of Computing Science
SE-901 87 Umeå, Sweden

The brain and interaction in a multimodal reality

Daniel Sjölie
daniel@cs.umu.se
Department of Computing Science
Umeå University
S-901 87 Umeå, Sweden

Abstract
This paper reviews the neural correlates of multimodal integration and the role it plays in the creation and maintenance of perception of reality. These issues are illuminated by reviewing concepts and research from a number of related subjects and we explore some of the relevant cognitive models, such as the *memory-prediction framework*. We further focus on how multimodal integration affects *reality-based interaction* (RBI) in general and *virtual reality* (VR) in particular. In this case the reality in question is generated by a computer and perception of reality may be unstable. In VR-related research the quality of the perception of reality is commonly referred to as presence and a review of the conditions for and effects of varying degrees of presence is presented. An increased understanding of the role of multimodal integration in the creation and maintenance of presence is one of the primary goals of this paper. The hope is that this will help us to understand and improve presence, something that we will show to be of great value. The effect of disturbances and failure in the multimodal integration on the perception of reality and presence is of particular interest. This is related to the concept of breaks in presence and prediction errors, to provide some framework for understanding. Also, the importance of understanding the neural correlates of these cognitive functions is related to the possible use of VR in combination with brain imaging, exemplified with *functional magnetic resonance imaging (fMRI)*. Finally, we discuss possible future work and possibilities to advance the understanding of the brain and reality in the context of human computer interaction.

The brain and interaction in a multimodal reality

The aim of this paper is to examine interaction in the context of perception and construction of reality from multimodal stimuli. Reality, as a concept, will be central throughout this paper. I believe that our perception of reality is fundamental to how our brain works, that this in turn affects how you should design efficient interaction paradigms and that the study of multimodal integration is an excellent way to approach these questions. Modalities, in this paper, can essentially be understood as the different perceptual senses, i.e., visual, auditory, tactile, etc.

One central idea is that the brain is an expert at dealing with the information contained in a representation of reality and that this makes it possible to create more efficient interaction systems. Areas reviewed demonstrate what the possibilities, conditions and problems with such an approach are and how we can proceed to investigate the issue further. This line of research can also be used to further study how meaning and representations are created from multimodal stimuli and providing a perspective firmly rooted in neuroscience.

Reality according to the brain

Intelligence and the brain have evolved to interpret reality and enable efficient interaction with reality. Based on the *memory-prediction framework* presented by Jeff Hawkins (2005), and the model for learning and inference presented by Karl Friston (2003), I believe that the fundamental principle at work is the ability to remember what has happened before, recognize a similar situation, and use the stored memories to make predictions about possible future events. In order to efficiently recognize meaningful aspects of the vastly varying situations of reality the brain needs to identify and classify patterns of coherency and correlations. If events always happen at the same time, at the same place or in a consistent relative position in space and time we can gain much by remembering this as an integrated concept. E.g., if we can integrate the visual sensation of the color of a ripe orange with the sensations of texture and hardness in the same spatial location this helps us predict that this is a good source of food and we can generate an expectation to experience a certain taste if we act to eat it. Hawkins calls these integrated concepts "invariant representations" since the defining property is that the same relationships and correlations can be recognized in many different forms and context, i.e., they do not vary. In the more formal model of Friston these representations are treated as *causes*, corresponding to a model in the brain of what has caused the sensations in question. E.g., the ripe orange is a possible cause of all the sensations described above. A very similar reasoning is also used by Dileep George and Hawkins when they develop the *hierarchical-temporal memory* platform (George & Hawkins, 2005) based on Hawkins ideas. Friston and George both focus on the importance of hierarchies, Bayesian probability networks and time as a teacher.

It should be obvious from the example with the ripe orange above that the integration of multiple modalities plays a key role in the initial creation of these invariant representations and in laying the foundation for an understanding of reality. The concept of understanding is

intimately tied to the ability to make predictions. If you are able to predict what can happen in a certain situation or with a certain object you have an understanding of the workings involved. In contrast, if something happens that you could not predict (not even as a possibility) you become confused and your understanding of the situation collapses. This view of understanding and reality is connected to a number of other relevant concepts in later sections of this paper. To quote Hawkins, "predictability is the very definition of reality", and when I use the word reality in this paper it should be read as "a consistent world that can be understood and predicted based on experience and recognition of patterns", in particular such a world that allows us to intuitively use our brain (evolved in this physical world) to recognize patterns and make predictions. It may be clarifying to consider the opposite of reality, i.e., the *unreal*. If something is unreal it means that it does not fit into your understanding of the world, it is inconsistent with the patterns you have learned to recognize and you have no basis for making predictions about this phenomenon.

The level of understanding introduced above is largely shared by all mammals. It is a feature of the neocortex and even a rat has invariant representations and the ability to make predictions. If a rat is allowed to navigate a maze several times and can always find a piece of cheese at the same location it will be able to recognize specific corners in the maze and predict that it will find a cheese if it takes a path that it remembers to have led to a cheese before. Similarly, according to Sverre Sjölander, if a dog sees a rabbit run behind a bush it has the ability to predict that the rabbit will probably come out at the other side and intercept it (Sjölander, 1999). Reptiles lack this ability and a snake that loses track of a mouse will not be able to predict what it might do and plan accordingly. Sjölander also remarks that we have no reason to believe that a snake has any concept of reality as a consistent, understandable, world; thereby supporting the definition of reality given above. The behavior of a snake shows no indication of any understanding of how mice work. The snake simply reacts to the sight or smell of the mouse and the reaction to such stimuli shows no evidence of integration between modalities. These examples illustrate the basic advantage of the neocortex that has driven its evolution and supported the success of mammals on earth.

In order to reach a higher level of understanding we need to introduce a hierarchical structure. The neocortex has a lot of hierarchical structure even at the lowest level (Friston (2003) includes a review), but the need now becomes apparent. In order for us to be able to understand what cars are and how they work with any measure of efficiency we need to build this understanding on other concepts (invariant representations). We need to understand that cars have an engine, tires, a body, doors and windows. This enables us to make efficient predictions about the capabilities of any specific car, without having to evaluate in detail if the round, black, cylinders below the car might be used to allow efficient transport. Greater understanding means an ability to efficiently predict the functions and uses of an object and a more complex hierarchy allows humans to understand and make use of more complex objects. We can predict a large number of specific ways to use wheels, doors or windows and this increases our understanding of cars. One notable effect of this reasoning is that the study of multimodal integration becomes even more interesting since the way in which the brain integrates different modalities to create invariant representations of objects that, e.g., sounds, looks and tastes in a consistent manner may very well have much in common with how the brain (the neocortex in particular) works on many (possibly all) levels.

One aspect of the implementation of this hierarchical prediction framework in the brain that both Hawkins and Friston emphasize the critical importance of is *feedback*, i.e., how higher levels in the hierarchy affect and prepare the lower levels based on predictions. The most obvious function of the hierarchy so far has been for each lower level to supply the higher level with invariant representations, used to construct even more complex representations on top of these. This corresponds to a forward feed of information, from our basic senses up to our perception of reality and complex objects. However, the brain is packed with feedback connections, distinct from the forward connections in both function and microstructure (Friston, 2003). In addition to demonstrating the fundamental hierarchical nature of the cortex this also suggests that these feedback connections must play an important part for the function of the brain. Friston argues formally that feedback is a necessary part of any model able to infer causes from input when the interpretation is contextually dependent. E.g., if you are walking along a street and see someone waving his arm, you cannot deduce if he is hailing a cab or swatting a wasp based only on observation of the arm movement (Kilner, Friston, & Frith, 2007). Hawkins starts by pointing out that we are very efficient at detecting anomalies and deviations from what we expect and continues to explain this phenomenon with feedback as a critical component. The idea is that feedback is used to prepare lower levels to receive the input that we expect, and that this allows the information going up the hierarchy to focus on the differences. If everything matches the expectations we don't need to change anything and working with assumptions like this, working with an internal model of reality, is what makes efficient thinking possible. This model is supported in a recent study (Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008) where the relation between repetition suppression and expectations is examined. Repetition suppression refers to the diminishing activation in response to repeated stimuli and influential theories has explained this effect as automatic consequences of the bottom-up flow of perceptual information (Desimone, 1996; Grill-Spector, Henson, & Martin, 2006). In this study, however, it is shown that the repetition suppression is modulated by the likelihood of a repetition and that the suppression was reduced when the repetition was unexpected. The authors explain this as consistent with a model of top-down predictions combined with bottom-up prediction errors, matching the model described above. In this case the interpretation of the result would be that repetition suppression in general corresponds to a diminished need to trigger prediction errors and that the unexpected repetitions did not match the expectations as well as the expected repetitions and thus triggered more prediction errors, i.e., more activity and a smaller repetition suppression effect. We return to the potential impact of these ideas on models for multimodal integration in later sections.

Finally, I should emphasize the importance of time in the frameworks presented here. As mentioned in passing above, both Friston and George/Hawkins focus on time and temporal sequences as the teacher allowing the brain to construct a model of reality without any (or little) prior knowledge. Temporal patterns are central to our ability to make predictions since we can only perceive a small part of the world at any moment. We need to recognize stimuli and events that are connected in time in order to be able to put these together as objects in the real world.

Reality-Based Interaction

The kind of interaction discussed in this paper is primarily interaction between humans and some kind of computer-controlled system. This subject is commonly known as Human Computer Interaction (HCI) and the concept of Reality-Based Interaction (RBI) was recently introduced in this context (Jacob et al., 2008). The central theme of RBI is that computers and the interaction styles should be adapted to humans and leverage the preexisting expertise of all humans to deal with the real world. The human brain has evolved to be able to understand and deal with reality and in many ways we can do this with great efficiency. By making interaction with a computer more like interaction with the real world it becomes possible to use familiar concepts to understand and predict the capabilities and functions of a computerized system. The Gulf of execution, the gap between a user's goal for action and the means to execute that goal, is reduced, and we can focus on *what* we want to do instead of on *how* each step must be executed. This means we can work with abstractions on a higher level, making it possible to control and interact with more complex systems. An increased understanding of how the brain constructs and maintains perception of reality and how multiple modalities are integrated in this process would be valuable in determining which methods are the most effective and what factors are potentially disruptive when designing a RBI-system. This is especially important, as different kinds of tradeoffs (Jacob et al., 2008) are often motivated in order to fulfill other goals or necessary because of technical limitations. Having control over what you are giving up and what you are gaining when doing such tradeoffs can be essential and this is discussed at length in a later section.

Another motivation for using reality as a basis for computer interaction is that the *ecological validity* can be increased. Ecological validity is important when you in some sense want to capture or reproduce a behavior or experience that is normally found in one specific setting, i.e., in one specific ecology, while you are actually present in another setting. One typical example is training simulators such as advanced flight simulators. In this case we want to reproduce the experience of flying a real airplane while avoiding all the dangers of doing this in the real world. A high ecological validity means that the constructed environment and interaction matches the one in the target ecology and, most importantly, that results and observations made in the constructed ecology are valid in the target ecology. Direct similarity between the settings is one method of gaining high ecological validity but exactly which aspects of a specific ecology are most important is not always obvious and depends on what you want to capture. In the case of a flight training simulator we are primarily interested in flying skill and we want to be confident that any skill at flying in the simulator means that you are also skilled at flying a real airplane. I.e., we want to be sure that the result carries over, or can be transferred, to the target ecology. In the specific context of training systems this effect is commonly known as *transfer*.

Ecological validity is also important for many kinds of research and medical purposes. For medical purposes we want to make sure that any systems intended for rehabilitation (a form of training) of patients have the desired effect in their everyday lives. Researchers need to consider ecological validity when doing testing in some form of controlled or limited environment (like a lab) and want to claim that any results are generally valid in the real world. This can be particularly challenging when studying the brain since it is hard to assert

that cognitive functions work in the same way when you try to break them down, e.g., into working memory, executive functions and spatial updating, in order to examine them closer as when the same functions are part of a complex real world task. One form of reality-based interaction that can be used to get higher ecological validity is *virtual reality* (VR), virtual worlds generated and simulated by a computer and connected to the user through an interface that aims to supplant the real world. One central vision that VR strives towards is a system which allows the user to be totally immersed in, and surrounded by, a generated reality that cannot be distinguished from the real world with perception nor interaction. This goal is often described using the terms *immersion* and *presence*. Immersion captures the physical and technical aspect of VR and the physical interface supplanting the real world is the deciding factor. Immersion increases with the proportion of sensory input that is generated by the virtual reality. If everything you see, hear, feel, etc, comes from the virtual reality you are totally immersed. Immersion is obviously increased in a multimodal system as more modalities and senses are covered. Presence on the other hand, is concerned with your mental experience of the system. In short, presence is high if you feel yourself to be present in a real environment, an environment that reacts and can be understood as a reality. An understanding of how multiple modalities are integrated and how a perception of reality is created and maintained is central to the study and improvement of presence. This issue plays a central role in the discussions of this paper.

Functional Brain Imaging and Virtual Reality

Functional Brain Imaging comes into play in two ways in this paper. First, brain imaging is a very important tool for exploring the function of the brain and much of the research reviewed has made use of brain imaging to explore how the brain works under many different conditions. In addition, I have a particular interest in brain imaging in combination with interaction and reality since my primary research revolves around the use of *virtual reality* (VR) together with *functional magnetic resonance imaging* (fMRI).

FMRI is based on the detection and analysis of the oxygenation of the blood in different areas of the brain. This can be detected and recorded with a *magnetic resonance* (MR) camera with a spatial resolution of a few millimeters (e.g. 4 mm) and a temporal resolution of one or two seconds (e.g. 1.5 s). When this data has been collected there are many available methods for analyzing it in order to find correlations between behavior or other known variables and neural activity deduced from the oxygenation of the blood. How to conduct this analysis and how to deduce the neural activity is under active research. See (Logothetis, 2008) for a current review of the capabilities and limitations of fMRI as a method.

The combination of VR and fMRI (VRfMRI) opens up new possibilities to create ecologically valid tests that can be used with fMRI and makes it possible to study how the brain functions in VR, thus enriching both fields. The MR-camera is a noisy and potentially uncomfortable place where your movement is restricted to assert the quality of data collection. It is not an ecologically valid environment for studying how the brain works in everyday life far from the confines of the camera. Remedying this is a challenge but using VR certainly moves us in the right direction. One way to improve further on this is to work to gain an increased understanding of presence and perception of reality, in order to employ the most efficient methods of fooling the brain into believing that it is in the context in which we want to study it. We also need to have a good grip on what the potential dangers are.

Multimodal system seems like an obvious development but without a proper understanding of multimodal integration the effects may be unpredictable.

Reality and the senses

Our perception of reality is fundamentally based on our senses. The senses are, per definition, our only way of getting stimuli from the surrounding environment. Using a term from above, we are *immersed* in a specific reality when the information we get from our senses corresponds to objects and events in this reality. This section focuses on how inputs entering into the brain via different senses affects each other and are integrated to form a perception of complex objects with associated smells, sounds, looks, etc. We review the existing literature and attempt to sort out related concepts, functions and phenomena. Two concepts that are widely used in this area are multisensory and cross-modal, both of which can refer to, e.g., functions or perception. These concepts are sometimes used in an overlapping manner but, primarily, multisensory simply means that multiple senses are involved, and the focus of cross-modal phenomena is the interactions between different senses/modalities.

The study of multimodal integration in the brain generally consists of comparing different measures of brain activity when receiving unimodal stimuli and when receiving multimodal stimuli. In this way areas of the brain that are heteromodal, that react to stimuli independently of the modality of the stimuli, have been located. This enables us to build models of feed forward networks where information from unimodal areas is integrated into multimodal representations as it converges while being fed upwards in the hierarchy. In this model, e.g., described in (Stein & Meredith, 1993), numerous areas of the cortex are considered to be heteromodal. The parietal cortex in particular is often referred to as an association area and among the subcortical areas the superior colliculus has been thoroughly researched. Feed forward integration of multimodal stimuli is certainly an important part of the picture, it is the best model we have for the hierarchical construction of invariant representations, but more recent research has made it increasingly clear that this is not the whole picture. There are interactions between modalities and areas of the brain previously considered unimodal that cannot described in this way. The emerging model of interactions among the parts of brain primarily handling specific modalities is not trivial, or complete, but the framework outlined in the section above can give some guidance when trying to conceptualize the relationships.

When reviewing the literature concerning multimodal integration one very quickly runs into the rather complicated concepts of attention and consciousness. The definitions of both are under some discussion and it can be hard to get a clear view of the border between them. One recent attempt to sort out the difference (Koch & Tsuchiya, 2007) argues that attention and consciousness are two distinct processes in the brain. However, while they specify that they are referring to top-down, selective, attention the examples they give of attention without consciousness are not obviously top-down. One example is priming, in this instance with words suppressed from conscious perception by a combination of forward and backward masking (Naccache, Blandin, & Dehaene, 2002). Another example is the fact that male and female nudes can attract attention even if they are rendered in such a way as not to reach conscious perception (Jiang, Costello, Fang, Huang, & He, 2006). I question the correctness of labeling these phenomena as top-down in this context. It is illuminating to compare these studies to research on multisensory interaction. Priming and attentional shifts are common in this literature. In one such study McDonald and colleagues (J J McDonald,

Teder-Sälejärvi, & S A Hillyard, 2000) presents evidence that sounds can trigger involuntary orientation of attention and that this in turn improves the quality of visual perception in a matching location, and in a later article (John J. McDonald, Teder-Salejarvi, Russo, & Steven A. Hillyard, 2003) they present strong support (based on activation timings measured with Event Related Potentials, ERP) for the hypothesis that this effect is due to feedback from multimodal areas in parietal cortex to unimodal areas in visual cortex. It is true that this feedback is top-down and could be labeled as top-down attention but growing evidence that this kind of feedback is common all over the cortex and can originate at many different levels puts the value of such a classification for this purpose into question. Is top-down attention always separate from consciousness or does it depend on from which level or area of the brain it is passed down? Defining attention and consciousness is not easy but it is too soon to establish that they are completely separate processes.

In a recent review of multisensory spatial interactions (Macaluso & Driver, 2005) the idea that this research can serve as "a window onto functional integration in the human brain" was held forward. These examples of functional integration based on a combination of the classical feed forward and predictive feedback provides encouraging support for frameworks attempting to describe the function of the brain and the neocortex in terms of memory and prediction, as described above. Further speculation along these lines might produce hypotheses with good descriptive and predictive properties, ripe for further research. As it is, the lack of clear definitions of attention and consciousness is a serious impediment to related research. It is maybe a little bit like the alchemists, working with the four elements, fire, water, air and earth. Researching concepts like these is certainly a good start and in many cases the issue at hand is specified with additional precision but the lack of a consensus makes this tedious and often lacking.

One way to be somewhat more specific concerning attention is to make a division between the reorienting or distracting function, and the focusing or attention fixating function. In a recent review of the reorienting system of the human brain (Corbetta, Patel, & Shulman, 2008) these functions are described as two separate functional-anatomical networks. A ventral frontoparietal network is responsible for interrupting and resetting ongoing brain activity when some external stimuli (e.g., threatening stimuli) requires reorientation and a dorsal frontoparietal network takes care of selection and suppress the ventral network when attention is focused. The function of these networks is described as supramodal, meaning that attention at this level can be directed by stimuli from any modality. They are also reacting primarily to stimuli based on behavioral relevance rather than sensory (unimodal) salience.

If attention mostly affects detection, quality of perception and reaction speed, the big remainder of multimodal integration phenomena is what you perceive, i.e., how you interpret what you see, hear, smell, etc. This is of course very much related to consciousness and this distinction between the "performance" of perception, as influenced by attention, and the interpretation and resulting conscious percept may be one of the best ways define a difference between attention and consciousness. I agree with Johan Eriksson that content (i.e., the "what") is a necessary part of a conscious state (Eriksson, 2007) but further discussion about the definitions of these concepts must be considered to be outside the scope of this paper. Instead, we focus on perception as it can be reported, i.e., what you can tell someone that you perceive. One classical example is the so called McGurk effect (McGurk & MacDonald, 1976) where the visual input from seeing the lip-movements of a person speaking affect what you hear so that the resulting perception of sound matches neither sound nor vision if these are manipulated to fool the subject. In a more recent study (Watkins, Shams, Tanaka, Haynes, & Rees, 2006) it is shown that the combination of a single flash with two auditory beeps is often perceived as two flashes. In this study they also used fMRI to investigate the neural

correlates of this illusion created by cross-modal interaction. The results show that activity in early visual cortex was increased by the concurrent auditory stimulation, demonstrating multisensory integration at the very first stages of sensory processing. This increase was present even when there was no illusion but when the subject perceived the illusion the activation was significantly greater and even matched activation if the subject was presented with two real flashes. This strongly supports the notion that what happens in the sensory cortices reflects the subjective experience and can very well be affected by other modalities. The study further associates this multimodal interaction with concurrent increases in activation in superior colliculus and the superior temporal gyrus, thus implicating these areas as involved in these multisensory interactions.

The possible neural correlates of multimodal integration can be divided into three rough categories depending on what kind of connections and correlations you are looking for. The first possibility is focused on the feed forward and feedback connections between different areas of the neocortex in correspondence to the hierarchical structure represented by these areas. Thinking along these lines we would make our starting-point the classical view of how unimodal input is fed upward into multimodal association areas such as parietal cortex and look for evidence of feedback moving down in the opposite direction.

A second perspective is to look for subcortical connections. There are a number of subcortical candidate areas to consider. Hawkins points to the thalamus as having a key role in the wiring of the neocortex and specifically for the detection and tracking of sequences and temporal correlations. The thalamus is connected to almost all parts of the neocortex and projects much of the incoming connections back to the same area. This means that the thalamus can function as a feedback-loop for most areas and pass the output of an area back as input, thereby enabling detection of correlations among what is and what just was. The thalamus also tends to send the returning signal back over a slightly larger area, giving each part of the cortex access to some information about what's going on in the neighborhood. This property is also a possible explanation for the findings you get when taking the third possible perspective. Another area of the brain that is similarly well connected is the claustrum, mentioned below.

The third perspective is to look at what happens in the borderland of unimodal areas. A study conducted on the rat brain (Wallace, R. Ramachandran, & Stein, 2004) has identified the existence of multisensory neurons (neurons firing in response to several stimuli trough multiple senses) throughout much of the sensory cortices with concentrations of above 50% in some areas along these borders. This again demonstrates that multimodal integration exists at the lowest level and there is no reason to think that these results are not relevant for the human brain as well. It is worth noting that the incidence of multisensory neurons in the central regions of each sensory area was low, signifying that early sensory perception still is primarily unimodal at this level.

Most of the research on multimodal interaction has been concerning spatial congruence and relations. Several of the studies referred above focus on stimuli that can be matched to the same spatial location. Other studies focus on temporal correlations, either with concurrent stimuli or with priming, although the latter is mostly combined with spatial correlations. One additional important aspect of multimodal integration is conceptual congruence. A recent study (Naghavi, Eriksson, Larsson, & Nyberg, 2007) examines this by presenting subjects with a picture and a sound simultaneously and vary whether or not these are conceptually related. The combination of a cat and a meowing sound would be such a conceptually related pair. Comparison of the activations for conditions with conceptually relevant combinations and conceptually irrelevant combinations showed an extra activation in the region of the claustrum and insula. These are areas that are of great interest from the second perspective of neural correlates for multimodal integration, as mentioned above. The

insula has been shown to contain mirror neurons, and is connected to several areas that are important for emotions and anticipation of feelings. However, the activation of the claustrum is the real gem here. The claustrum has been implicated as the seat of consciousness (Crick & Koch, 2005) and has reciprocal connections to almost all parts of cortex making it a prime candidate for integration of multimodal input.

Selecting realities

Our perception of reality relies on the predictions we make about the workings of our perceived surroundings. Recall the concepts of presence and immersion, introduced above. When we understand the environment that we are immersed in this leads to a subjective feeling of being present in a real world, in a reality. Recall the definition of reality given above, "predictability is the very definition of reality", and that being able to predict something means that you understand it, that you are familiar with it and that you can handle it expertly. In this section we delve deeper into these ideas, in particular in the context of Virtual Reality (VR), where the generated reality has to contend with an external (real) reality and there is some question as to which reality is perceived.

Presence, hypotheses and breaks

One description of presence that fits very nicely with the framework presented so far is based on the selection of a hypothesis about where you are and what you are experiencing (Slater, 2002). This hypothesis is the basis for your predictions and understanding of the world you feel yourself to be present in and it is selected to match and predict observations. Slater also introduces the concept of *breaks in presence* (BIPs), describing how you can be thrown out of presence by a specific event that fails to match your hypothesis and forces you to reevaluate and reject your perception of reality leading to either a switch of selected reality or confusion. The maintenance of presence, and thus a perception of reality, depends heavily on avoiding BIPs, or in other words, on avoiding critical prediction errors. Exactly which prediction errors are critical is an area for future research and one possible study is outlined below.

It should be noted that a key description of presence is which reality you are reacting to. If your perception is in some way affected by several realities (e.g., the real reality and a virtual reality) your behavior is the decisive measure of which reality you are present in. You cannot be present in several realities at the same time, but you might be able to construct a perception of reality that is a combination of input from what might be seen as several alternative realities to an outside observer. Also, since presence is a state of mind, if you are confused about what reality you are in you are not really present in *any* reality. Surprising events in the physical world do not improve your sense of being present in this reality; it makes you doubt whether you have a correct understanding of the world you are in and question the reality of these events and the world they occur in, thus leading to a decreased sense of presence.

Especially important aspects of this model of presence are that the selected hypothesis is continuously evaluated through a top-down process and that our brain is generally eager to accept a hypothesis and fond of maintaining it. This continuous top-down evaluation can be described in the framework presented above as the constant generation of predictions from higher levels in the cortical hierarchy. It is interesting to note that this function seems to be directly related to the idea of scan-sensing, i.e., the fact that the way in which you look at an

object depends on what you think that you are looking at. When you are looking at an obscured face your eyes moves in a particular pattern, one different from how they would move if you had not yet detected that it was a face. This is a clear example of how predictions affect our behavior directly and Hawkins and Friston both extend their reasoning (Hawkins, 2005; Friston, 2003) to claim that our motor control in general is essentially based on predictions and the drive to confirm them.

The eagerness to accept and fondness of maintaining a hypothesis can be related to the idea that the recognition of sequences and contexts are the fundamental function of the cortex. It is the job of our brain to recognize what it experiences and when faced with novel stimuli from a strange environment it stills strives to match this to previous experience. Considering the variation of objects and events in the real world that the brain manages to create invariant representations for, it does not seem strange that it accepts many flawed representations of similar objects (such as animated characters in a video game) as matching and continues to work on that hypothesis. This effect is evident in much of the VR related research as reports of presence are generally high even when environments may depart from reality in big ways. As illustrated by fact it should also be further noted that anomalies in a particular reality are not equal in their significance or tendency to trigger breaks in presence and it might not be apparent which events are most disturbing. This can, e.g., be related to the discussion about attention and reorienting above, in particular to the review (Corbetta et al., 2008) showing that the behavioral relevance of stimuli is the deciding factor concerning which stimuli triggers reorienting and shifts of attention. A concrete example is that anomalous movement around the eyes and mouth of a human would be far more disturbing that deviations in the general shape of the body (Slater, 2002).

Out-of-body Experiences

One powerful and very relevant example of the precariousness of perceived reality is given by recent research into the out-of-body experience (OBE) phenomenon. When we attempt to immerse a person in a virtual reality and want to attain the experience of presence in such a constructed location, getting the person to forget or disregard the location and context of her real body is an essential component. The definition of OBEs generally includes seeing your real body from the outside but the dissociation from the real body is a key commonality. In one recent study (Ehrsson, 2007) the subject is made to experience the illusion of being outside of her own body. This illusion depends on the integration of multimodal inputs in the form of correlated visual and tactile stimuli. The normal visual input is replaced by a combination of a head mounted display (HMD) and a camera, presenting the subject with a view of her own back by placing the camera two meters behind the subject. Tactile stimulation was correlated into this view by simultaneously rubbing the chest of the subject with the rod and moving an identical rod in a corresponding motion below the camera, as if rubbing an illusory body with eyes in the position of the camera. When the participants were interrogated (using questionnaires with control questions) after two minutes of such stimulation they reported a significant experience of sitting behind their physical bodies. These reports were further confirmed by measuring the emotional response (using skin conductance response) when their illusory body was “hurt” by hitting the space below the camera with a hammer. This showed a significant increase in emotional response when the illusion was reported as compared to when the multisensory stimuli was asynchronous (the chest of the real and illusory body was stimulated alternately instead of together) and the illusion was absent.

Research on this subject is ongoing and a newspaper article (Wänerholm, 2008) very recently describes how a study manages to fool subjects into accepting the body of a manikin

as their own. Much of the research in this area is based on or influenced by the well established "rubber hand illusion" (RHI) (Botvinick & Cohen, 1998) where subjects were made to experience a rubber hand as their own after a period of synchronized tactile stimuli to both hands while viewing only the rubber hand. This phenomenon has also been investigated to discern the neural correlates and results (Ehrsson, Spence, & Passingham, 2004) show that activity in the premotor cortex was related to the feeling of ownership of the rubber hand and suggests that a mechanism based on multisensory integration on this area is involved. In related work (V. S. Ramachandran, Rogers-Ramachandran, & Cobb, 1995) patients suffering from phantom limbs reported feeling touch applied to the intact arm as if it were applied to the missing arm when a mirror was used to make them see the intact arm in the place of the missing arm.

Normally the experience of an OBE is a sign of a clinical condition and a disturbance of normal brain functioning. Most research investigating the neural correlates of OBEs have this perspective and based on the study of such disturbances support has gathered for the conclusion that the this kind of OBEs depends primarily on ambiguous input from different sensory systems (Blanke, Landis, Spinelli, & Seeck, 2004) and/or failure to integrate multisensory stimuli (Blanke & Arzy, 2005), e.g., in the area of the temporo-parietal junction. This further supports the idea that multisensory integration is the key to understanding our perception of reality. In the cases described above, where OBEs were triggered in healthy subjects, it is reasonable to speculate that the creation of a more likely hypothesis, based on the integration of correlated multisensory input, managed to overrule the previous hypothesis about the presence of the self and the body. It also constitutes convincing support for the idea that the experience of presence can be fully determined by the perceptual process and the input from the senses.

Virtual presence

What are the functions and deciding factors for creating and maintaining presence in a virtual environment? Can we draw on the results above in the context of virtual reality? One similar study (Lenggenhager, Tadi, Metzinger, & Blanke, 2007) presents a corresponding result first in a setup very similar to the one with the camera filming the back described above and then replicates this with the real body replaced by a fake, virtual, body. As above, the illusion is triggered by synchronized tactile stimulation and compared to a condition with asynchronous tactile stimuli. In this case the effect is measured primarily by examination of the proprioceptive drift, i.e., the tendency to indicate the position of your hand (in the case of RHI) or body as shifted towards the position of the illusion of your hand or body. The results presented show that the real and fake body were both equally effective in the creation of an OBE illusion. However, an additional condition where the bodily representations were replaced with a virtual object with no bodily features (e.g., a box), failed to produce the same effect from synchronous stimuli. This distinction between the reaction to bodily stimuli and an object suggests that higher level representations of the body come into play.

In another study (Slater, Perez-Marcos, Ehrsson, & Sanchez-Vives, 2008) the focus is specifically on investigating the RHI phenomenon in VR by replacing the physical fake hand with an entirely virtual hand. In this case the visual representation of tactile stimuli was included in the virtual rendering so that a virtual ball touched the virtual hand at the same time as a real ball was touched to the real hand, out of sight. In addition to presenting results confirming that the illusion can be replicated in an entirely virtual setting with subjective reports (questionnaires) and proprioceptive drift measurements of muscle activity in the real arm (measured with electromyogram/EMG) while the virtual hand was rotated are included. This data shows a clear correlation between the level of illusion (from questionnaires) and

muscle activity when the virtual arm is rotated as compared to muscle activity when the virtual arm is still, thus demonstrating the profound effect of the adjusted perception of reality (hypothesis) on the brain.

Further examples of the effect and importance of presence in a virtual environment can be found in applications of VR to reduce and manage pain. In one influential study (Hoffman, Patterson, et al., 2004) the use of VR to distract a single burn victim from the pain is presented with results showing that both the experienced intensity and the unpleasantness of the pain was reduced and less time was spent thinking about the experienced pain. The virtual environment was specifically designed to distract from the burns by making it into a "SnowWorld" where you travel through an icy 3D canyon while throwing snowballs at passing objects. It is not clear what the relative significance of the conceptual incongruence between the burning pain sensation and presence within a freezing cold virtual reality is as the focus of the study is in the general distracting properties of VR. The subjective experience of pain requires attention and thus the potential of immersive VR to engage attention and divert it from the real world in general is a key value. This initial study is followed up by examining the experience of presence in this virtual environment (VE) during fMRI (Hoffman, T. Richards, Coda, A. Richards, & Sharar, 2003), i.e., while confined by the constraints and limitations mentioned in the introduction of this paper and with further studies (Hoffman et al., 2006; Hoffman, T. L. Richards, et al., 2004) using this VRfMRI setup to investigate the neural correlates of VR pain management. The study focusing on presence during fMRI mainly presents the VRfMRI system used and establishes that presence was rated higher in the VE in an unobstructed view of the VE than if a white cross was introduced to obstruct the view. Neural correlates presented in the further studies clearly show the effects of immersive VR on brain activity in brain areas commonly related to pain. Pain activation in these areas was first confirmed by fMRI with laboratory thermal pain and five areas of interest were selected. The following comparison of activation in these areas with and without VR distraction (with the order of these conditions randomized in the same session) showed significant reduction of brain activity in all areas. In other studies the use of VR for pain management has been used to distract children during intravenous placement (needle insertion) (Gold, Kim, Kant, Joseph, & Rizzo, 2006) and cold pressor pain (Dahlquist et al., 2008). The first of these studies show a clear improvement of subjective experience with VR (a fourfold decrease of affective pain) while the second study deals with the additional gain of using a head mounted display and present results suggesting that this is effective for some children, but not all.

Future work

We intend to investigate further how the occurrence of conflicting stimuli from different modalities can affect presence in a virtual world, perception of such a reality and consequently the ecological validity of the brain activation in such a setting. If we want to investigate how the brain works when performing tasks in a complex setting such as a downtown area of a city we need to establish to which degree the brain works as if it was in such a setting. As described above, we need to know that you are reacting to the virtual environment and not to any outside stimuli such as distractions from the MR-camera and if this is uncertain we need to know how conflicting stimuli might affect the resulting brain activation. In order to tackle this issue we want to examine how isolated stimuli events that are not in agreement with the rest of your experience affect you. The first step will be to design and conduct a study measuring reaction speeds and behavioral data when the stimuli

in question does or does not disrupt your general hypothesis about the workings of your current world. We believe that a comparison based on sounds with a varying degree of conceptual relevance for your current reality is a good starting point. There are three basic categories of sounds that would be relevant to test, one category matching the virtual reality, one matching the outer (real) reality and one category matching neither. Examples of such sounds if you are immersed in a VE representing an expansive forest while in an indoor laboratory in reality would be the sound of singing birds (matching the VE), the sound of chairs being moved (matching the real world) and the sound heavy traffic (matching neither). Virtual sounds should increase presence in the VE and support efficient reaction to virtual stimuli and interaction with the VE. Real sounds should result in a BIP and a subsequent decrease in efficiency when interacting with the VE. Sounds matching neither reality should result in confusion and probably an even greater reduction of interaction efficiency.

One of the challenges in designing this study will be the timing and maintenance of these desired effects. If too short sounds are used it is probable that presence will switch back to the VE or that it will be unpredictable. It is also important to consider the effect of repeated sounds. If the sound of traffic is used to induce confusion once it is probable that the effect will not be the same if the sound is repeated later. It would be interesting to construct a design were the effect of such repeated sounds could be compared to the effect of novel sounds in the same category and also to compare the effect of repeated sounds in different categories.

If this behavioral study yields interesting result we will move on to a study using VRfMRI to investigate the neural correlates for these variations in presence and resulting efficiency. Expectations concerning the resulting brain activity can be based on the fact that non-matching sounds should result in reduced multimodal integration and consequently in increased confusion and reduced understanding. The assumption that understanding means that the brain is working more efficiently and that much of the activity in the brain constitutes new predictions and prediction errors, both of which are needed to a lesser extent when the sound matches the current presence hypothesis, further predict that non-matching sounds should in general result in greater brain activation. The brain activity should in particular be greater than for matching sounds at the level where predictions fail and a break in presence is initiated. This level should be part of the auditory processing hierarchy, quite possibly at the lowest level.

Another possible hypothesis based more on the tendency to maintain the current presence hypothesis is that non-matching stimuli will be suppressed initially. This might lead to an initial decrease in activation in areas related to multimodal integration of sound but sustained non-matching stimuli should force recognition of the sound and a consequent reevaluation of understanding and a break in presence, as discussed above. Investigating the limit for when a presence hypothesis switches from suppressing contradictory stimuli to breaking down is interesting and probably can be addressed in several future studies.

Summary and conclusions

This paper attempts to tie together research spanning over a rather large set of fields, but all related to the perception of reality, multimodal integration and interaction in this context. The prevalence of concepts such as consciousness and attention in this research is at the same time a confounding factor and an exciting possibility. I have tried to provide some structure by referring back to a framework of prediction and representation described in the first section of this paper. I believe that this framework fits well with a lot of this research and

that such an overarching model is a great aid for understanding the related concepts and placing them into a context. In the future I would like to contribute to developing and test similar models.

The key motivation for the review of the issues presented in this paper is the quest for efficient interaction and ecological validity. Interaction designed to take advantage of the fact that the human brain has evolved to deal with a multimodal reality holds great potential for increased efficiency. The human brain is familiar with the conditions of reality, we understand reality and we are experts in interacting with reality. Virtual reality and reality-based interaction are two promising ways forward and the presented review of presence and how our perception of reality can be constructed and manipulated in a virtual reality constitutes an important foundation for future work. We can see that it is possible to make humans believe that they are present in a constructed reality and even let go of the connection to their own body and we see how the integration of stimuli from several modalities is essential for achieving such effects. Systems taking advantage of these results have the potential to make use of the full range of natural human interaction capabilities and the increased ecological validity is of key value to the use of VR and VRfMRI, making it possible to study the functions of the brain in complex settings and enabling the development of better diagnostic tools for detection of cognitive degeneration.

References

Blanke, O., & Arzy, S. (2005). The out-of-body experience: disturbed self-processing at the temporo-parietal junction. The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 11(1), 16-24. doi: 11/1/16.

Blanke, O., Landis, T., Spinelli, L., & Seeck, M. (2004). Out-of-body experience and autoscopy of neurological origin. Brain: A Journal of Neurology, 127(Pt 2), 243-58. doi: 14662516.

Botvinick, M., & Cohen, J. (1998). Rubber hands `feel' touch that eyes see. Nature, 391(6669), 756. doi: 10.1038/35784.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. Neuron, Vol 58, 306-324.

Crick, F. C., & Koch, C. (2005). What is the function of the claustrum? Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1458), 1271–1279. doi: 10.1098/rstb.2005.1661.

Dahlquist, L. M., Weiss, K. E., Dillinger Clendaniel, L., Law, E. F., Ackerman, C. S., & McKenna, K. D. (2008). Effects of Videogame Distraction using a Virtual Reality Type Head-Mounted Display Helmet on Cold Pressor Pain in Children. J. Pediatr. Psychol., jsn023. doi: 10.1093/jpepsy/jsn023.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in□attention. Proceedings of the National Academy of Sciences of the United States of America, 93(24), 13494-13499. doi: VL - 93.

Ehrsson, H. H. (2007). The Experimental Induction of Out-of-Body Experiences. Science, 317(5841), 1048. doi: 10.1126/science.1142175.

Ehrsson, H. H., Spence, C., & Passingham, R. E. (2004). That's My Hand! Activity in Premotor Cortex Reflects Feeling of Ownership of a Limb. Science, 305(5685), 875-877. doi: 10.1126/science.1097011.

Eriksson, J. (2007). The conscious brain: Empirical investigations of the neural correlates of perceptual awareness. Text, . Retrieved November 19, 2008, from http://www.diva-portal.org/umu/abstract.xsql?dbid=1430.

Friston, K. (2003). Learning and inference in the brain. Neural Networks, 16(9), 1325-1352. doi: 10.1016/j.neunet.2003.06.005.

George, D., & Hawkins, J. (2005). A hierarchical Bayesian model of invariant pattern recognition in the visual cortex (Vol. 3, pp. 1812-1817 vol. 3). doi: 10.1109/IJCNN.2005.1556155.

Gold, J. I., Kim, S. H., Kant, A. J., Joseph, M. H., & Rizzo, A. S. (2006). Effectiveness of virtual reality for pediatric pain distraction during i.v. placement. Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, 9(2), 207-12. doi: 10.1089/cpb.2006.9.207.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends in Cognitive Sciences, 10(1), 14-23. doi: 10.1016/j.tics.2005.11.006.

Hawkins, J. (2005). On Intelligence. Owl Books.

Hoffman, H. G., Richards, T., Coda, B., Richards, A., & Sharar, S. R. (2003). The illusion of presence in immersive virtual reality during an fMRI brain scan. Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, 6(2), 127-31. doi: 12804024.

Hoffman, H. G., Richards, T. L., Bills, A. R., Van Oostrom, T., Magula, J., Seibel, E. J., et al. (2006). Using FMRI to study the neural correlates of virtual reality analgesia. CNS Spectrums, 11(1), 45-51. doi: 16400255.

Hoffman, H. G., Richards, T. L., et al. (2004). Modulation of thermal pain-related brain activity with virtual reality: evidence from fMRI. Neuroreport, 15(8), 1245-8. doi: 15167542.

Hoffman, H. G., Patterson, D. R., et al. (2004). Water-friendly virtual reality pain control during wound care. Journal of Clinical Psychology, 60(2), 189-195. doi: 10.1002/jclp.10244.

Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., et al. (2008). Reality-based interaction: a framework for post-WIMP interfaces.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. Proceedings of the National Academy of Sciences of the United States of America, 103(45), 17048–17052. doi: 10.1073/pnas.0605678103.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. Cognitive Processing, 8(3), 159-66. doi: 10.1007/s10339-007-0170-2.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. Trends in Cognitive Sciences, 11(1), 16-22. doi: 10.1016/j.tics.2006.10.012.

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video Ergo Sum: Manipulating Bodily Self-Consciousness. Science, 317(5841), 1096-1099. doi: 10.1126/science.1143439.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. Nature, 453(7197), 869-878. doi: 10.1038/nature06976.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. Trends in Neurosciences, 28(5), 264-271. doi: 10.1016/j.tins.2005.03.008.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. Nature, 407(6806), 906-8. doi: 11057669.

McDonald, J. J., Teder-Salejarvi, W. A., Russo, F. D., & Hillyard, S. A. (2003). Neural Substrates of Perceptual Enhancement by Cross-Modal Spatial Attention. Journal of Cognitive Neuroscience, 15(1), 10-19. doi: 10.1162/089892903321107783.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264(5588), 746-8. doi: 1012311.

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. Psychological Science: A Journal of the American Psychological Society / APS, 13(5), 416-24. doi: 12219807.

Naghavi, H. R., Eriksson, J., Larsson, A., & Nyberg, L. (2007). The claustrum/insula region integrates conceptually related sounds and pictures. Neuroscience Letters, 422(1), 77-80. doi: S0304-3940(07)00668-4.

Ramachandran, V. S., Rogers-Ramachandran, D., & Cobb, S. (1995). Touching the phantom limb. Nature, 377(6549), 489-490. doi: 10.1038/377489a0.

Sjölander, S. (1999). How Animals Handle Reality—The Adaptive Aspect of Representation. Understanding Representation in the Cognitive Sciences: Does Representation Need Reality?

Slater, M. (2002). Presence and the sixth sense. Presence: Teleoperators & Virtual Environments, 11(4), 435-439. doi: 10.1162/105474602760204327.

Slater, M., Perez-Marcos, D., Ehrsson, H. H., & Sanchez-Vives, M. V. (2008). Towards a Digital Body: The Virtual Arm Illusion. Frontiers in Human Neuroscience, 2, 6. doi: 10.3389/neuro.09.006.2008.

Stein, B. E., & Meredith, M. A. (1993). The Merging of the Senses (p. 211).

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci, 11(9), 1004-1006. doi: 10.1038/nn.2163.

Wallace, M. T., Ramachandran, R., & Stein, B. E. (2004). A revised view of sensory cortical parcellation. Proceedings of the National Academy of Sciences of the United States of America, 101(7), 2167–2172. doi: 10.1073/pnas.0305697101.

Watkins, S., Shams, L., Tanaka, S., Haynes, J., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. NeuroImage, 31(3), 1247-56. doi: S1053-8119(06)00045-0.

Wänerholm, M. (2008, December 2). Skaka hand med dig själv. Dagens Nyheter.