

The Role of Effector Physicality and Risk Perception in Virtual Environments

Shulan Lu (Shulan.Lu@tamuc.edu)

Derek Harter (Derek.Harter@tamuc.edu)

Gang Wu (gangwu6@gmail.com)

Pratyush Kotturu (pratyush.kotturu@gmail.com)

Texas A&M University – Commerce

Commerce, TX 75429 USA

Abstract

Research has consistently demonstrated that people treat digital technology-based environments such as VR as if they were real. This is consistent with neural reuse and predictive processing theories. Neural circuits that have developed to perform real world actions are reused when performing tasks in computer mediated environments. The current research investigates some of the factors that could support users in leveraging their existing real world representations. A reasonable hypothesis is that users are more likely to emulate existing real world processing if technological artifacts are congruent with their experiential basis. This work investigates the perceived cues of task risks, movement realism and effector realism in performing actions. Effector design is manipulated (gesturing, wand, vs. knife), and participants cut a vegetable in a simulated environment. Participants evoked real world sensory motor contingency when technological artifacts are congruent with their experiential basis.

Keywords: embodied cognition; risk perception; computer mediated learning; danger avoidance; effector; controller

Introduction

Increasingly technological systems have begun to develop new interactive styles that leverage the richness of humans' real world interactions. For example, systems using low cost full body motion tracking, such as Kinect, have been made available. There is also a breakthrough in eye gaze based interactive system such as LC technologies' eye gaze edge tracking. Because of this departure from WIMP interfaces, a significant question arises as to whether and how gestural interactions, or in some cases intention driven touchless interactions, can evoke representations that are similar enough to perceiving and enacting actions in the real world in order to train up responses and habits that would be able to later get deployed in real world practices. If not, what differences might there be?

A myriad of theoretical approaches have been proposed to guide the design of systems that support users embodying themselves in the environment and participating in the interactions meaningfully. One of the central themes of this embodied interactive movement is to encourage the alignment between the representations being constructed for the digital world and the relevant experiential basis, making digital artifacts part of the background in the formation of representations instead of being in the foreground (Dourish,

2001; Hornecker, 2011; Ishii, 2008; Jacob, et al., 2008; Lu, Harter, Kosito & Kotturu, 2014; Slater, 2009). By judiciously re-representing the key elements in physical reality, as well as tapping into visual-perceptual cues, such digital-physical systems create a new interface interaction paradigm that leverages existing embodied proprioceptive abilities and motor skills we all develop and employ in the real world. This movement is consistent with insights from embodied and grounded cognitive science (Kirsh & David, 2013).

Recent views of embodied cognition are exploring the high level neural mechanisms that may be critical to our embodied cognitive abilities. For example, views of cognition as being hierarchical predictive machinery, where higher level layers predict activity of lower layers, and the lower layers send feedback in the form of error signals of the predictions have been proposed (Clark, 2013; Anderson, Richardson, & Chemero, 2012; Barrett & Simmons, 2015). These predicative theories suggest that more abstract concepts and higher level abilities, such as keeping track of goal states, are built up through the testing and refining of predictive mechanisms. The predictions and error signals are fundamentally bidirectional, higher levels generate predictions of the neural patterns of activity of lower layers, and mismatches generate error signals that are propagated back up the hierarchy which can be used to refine the predictive machinery.

This brain as active predictive machine view suggests that the sensory repertoire gathered from past experiences and the current sensory/perceptual inputs constrain the computation of probabilities that underlie neural representations. Such predictive views of embodied cognition are especially relevant to understanding human performance in computer mediated environments. In a computer mediated environment, we use predictive machinery that is evolved and developed to work with other (usually real world) experiences in order to interact with the digital environment (Lu & Harter, 2016).

The reuse and redeployment of neural circuits is expected (according to neural reuse theories) in order for perceptual predictions to be as efficient and accurate as possible in computer mediated environments (Anderson M. L., 2010). There are two mechanisms by which neural circuits are commonly reused, especially in the context of learning to

use a computer mediated environment for some task. In one type of reuse, new types of higher-level prediction abstractions will be created to learn predictions of low-level circuitry that is essentially being used for the purpose it was originally developed for. For example, in order to interpret visual objects being depicted in a virtual reality, they are of course designed to be visually similar to their real world counterparts. Another type of reuse is where low-level circuitry is put to a novel function by existing higher-level abstractions to cope with the differences in an unfamiliar computer mediated experience. For example, we may be experiencing a common task in a simulated environment, such as moving objects around to complete some goal, but our low-level motor actions needed in order to interact with the virtual world use some sort of input effector like a joystick rather than our own hands to perform the task.

In cognition, this predictive machinery results in a tight coupling of what is available in the environment (such as the fidelity of the environment) and the sensory motor contingency that gets triggered in a user. Central to the argument in the current work is the bi-directionality of this coupling. For example, user's movements can modify which aspects of the environment are attended to and reflexively tweak the run-time representations that are used for selecting the next action. However, previous research has been inconclusive to this prediction and the existing research paradigms are not conducive to understanding these bi-directional interactions as they unfold. In existing studies, researchers examined explicit game performance measures and player subjective reports including perceived mental workload, and did not look into real time processing measures (Freeman, et al., 2012; Reinhardt & Hurtienne, 2018). In yet another study, video clips of transitive actions were examined and participants reported the habitual actions were perceived to be easier and more natural to understand (Grandhi, Joue, & Mittelberg, 2011).

In recent work on immersive virtual reality, researchers have demonstrated the current state of the art in terms of providing tracking of handheld effectors in a typical head-mounted display (HMD) virtual reality system (Pandey, Pidlypenskyi, Yang, & Kaeser-Chen, 2018). Tracking the position of the handheld effectors is of course important in theory in order to provide an immersive experience not only of seeing the environment, but of having your body (hands and arms) embodied and perceptible within the environment. This is relevant to our current study, as it shows what may be possible in virtual reality to enable embodying hand movements and interactions. For example, the reported image-based markerless 6 degrees of freedom tracking of handheld effectors demonstrated much more reliable tracking than current virtual reality systems can achieve without additional sensors embedded in the handheld effectors. In fact, though not discussed in this article, it would seem that this method could be applied equally well to tracking the user's hands, even without holding a effector. In the research report, the authors

showed that using machine learning and dual visual images, such a system can be trained to track and localize the handheld effectors with very good localization accuracy.

To what extent can users perceive the avatars in extra personal space to be their own bodies? This predictive machinery points to the importance of the visual motor correlations in embodying onto the avatar. For example, an illusory body ownership can be created over an invisible body via visual-motor synchronization (Kondo et al., 2018). While wearing a HMD, participants saw left and right white gloves and socks in front of them, at a distance of 2m, moved along in a virtual room. The visual-motor synchronicity of hands and feet were adequate to create the illusion that the moving virtual gloves and socks were part of participant's own body. This illustrates that humans are more fluid in integrating real and virtual environments than we thought previously. Are there some minimal or necessary conditions where users could blur the boundaries of real and virtual environments and perceive the actions of the avatar to be part of their extended personal space? For example, humans perceive their own mirror reflections to be part of their extra personal space. The question here is the extent to which the visual motor correlations impact users' projecting themselves into the virtual environments and act as if they themselves would be impacted by the consequence of the actions. Given that users are shut off from the real world while using the HMD, there are a number of advantages in examining the integration of real and low-cost simulated environments.

In this research, we look into real time processing measures as we manipulate an effector used by a participant in a simulated environment on a simulated task. We vary the effector to become more congruous with the real world tool they might use to do the same task. In particular, we set up a simple task to cut objects with a knife, and test certain *implicit* task measures as users perform the task but with an empty hand, vs. when holding a wand, vs. when holding a prop knife to interact with the simulated environment. Previous studies did not find realism in effectors resulted in significant differences in performance metrics, and occasionally reported some differences in subjective ratings (Freeman, et al., 2012; Reinhardt & Hurtienne, 2018). We think the explicit performance metrics and subjective evaluations of the user experiences could result from participants' strategic decision making. In the current study, we contrast the significant dimensional differences among effectors and make predictions as to whether there might be implicit differences in participants' behavioral repertoire, which is less likely to be modulated explicitly. We will explore the following hypotheses.

Physicality Hypothesis

Given that the knife and the wand in our study are matched in terms of weight and length, the physical properties of device-based effector vs. open hand are significantly different. Thus the *implicit* task performance in terms of cut

location (i.e., the cut location index) will be expected to be different between device-based effector (knife or wand) vs. open hand gesturing. We do not make predictions in terms of total time on task. It is reasonable to think that it takes longer time to cut when holding an actual physical object. However, if holding a physical object primes an awareness of the risk, then it is possible that the total time on task will be longer.

Risk Perception Hypothesis

Given that the knife is the only effector that could trigger the perception of risk (Aneli, Borghi, & Nicoletti, 2012; Brogni, Caldwell, & Slater, 2011; Liu, Cao, Chen, & Wang, 2017; Zhao, 2017), there would be significant differences in total time on task between the knife condition vs. the non-knife conditions (wand, and open-hand). Also the trajectories people take in moving the effector may differ between these, for example by being less smooth.

Method

The low-cost desktop virtual environments we developed for the experiments reported here aim to emulate a stationary work area, where the avatar puppets the motions of the user's arm in the real world, to allow the user to manipulate objects through the avatar's actions. A typical example we have implemented is a kitchen food preparation area, where the user has control of one arm of the avatar in the virtual space to manipulate knives, bowls, food and other objects. The user can have full control of the arm(s), and in more immersive versions can also control head gaze and direction.

We use the hands-free capability of the Kinect to test different conditions of physical embodiment in a vegetable cutting task, where the user has a (prop) knife versus a wand or a tracked empty hand when controlling an avatar with a virtual knife in the virtual environment. The Kinect device provides position information of the user's hand in real space, which is transmitted to the running Blender program as a set of three position coordinates. We have developed the framework to gather this positioning information reported by the Kinect, and then transmit them to a running Blender simulation. We recorded the effector position in pixels every 10 ms.

Participants

There were 53 undergraduate students recruited from a State University in the United States (Mean age = 24 years), of which 57% were female and 43% were male. In the data reported below, 2 participants' data were trimmed. Participants did not report previous exposure to such a task.

Experimental Design

We used an effector (knife condition, wand condition, vs. open hand condition) between subjects design. The weight and length of the wand were matched with those of the prop knife. Participants were randomly assigned to each of the experimental conditions.

Procedure The height of the monitor was positioned such that the location of the eyes and head of the avatar in the environment was consistent with the location of the human participant's eyes and head in the real world. The food preparation station was positioned right above the waist of a user of average height.

Once participants were successfully calibrated in Kinect, they went through a phase familiarizing themselves with Kinect. For the experimental trial, participants were given the following instructions: (1) they would see a cucumber being cut; (2) kitchen bell tone would signal their turn to make a cut; (3) make the cut where they desire. Participants saw the avatar cut the cucumber, however, they were not told where to cut exactly. As indicated in Figure 1, the length of the cucumber that remained to be cut was not significantly longer than the previous cuts made by the avatar. The idea is that how close the participant cut to the avatar's left hand fingertip would provide an indication of the extent to which participant treated the action as real (i.e., the temporary blurring the boundary of the real and situated environments). In addition, an experimenter indicated the appropriate starting position to facilitate accurate Kinect tracking.

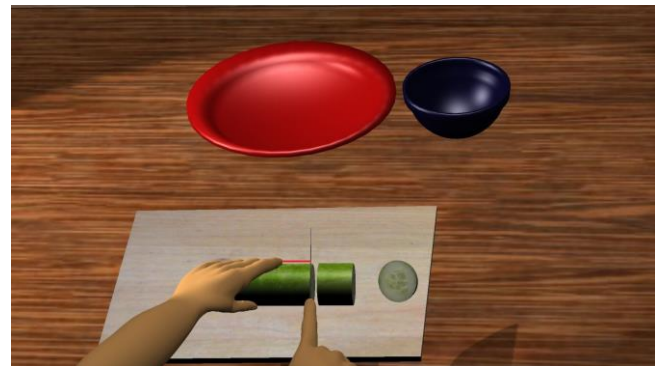


Figure 1. The cut location index is measured in blender pixel units, the distance from the avatars left hand holding the vegetable in the current task, to where the actual cut occurs indicated by the subject's actions in the experiment.

Results

We computed the following measures: (a) the cut location index, which is the distance in pixel coordinates from the participant cut locations to the finger tip of the avatar left hand in the computer mediated environment; and (b) the total task time, which was the total time from when the kitchen bell tone occurred indicating that it was the participant's turn, to when the participant moved the knife in the virtual environment in such a way that it indicated a cut should occur on the vegetable and the action to cut the vegetable was completed. In Figure 1, we depict the cut location index measure (in blender virtual environment pixel units). The measurement indicated in the figure was taken as the actual distance in 3 dimensions from the tip of the avatars fingertip, to the tip of the location where the cucumber cut location began on the vegetable being cut in

the experiment. The higher the value the cut location index is, the less risk there is to being injured.

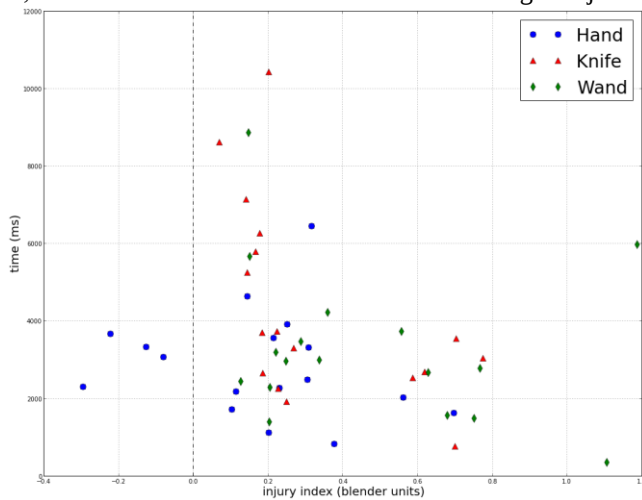


Figure 2. The cut location index scatter plot. Distance (in blender units) of the cut from the avatars hand vs. total task time when cut was made, for the 3 experimental conditions.

In Figure 2, we visualize the results of the cut location index measure together with the time to task measure. In this figure, we indicate the 3 different conditions (open hand, holding a prop knife and wand). Notice that for the knife condition especially subjects take the longest to complete the cut the closer the cut they are attempting to perform is to the avatar's left hand (which might result in potential injury, at a distance of 0.0 or less from the hand). Interestingly as well, all subjects who actually caused an injury to the hand, i.e., cuts that actually went into the avatars finger, were in the most incongruous condition, where the user in reality had an empty hand, but were controlling an avatar wielding a knife in the virtual environment. In general cuts that were more accurate and closer to the hand (without actually injuring the hand) usually took the most amount of time to make.

In Figure 3, we show the average cut location index for participants in each condition, along with 95% confidence intervals. The planned contrast showed that participants cut significantly closer to the fingertip when gesturing open hand than holding an effector, $t(48) = 2.61$, $p = 0.012$. This is consistent with the physicality hypothesis. Also of note, open hand performance on the cut location index showed the closest location index (e.g. cuts that were closest to the finger).

In Figure 4 we summarize the total task time measure. The planned contrast showed that participants used significantly longer time to cut with a knife than the other two non-knife forms, $t(48) = 2.06$, $p = 0.045$. This is consistent with the risk perception hypothesis. So while time to perform the cutting task with knives was significantly slower than the non-knife conditions. Users in the open hand condition might seem to be closest to the avatar's left hand fingertip,

but they were significantly more likely to actually cause injuries to the avatar hand in this condition.

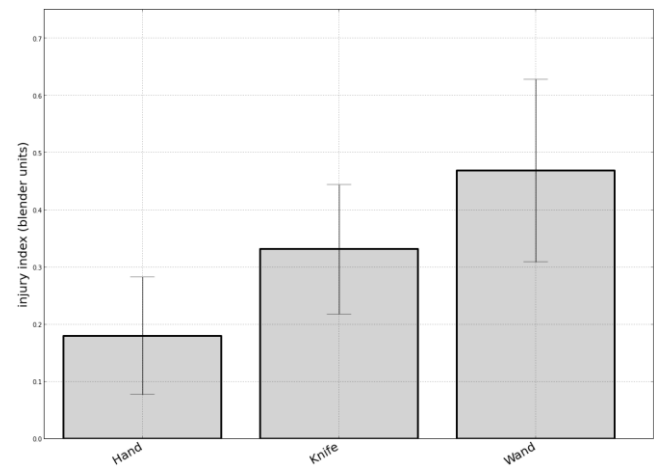


Figure 3. Summary of the cut location index measure. Mean cut locations are shown with whiskers indicating 95% confidence interval limits of the means.

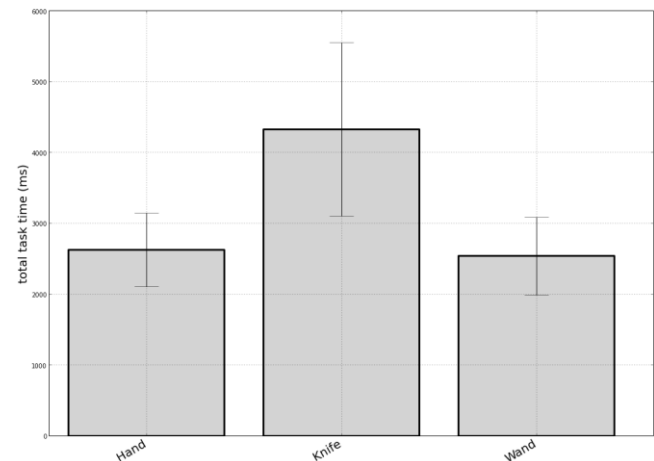


Figure 4. Summary of total task time measure. Figure indicates mean total task time for the three experimental conditions, with 95% confidence interval shown.

Conclusion

The experiment showed that effector congruence in the simulated environment does have some significant effects on task performance. Participants are more likely to treat the task as they would in the real world when the effector they use is the most realistic, and most in line with existing neural circuitry that would typically be employed to accomplish the task. For example, users were much less likely to cause injury to the virtual hand, when they were holding an object in their hand. We interpret this to mean that existing neural circuits and predictive machinery are more likely to be invoked in these more congruous conditions. Thus caution and appropriate location of the virtual knife in space were more likely to be achieved in order not to injure the virtual avatar. The most cautious behavior, in terms of time taken on the task, occurs when

holding a prop knife to manipulate the virtual environment. People using the knife took longer than people using a wand or not holding a prop.

What constitutes better performance when performing common food preparation using a knife? Speed and accuracy, as well as safety are all factors we would identify as important in separating novice level from expert level kitchen workers. Professional chefs are probably able to exceed on all three metrics, cutting quickly and accurately, but rarely if ever causing injury to themselves when using their dangerous tools.

We have done some analysis on the planning and execution of the task as indicated in the motor coordination measures of our participants in this simulated task. For example, in Figure 5 we show an analysis of the smoothness (or its absence of jerkiness) of the actual trajectories of the avatars hands in the virtual environment being controlled by the subjects hand movements through the Kinect effector. We have broken down the trajectories into 5 segments, and used the third time derivative (Hogan, 1984) to measure the smoothness of their trajectories. The whiskers represent 95% confidence intervals of the smoothness measure for each of the 5 trajectory segments. Knife performance differed significantly on this smoothness measure, especially in the middle part of the motion of the virtual knife on the task. These motion analysis measures show how different participants are treating the task when using the more congruent effector.

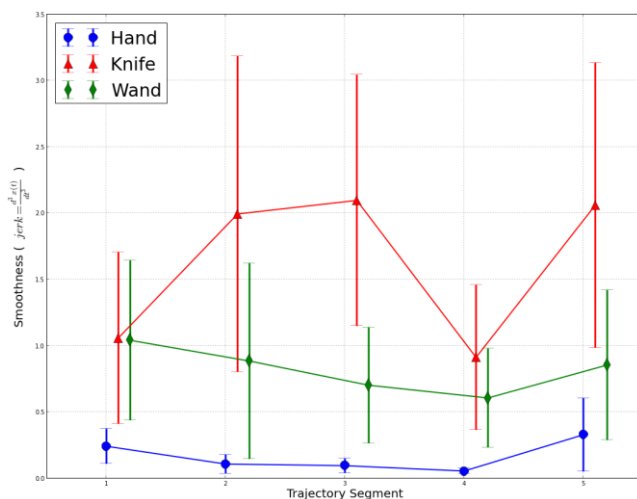


Figure 5. Measure of smoothness of subject's motion of the knife in the virtual environment, broken up into 5 equal length segments. Whiskers indicate 95% confidence interval on the smoothness measure.

Unlike the previous studies, the current results speak to the importance of looking into real time task planning and execution and showcase a paradigm in observing movements in developing embodied design (Fdili Alaoui, et al., 2015). The equivalence on the performance measures

does not necessarily speak to the ongoing differences in the users' minds. Let us draw an analogy. When people use a sharp vs a dull knife to prepare food, people would take more time and be more cautious with the sharp knife, but this does not mean people would slice into their finger tips or would not be able to use the full available length of the food being prepared. A simpler view of the effector risk perception hypothesis is that users would produce different vegetable cuts while using different effectors. The implication of this simpler view is that slight differences in effector or other aspects of the environments would lead to significant differences in action outcome. Such a simplified view is in effect inconsistent with the functional redeployment theory of cognition. It is also inconsistent with the finding in virtual reality that people treat the virtual environment as if it were real even though they know it is not real (Bailenson, 2018).

The paradigm developed in the current study shows potential to examine how the bi-directional interactions of changes in simulated environments influence subsequent user actions and vice versa as they unfold in real time (Dawley & Dede, 2014). Through systematic comparison, the current study give insight into what critical ingredients of intuitive touchless interactions should involve (Chattopadhyay & Debaleena, 2015; Gillies & Marco, 2016). When the congruency between the effector in the virtual environments and the tool used in the real world tasks supports the low-level visual motor contingency, users are more likely to incorporate the extra personal space into their behavioral repertoire.

Predictive views of embodied cognition that take into account how neural circuits are likely to be reused when experiencing a simulated environment are a rich conceptual framework to better understand how to improve immersion and learning outcomes when training in simulated environments. This theory could address a number of thorny issues. For example, why environments with minimal realism can still trigger the experience of immersion and why often environments with varying degrees of realism do not get rated differently when it comes to the subjective reports of user interaction experience. A system that provides less support to align with real world sensory motor contingencies, the more perceptual prediction errors will be generated along the way and the more hierarchical adjustments will have to be made to compensate for errors. This would lead to the greater probability of errors on the task and less satisfaction as reflected in the subjective reports of user experience.

Acknowledgements

This work is supported by a grant from US National Science Foundation (IIS-0742109, 0916749). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of US National Science Foundation. We thank Sarah Wang and Paweena Kosito in assisting various

stages of this work, in particular in the phase of data collection.

References

- Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 33(04), 245-266.
- Anderson, M., Richardson, M., & Chemero, A. (2012, 10 1). Eroding the Boundaries of Cognition: Implications of Embodiments. *Topics in Cognitive Science*, 4(4), 717-730.
- Aneli, F., Borghi, A., & Nicoletti, R. (2012). Grasping the pain: Motor resonance with dangerous affordances. *Consciousness and Cognition*, 21, 1627-1639.
- Barrett, L., & Simmons, W. (2015, 7 28). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16(7), 419-429.
- Bailenson, J. (n.d.). *Experience on demand: what virtual reality is, how it works, and what it can do*. New York: W.W. Norton
- Broggi, A., Caldwell, D., & Slater, M. (2011). Touching sharp virtual objects produces a haptic illusion. In R. Shunmaker (Ed.), *Lecture Notes in Computer Science: Virtual and Mixed Reality* (pp. 234-242). Berlin Heidelberg: Springer-Verlag.
- Chattopadhyay, D., & Debaleena. (2015). Toward Motor. *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces - ITS '15* (pp. 445-450). New York, New York, USA: ACM Press.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 33(4), 181-204.
- Dawley, L., & Dede, C. (2014). Situated Learning in Virtual Worlds and Immersive Simulations. In L. Dawley, & C. Dede, *Handbook of Research on Educational Communications and Technology* (pp. 723-734). New York, NY: Springer New York.
- Dourish, P. (2001). *Where the Action is: The Foundations of Embodied Interaction*. MIT Press.
- Fdili Alaoui, S., Schiphorst, T., Cuykendall, S., Carlson, K., Studd, K., & Bradley, K. (2015). Strategies for Embodied Design. *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition - C&C '15* (pp. 121-130). New York, New York, USA: ACM Press.
- Freeman, D., Hilliges, O., Sellen, A., O'Hara, K., Izadi, S., & Wood, K. (2012). The role of physical effectors in motion video gaming. *Proceedings of the Designing Interactive Systems Conference* (pp. 701-710). ACM.
- Gillies, M., & Marco. (2016). What is Movement Interaction in Virtual Reality for? *Proceedings of the 3rd International Symposium on Movement and Computing - MOCO '16* (pp. 1-4). New York, New York, USA: ACM Press.
- Grandhi, S. A., Joue, G., & Mittelberg, I. (2011). Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 821-824). ACM.
- Hogan, N. (1984). An organizing principle for a class of voluntary movements. *The Journal of Neuroscience*, 4(11), 2745-2754.
- Hornecker, E. (2011). The Role of Physicality in Tangible and Embodied Interactions. *Interactions*, 19-23.
- Ishii, H. (2008). Tangible bits: beyond pixels. *Proceedings of the 2nd international conference on Tangible and embedded interaction* (pp. xv-xxv). ACM.
- Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., & Zigelbaum, J. (2008). Reality-based interaction: a framework for post-WIMP interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 201-210). ACM.
- Kirsh, D., & David. (2013, 3 1). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction*, 20(1), 1-30.
- Kondo, R., Sugimoto, M., Minamizawa, K., Hoshi, T., Inami, M., & Kitazaki, M. (2018). Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity. *Scientific reports*, 8(1), 7541.
- Liu, P., Cao, R., Chen, X., & Wang, Y. (2017, 6 1). Response inhibition or evaluation of danger? An event-related potential study regarding the origin of the motor interference effect from dangerous objects. *Brain Research*, 1664, 63-73.
- Lu, S., & Harter, D. (2016). Toward a cognitive processing theory of player's experience of computer mediated environments. *Proceedings of CHI Play '16*. Austin, TX: ACM Sheridan Printing .
- Lu, S., Harter, D., Kosito, P., & Kotturu, P. (2014). Developing low-cost training environments: How do effector and visual realism influence the perceptual grounding of actions? *Journal of Cognitive Education and Psychology*, 3-18.
- Pandey, R., Pidlypenskyi, P., Yang, S., & Kaeser-Chen, C. (2018). Efficient 6-DoF Tracking of Handheld Objects from an Egocentric Viewpoint. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 416-431).
- Reinhardt, D., & Hurtienne, J. (2018). The impact of tangible props on gaming performance and experience in gestural interaction. Stockholm, Sweden: ACM.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549-3557.
- Toussaint, M. (2009). Probabilistic inferences as a model of planned behavior. *Künstliche Intelligenz*, 3, 23-29.
- Zhao, L. (2017). Separate pathways for the processing of affordance of neutral and dangerous object. *Current Psychology*, 36(4), 833-839.