What is Movement Interaction in Virtual Reality for?

Marco Gillies
Dept. of Computing, Goldsmiths, University of London
New Cross, London, UK
m.gillies@gold.ac.uk

ABSTRACT

This paper raises the question of why movement base interaction is important in Virtual Reality (VR). This is an important question as new VR hardware is increasingly being released together with movement interfaces. Slater's view is that VR reproduces the sensorimotor contingencies present in our interactions with the real world. This provides a powerful justification, but when the contingencies are not perfectly reproduced, they can result in interfaces that lack important features of established interaction design: discoverability memorability, and feedback. However, Embodied Cognition suggests that these imperfect reproductions can still have value if they allow us to reproduce our cognitive and emotional engagement with the world and our movements.

Author Keywords

Embodied interaction; virtual reality; full body interfaces

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

General Terms

Human Factors; Design;

INTRODUCTION

The recent resurgence of Virtual Reality (VR) display technology has been accompanied (slightly later) by a number of movement interface devices. The HTC VIVE will be released with two hand trackers and after the release of the Oculus Rift, OculusVR will be releasing hand trackers to go with it, and both systems include head trackers. These releases underline the idea that movement interaction is a necessary, or at least very important, part of VR. This is a long established idea and is well supported by a number of researchers such as Slater [11] and Jacob [6] and numerous studies (some of why are described below). However, if we are to design movement interaction for VR we need to understand why movement is important in VR. Without this understanding it would be easy to create interfaces that are not effective, and which may even be detrimental relative to a traditional game controller.

After describing examples of movement interaction in VR, this paper will outline Slater's theory of Place Illusion, to which movement interaction is key. After that we describe Norman's criticism of movement interfaces from the point of view of User Centred Design. Finally, we discuss Embodied Cognition which can bring new insights into the importance of movement interaction and points to a possible explanation of its value.

EXAMPLES

Before we look at the effects and purposes of movement interfaces in Virtual Reality it is worth describing some examples of how movement is used in current VR interaction.

Head tracking

Head tracking is one of the most basic movement interactions, and one that present in almost all VR systems that can plausibly claim to be immersive (in the sense used by Slater [11]: a system whose physical characteristics make it capable of producing the illusion of being in another place) . When a user rotates their head a tracker attached to the head mounted display or stereoscopic glasses recognizes this movement and rotates the view of the virtual world producing the effect that the user seems to be able to look around the world by moving their head. Most systems also support positional head movements, though some cheaper systems such as google cardboard and Samsung GearVR do not. Slater [11] proposes that this tight relationship between head movements and view is one of the key factors that makes a VR system immersive.

Walking

Positional head tracking can allow for a limited range of movement with a virtual environment, but unless the environment is very small, we need some form of large scale navigation to fully explore it. Most systems use a navigation system based on traditional computer games, for example a joystick. However, some have also used head tracking to enable users to walk around an environment just as they would in the real world. Usoh, et al. [12] found that walking created a stronger sense of presence than using a joystick. However, walking requires a physical space and tracking volume at least as large as the virtual space, making it impractical in most cases. For this reason they also propose an intermediary system in which a neural network is used to recognise when a user is walking "in place" (taking steps without moving forward or backward). They found that this method also created

MOCO'16, July 05 - 06, 2016, Thessaloniki, GA, Greece.

ACM ISBN 978-1-4503-4307-7/16/07...$15.00
DOI: **http://dx.doi.org/10.1145/2948910.2948951**

a greater sense of presence than joystick navigation but not as great as real walking.

Object manipulation

Another key interaction is to select and manipulate objects in the world. Bowman and Hodges [1] have investigated object manipulation at a distance (further than would normally be possible by extending your arm). They compared a movement based augmented arm extension method where the arm stretches more than it usually would, with a ray casting method, where a line is cast from the users hand and an object is selected if it intersects that line. Raycasting was found to be preferable for selecting at a distance (a rather unnatural task), but once the object is selected, directly mapping hand movements and rotations onto object manipulations was preferred by all participants.

Body language

The above three example are all relatively low level behaviors involving direct physical interaction with the world. However, virtual reality can also support more complex interactions. For example, there has been considerable work on simulating interactions with animated virtual humans. While speech is an important part of this, movement is also key in order to simulate realistic body language. For example, Gillies and Slater [4] recognize a number of behaviors by the user such as speaking, posture shifts and moving in and out of the the virtual human's person space. These result in response from the virtual human such as nodding or moving to maintain a comfortable conversational distance. Huang et al. [5] propose a more sophisticated machine learning system that use features based on speech prosody and body movement to find appropriate times for a virtual character to provide backchannel feedback while listening (e.g. nodding or saying "uhuh"). These examples are much more indirect than the ones discussed above. While head tracking requires an immediate one-to-one response, body language response can happen at varying and can take different forms and often do not happen at all without the interaction seeming strange.

These are a number of examples where movement interaction appears to increase the sense of presence in VR. We should now ask why is this?

SLATER'S SENSORIMOTOR ACCOUNT

Why is rotating the view by turning your head better than using a mouse? Why does walking create more of a sense of presence than using a joystick. Slater has spent decades researching these questions and has proposed a comprehensive theory [11]. He defines that a virtual reality is immersive if it reproduces the same sensorimotor contingencies as the real world. Sensorimotor contingencies are the key element of O'Regan and Noë's theory of perception [10], they propose that the perceptual experience of the world is not purely due to sensory input but the relationship between motor actions and the resulting sensations. Thus turning your head and having the view of the world update is a key sensorimotor contingency and Slater would propose that reproducing this contingency in VR will make a system more immersive. Slater's theory states that if some one experiences a VR system that supports many real world sensorimotor contingencies they are more likely to experience a form of presence called "Place Illusion" in which they feel they are present in a different place from the one they physically inhabit.

Place illusion accounts for many of the examples we have discussed above. Turning our head and our view of the world changing is a very important sensorimotor contingency in the real world and reproducing it in VR is likely to create place illusion. Similarly the sensorimotor relationship between the movements associated with walking and the larger scale transformation of our view point is a key part of the world, and, in fact, when they are decoupled they often result in nausea. Similarly, O'Regan and Noë's[10] account would propose that the relationship between our hand movements and the perceptual changes in viewpoint of an object we are holding define what it feels like to manipulate an object. However, higher level interactions such as body language interaction with other people are more complex. They do not take the form of direct, immediate mappings of movement and sensory information. Slater [11] extends his theory to handle these forms of interaction by proposing a second form of presence called "Plausibility Illusion". This covers the case where interactions are less direct but are none the less correlated in natural ways in the real world. For example, we do not expect a person we are talking to to immediately respond to the slightest movement we make, as we would expect of an object we are holding in our hand. But we do expect their behavior to more roughly correlate to ours by showing acknowledgement of what we are saying and using facial expressions that are appropriate to the emotional tone of what we are saying.

Slater's theory is closely related to Jacob's [6] concept of "Reality Based Interactions". This is a way of characterizing a wide range of new interaction techniques including virtual reality and movement based interaction. These techniques are successful, according to Jacob, because they allow us to use the skills we know from the real world, including interacting with physical object, our body, our environment and other people. Since these skills are in large part related to the sensorimotor contingencies we experience in the real world and the correlations between our behaviour and that of other people and things in our world, Jacob's theory is compatible with Slater's.

To summarize, Slater proposes (and Jacob supports him) that presence is created, at least in part, when a virtual environment responds to our behavior in the same way as the real world. On a technical level this requires that a system can *sense* our behavior and *respond* to it in an appropriate way.

CRITIQUES OF MOVEMENT INTERACTION

Both Slater and Jacob support the idea that movement interaction can be natural if it reproduces the interactions and sensorimotor contingencies that we expect form the real world. However, there have also been several important criticisms of movement based interaction. Norman [9] criticizes many current gestural interfaces because they lack a number of features of good interfaces. For example they are not discoverable in the sense that it is hard to know what gesture to do and we

often cannot look it up in the way we would look at a menu to see what options are on it. This also makes them less memorable: there is no support in remembering the correct gesture. Finally, many gestures do not provide good feedback on whether they are being performed correctly.

Do these criticisms apply to the kind of movement based interfaces used in VR? It can certainly be argued (and I have argued in the past [3]) that movement interfaces can overcome the problems cited by Norman if they implement the sensorimotor contingencies we expect from the real world. We know what to do and can remember them because we have learned these forms of interaction from early childhood. They provide feedback because our sensory experience is directly tied to our motor actions. This clearly applies to the examples of head tracking and direct walking. We are simply doing things we have learned since childhood and we have immediate feedback that we are doing them correctly because our view of the world is constantly updated. However, this argument is not so compelling in the case of walking in place or body language interaction. Walking in place is not an action we do every day, it is a gestural interface that is similar to a day to day action. It therefore has to be learned and is not necessarily memorable in the same way. Also, direct walking is implemented very simply with a position tracker that can be relied on to be mostly accurate. Walking in place, on the other hand is implemented by recognizing the walking gesture via a complex neural networks. This is unlikely to be as accurate. It might fail to recognize actions that are intended to be walking, but more importantly is also likely to recognize other movements that cause similar head movements but are not walking (another problem identified by Norman). Body language is far more complex than walking in place, not only are the movements involved quite subtle, but many different movements can have the same meaning. The meaning of body language can also rely on the combination of many cues some of which might not be sensed, for example, the emotional tone might combine body posture, the content of speech, tone of voice and facial expression (which might be impossible to sense if a users is wearing a head mounted display). All of this means that sensing body language is likely to be very inaccurate. What is worse is that, since the responses are indirect, it might not even be clear whether a behavior has been recognised correctly or not.

The end effect is often that, though the computer is supposed to sense what a person is doing, in fact, the person is constrained to a limited set of actions that a computer can perform (this is a well known issue with natural language interfaces). That is supposed to be a natural sensorimotor mapping becomes an exercise in guessing the correct action without any of the visual exploration or feedback that a graphical interface provides.

So we cannot simply treat virtual reality interaction as a process of sensing a person's behaviour and providing a sensory response that matches the real world. Any but the simplest interactions and mapping will require a clear set of behaviours from users and these behaviours must be discoverable, memorable and provide sufficient feedback to be learnable. VR interaction is still a human computer interaction and it requires a process of user centred design (in the sense used by Norman[8]) that follow the same rules as current interaction design.

EMBODIED COGNITION AND POWER POSES

While Slater's arguments suggest that movement interaction can support presence if the sensing is near perfect, Norman's critique shows that in cases where the recreation of sensorimotor contingencies is unreliable or imperfect we should return to a more traditional view of interaction design that values the learnability and memorability of interfaces. A movement interface therefore remains an interface and not a reproduction of the real world. In these cases is there still value in movement interaction or should we revert to using traditional interface devices?

To state the question more exactly: if it is not possible to exactly reproduce sensorimotor contigencies, but instead we must have a constrained interface, is there value in using a movement interface rather than a more traditional one?

The theory of embodied cognition can help us understand this issue. According to Kirsh [7] much of our cognition occurs in our perceptuo-motor system: we think, at least in part, by sensing and manipulating the world around us. When assembling flat pack furniture, we do not treat the problem of how two pieces fit together in an abstract manner. We look at the particular shapes of the two pieces and then rotate them in our hands to better understanding the shapes from different angles and also how they could fit together in different ways. This perceptuo-motor thinking is not restricted to physical problems, our mathematical reasoning is almost always supported by physical actions, from counting on our fingers to using pen and paper for advanced calculations. This way in which our cognition is embodied means that the way we physically perform actions matters a lot cognitively, it affects how we think and how we experience. Kirsh [7] stresses that using different tools has a fundamental effect both on our cognitions (a pencil and paper changes our ability to do mathematics) and our perceptions (when cooking, using a wooden spoon allows us to feel food stuck to the bottom of a pan in a way that we simply cannot access using our normal senses). This has the implication that how we interact with virtual reality will have a fundamental cognitive and experiential effect on us.

This also includes emotion. For example, Carney et al. [2] performed an experiment in which participants were placed in poses that are normally considered either powerful (for example, feet up on a table and leaning back in a chair) or weak (sitting straight with head bowed and hands in their lap). The poses were not described in terms of power, the participants were simply given detailed instructions of how positions their body. The participants in the more powerful poses had increased sense of power both in terms of hormone levels and in terms of risk taking behaviour. In an earlier study, Wells and Petty [13] told participants that they were supposed to test the headphones and how they performed while listeners moved. Participants were asked to either move their heads vertically or horizontally. The vertical head movements were designed to be very similar to nodding, a signal for agreement, and the

horizontal movements were similar to head shaking, a signal for disagreement, though participants were not made aware of this similarity. They found that simply making nodding movements made participants more likely to agree with the a person speaking through the headphones while head shaking had the opposite effect. Both of these studies show that movements can have a real effect on people's emotions and social responses. This suggests an interesting design strategy for movement interfaces: create movements that will emotionally involve participants more in the VR experience. For example, using a head nod to agree with a character may create a stronger social and emotional effect than pressing a button. The fact that in both studies the participants were unaware of the social signals they were doing, but were simply instructed to perform certain movements, suggests that the interfaces do not need to recognize a full range of social signals, it is enough to have a clearly defined gesture, which nonetheless mimics a common social signal.

Kirsh presents an intriguing study [7] which might provide support for this idea. He study professional dancers who were asked to practice a piece in one of three ways: practicing the piece in full, imagining the piece without moving and a third method called "marking" in which the dancer does a reduced version of the movement (a common practice technique in dance). Full practice was better than imagination but interestingly marking was best of all. This opens up the possibility that, at least in some cases, a reduced movement interface, such as walking in place, might actually be better in some way than a full movement interface. It is far too early to draw this conclusion for definite, and the mechanisms are not yet fully understood, but this does suggest a very interesting area for future research.

CONCLUSION

No single theory seems to total explain the importance of movement interaction in VR. While Slater's theory is very important, and characterizes some very important elements of VR interaction, simply attempting to reproduce real world sensorimotor contingencies can break down when they are too complex to model and recognise easily in code (or an not feasible for other reasons, like direct walking). At this point it is important to remember that VR interfaces are still interfaces, and we also need to take account of more traditional interaction design values if we are to make VR usable. If an interface cannot perfectly reproduce real world sensori-motor contingecies, is there still value in it being similar to the real world. Embodied Cognition seems to provide an answer: a simplified movement interface can still be important if it supports the embodied aspects of our cognition and emotions effectively. The interface should trigger thoughts and emotions in a similar way to real world movement, even if the movement is not identical.

REFERENCES

1. Bowman, D. A., and Hodges, L. F. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics - SI3D '97*, ACM Press (4 1997), 35–ff.
2. Carney, D. R., Cuddy, A. J. C., and Yap, A. J. Power posing: brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological science : a journal of the American Psychological Society / APS 21*, 10 (2010), 1363–1368.
3. Gillies, M., and Kleinsmith, A. Non-representational interaction design. In *Contemporary Sensorimotor Theory*, J. M. Bishop and A. Martin, Eds. Springer-Verlag, 2014, 201–208.
4. Gillies, M., and Slater, M. Non-verbal communication for correlational characters. In *International Conference on Presence*, M. Slater, Ed. (9 2005).
5. Huang, L., Morency, L.-P., and Gratch, J. Virtual rapport 2.0. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents*, IVA'11, Springer-Verlag (2011), 68–79.
6. Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J. Reality-based interaction: a framework for post-wimp interfaces. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM (2008), 201–210.
7. Kirsh, D. Embodied cognition and the magical future of interaction design. *ACM Trans. Comput.-Hum. Interact. 20*, 1 (4 2013), 3:1–3:30.
8. Norman, D. A. *The Design of Everyday Things*. Basic Books, Inc., 2002.
9. Norman, D. A. Natural user interfaces are not natural. *interactions 17*, 3 (5 2010), 6–10.
10. O'Regan, J. K., and Noë, A. A sensorimotor account of vision and visual consciousness. *The Behavioral and brain sciences 24*, 5 (2001), 939–973.
11. Slater, M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos Trans R Soc Lond B Biol Sci 364*, 1535 (12 2009), 3549–3557.
12. Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., and Brooks, F. P. Walking ¿ walking-in-place ¿ flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99*, ACM Press (7 1999), 359–364.
13. Wells, G. L., and Petty, R. E. The effects of over head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology 1*, 3 (6 2010), 219–230.