The Uncanny Valley of Embodied Interaction Design

Francesco Cafaro[1], Leilah Lyons[1,2], Jessica Roberts[2], Josh Radinsky[2]
[1]Computer Science [2]Learning Sciences
University of Illinois at Chicago
{fcafar2, llyons, jrober31, joshuar} @uic.edu

ABSTRACT

The "Uncanny Valley" theory explains the counter-intuitive phenomenon where people may get suddenly uncomfortable with an artificial entity when it becomes very similar to humans. We propose the existence of an "uncanny valley" for embodied interaction, when a user's body motions in the physical space (the *locus* of interaction) are incompletely mapped into effects in the virtual space (the *focus* of interaction). It is generally assumed that this mapping should be as veridical as possible to promote seamless embodied interaction. Many design factors (e.g., synchronicity, sensitivity, shared realism) contribute to veridical locus-focus mapping. We intentionally varied the level of veridicality of these different factors, affecting how the user's movements were mapped to virtual effects. Our results indicate that there is a dip (valley) in user preferences when the design contains mixed degrees of veridicality. Thus, when one veridical dimension is limited, designers should likewise reduce the veridicality of other dimensions.

Author Keywords

Embodied Interaction; Uncanny Valley; Design; Realism; Locus of Interaction; Focus of Interaction.

ACM Classification Keywords

H.1.2 User/Machine Systems: Human Factors. H.5.2 User Interfaces: Interaction styles, Theory and methods.

INTRODUCTION

In "Bukimi No Tani" ("The Uncanny Valley") [10] Japanese roboticist Masahiro Mori observed that viewing a prosthetic hand may trigger feelings of eeriness and repulsion, whereas seeing a human hand or a mechanical hand does not (Figure 1). Mori's theory has since been applied to areas like the design of 3D videogames and animated movies [6]. These examples all involve taking a virtual entity and attempting to bring its appearance closer to that of a human, in other words, to more effectively bring the virtual into our human world. Mapping the user's body (e.g. actions, gestures, and movements) into a virtual world is a central component of many embodied interaction designs. Embodied interaction [3] should create a "seamless connection" between our bodies and the virtual space. But designers are often limited by the capacities of current interactive technologies to attain this seamlessness. For example, with many whole-body interaction designs, there is usually a physical separation between what we dub the *locus* of the interaction (the human body) and the *focus* of the interaction (typically a display screen): they comprise two parallel worlds which need to be mapped on to one another [1]. To use the system, users need to understand which bodily action triggers a specific effect (animation, sound, etc.). Mapping gestures and movements into effects that feel natural and familiar to the users is a nontrivial challenge [13].

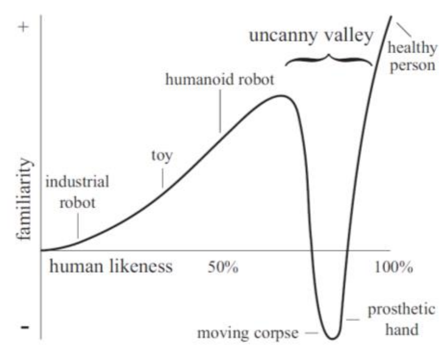

Figure 1. Mori's Uncanny Valley (as illustrated by MacDorman et al. [7]). In robotics, increased human-likeness does not linearly correspond to a higher feeling of familiarity: if a robot becomes "too" human, it may look repulsive.

Attaining an embodied interaction design which is "familiar" (i.e. the feeling of being "at home," "at ease" [5]) to the user has often involved increasing the "realism," or veridicality of the locus-focus mapping, of that interaction. For instance, the handle bar metaphor introduced in [12] for performing mid-air gestures works because of the "physical familiarity" with a rotisserie skewer. Thus, in order to bring the focus of the interaction closer "to the world of the user" [3], many interaction designers recommend seeking a veridical mapping between user actions and effects on the screen. For instance, [4] provides a framework to increase the "realism" of embodied experiences; [8] presents a technique to incorporate the user's body orientation and posture into a virtual 3D world. The study presented in this paper was designed to examine these assumptions that maximizing the veridicality of the body-virtual mapping reliably improves embodied interaction experiences.

DIS '14, June 21 - 25 2014, Vancouver, BC, Canada.

ACM 978-1-4503-2902-6/14/06…$15.00.
http://dx.doi.org/10.1145/2598510.2598593

BACKGROUND AND RELATED WORK

Mori's theory of the "Uncanny Valley" [10] is based on Jentsch's description of physiological state of "uncanny": an uncomfortable "lack of orientation," which arises when we experience something "uncertain" [5]. In this paper, we use the word "uncanny" with its original meaning of "disorienting," "uncertain," "unfamiliar" [5] to refer to the disorientation that a user may experience when her/his mental model of the input-output relationship is disaffirmed. We do not imply any reference to the sense of "eeriness" found in Freud's use of the term "uncanny" [4].

The Uncanny and the Perceptual Magnet Effect

According to [11], the uncanny effect may be caused by conflicting sensory clues. [9] introduces a probabilistic model based on the "perceptual magnet effect": people judge perceptual stimuli which are close to different category boundaries as more dissimilar than stimuli that are away from that boundary. This may produce non-monotonic distributions of the "familiarity" of a stimulus [9] –i.e. the uncanny valley shown in Figure 1.

Reality-Based Interaction (RBI)

In [4], Jacob et al. introduce a framework for bringing "reality" into the design of "post-WIMP interfaces" –a broad class that includes mixed and augmented reality, tangible interaction, and ubiquitous computing. Interfaces should be as "realistic" as possible, i.e. they should resemble what people experience in the real world, except in circumstances where trading off a small "degree of realism" can favor "other desired qualities," such as the expressive power of the system, its ergonomics, the efficiency, etc. [4]. In particular, they suggest designers should analyze these trade-offs along four different "themes from the real world": "Naïve Physics," i.e., people's common sense understanding of the physical rules that hold in the real world; "Body Awareness & Skills," i.e., users' awareness of their own physical bodies and of what they can do; "Environment Awareness & Skills," i.e., how we navigate our surrounding space; "Social Awareness & Skills," i.e., user's awareness of the presence of others [4].

In this paper, we uncover that there may be some risk to trading in a small "degree of realism." Our hypothesis is that an "uncanny valley" of embodied interaction may introduce a non-linear relationship between the traded realism and the user's expectations. We recommend that when a designer compromises one factor influencing the veridical mapping between user's actions and effects, the veridicality of the other factors should also be reconsidered.

METHODS

The study took place at a small museum and cultural center in Chicago. Nine museum visitors participated in the study.

Use Scenario: Whole-Body Interaction with Census Data

The testbed for this study is a prototype museum exhibit (*CoCensus*) designed to allow visitors to explore a visualization of data (e.g., US Census data) via embodied interaction [2]. When entering the exhibit, a user selects a data subset to control (e.g., "residents with German ancestry"). Approaching a 65" LCD screen, she/he sees "her/his" data subset on a display (Figure 2).

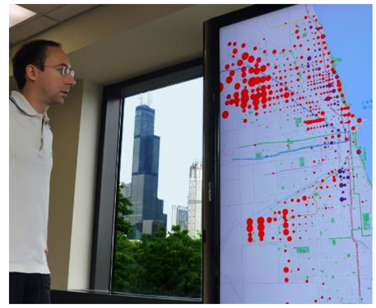

Figure 2. One user interacts with *CoCensus*. The map shows scaled centroids ("bubbles") representing the population of his self-identified ethnic group.

Design of the Study

Only one user at a time was allowed in the room, to avoid influencing other participants. We selected a very specific body movement, jumping, because it is a very intentional action on the part of the user (whereas lateral movement may be less conscious in exhibit galleries). The effect of this body movement on the virtual world is to make the user's data subset move vertically (when the user jumps, his or her data bubbles "jump" on the map, and then fall down to their original places). The design rationale is explained in detail in [2], but briefly: moving the data subset away from its georeferenced "root" location helps visitors notice distribution patterns. We designed and implemented 12 different versions of this "jump" effect by manipulating three different factors affecting the veridicality of the locus-focus mapping: (1) sensitivity to the nuance of the user's motion, (2) temporal synchronicity, and (3) shared physical realism. The first two factors explore "*Body Awareness & Skills"* realism [4], by manipulating the "tightness" of the connection between the user's body (locus) and the virtual realm (focus). The third factor explores "*Naïve Physics"* [4]. We left factors related to "Environment Awareness & Skills," and "Social Awareness & Skills" to future work. The specific manipulations tested were:

Sensitivity to User's Motion. The mapping between the path of the user's motion (here, their body's y-coordinate in the physical world as they jump) and the path of the effect (the y-coordinate of the data bubbles on the screen). (1) Threshold. The animation of the user's data subset is triggered when the user's jump crosses a pre-defined threshold. After the trigger, the animation will follow its path and the user cannot control it. (2) Absolute. The system continuously tracks the user. Maximum height of the animation depends on the height of the user's jump.

Temporal Synchronicity. The mapping of the start and stop of the user's motion to the start and stop of the effect. (1) Asynchronous. There is a short lag between the jump and the animation, which begins a few *ms* after the user starts jumping and ends 1 to 3 seconds after the user landed (the duration differs because it depends on both the Motion

Sensitivity and Shared Physical Realism conditions). (2) Synchronous. The animation timing is closely mapped to the jumping motion. The animation starts when the user's jump begins, and ends when the user's feet touch the ground.

Shared Physical Realism. How much the virtual objects respected the "physics" users expect from real-world objects. We tested three different levels of "shared physical realism" in the animation of the bubbles: (1) Low Shared Physical Realism. All bubbles, regardless of their size, rise up with the same speed, rise to the same height, and return back simultaneously to their original positions – effectively moving as a unit. (2) Medium Shared Physical Realism. All the bubbles rise to the same height at the same time, but they then fall with different speeds proportional to their mass: bubbles with larger masses return to their original positions faster than those with smaller masses. (3) High Shared Physical Realism. The bubbles are given the same initial forces. As F=ma, their acceleration depends on their mass in the rising phase. Each bubble thus rises to a different height depending on its mass, and simulated "air resistance" causes the smaller bubbles to drift down while larger bubbles quickly fall.

Experimental Procedure

Visitors were asked to jump and take note of the system response for each of the 2x2x3=12 virtual effects. We randomized the presentation order. After making a few jumps per interaction effect, visitors were asked to rate the effect on a scale from one to three (three being most satisfied). Interviews were fully recorded and transcribed.

It is worth noting that we structured the study as a 2x2x3 to explore mutual interaction effects. Similar to [4], we believe that "realistic experiences" are a combination of multiple perceptual stimuli –we were not seeking an uncanny valley for each single realism factor.

RESULTS AND DISCUSSION

A three-way ANOVA was conducted to examine the effect of Motion Sensitivity, Temporal Synchronicity and Shared Physical Realism on the satisfaction score.

When analyzing each realism factor, we observed that: (1) Users seem to prefer the less-realistic Threshold condition, X=2.26, SD=0.70 compared to X=1.89, SD=0.74 for Absolute, $F(1, 53)=9.680$, $p<0.002$ –we believe this result is due to the physical effort involved with higher jumps. (2) No significant differences ($F(1, 53)=0.320$, $p>0.573$) were found between the scores of Synchronous (X=2.09, SD=0.78) and Asynchronous effects (X=2.06, SD=0.70). (3) Users expressed a linearly increasing preference for higher levels of Shared Physical Realism: X=1.86, SD=0.75 for the low condition; X=2.08, SD=0.70 for the medium; X=2.28, SD=0.73 for the high ($F(2,35)=6.221$, $p<0.003$).

However, the analysis of the mutual interaction of the three realism factors unveils an interesting phenomenon. The ANOVA showed a statistically significant interaction between Motion Sensitivity and Temporal Synchronicity ($F(1, 26)=5.520$, $p<0.025$), and between Temporal Synchronicity and Shared Physical Realism ($F(2, 17)=4.340$, $p<0.016$).

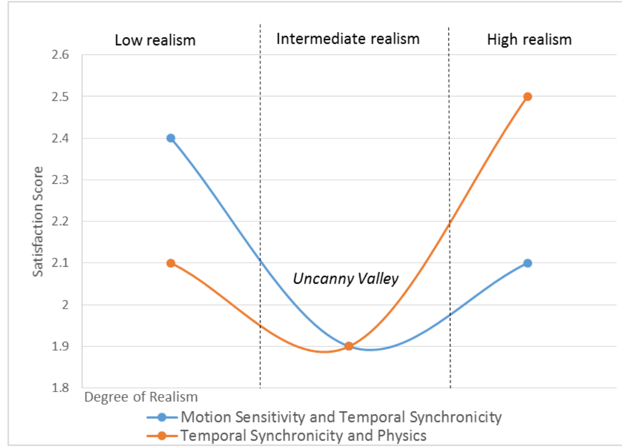

Figure 3. The blue curve depicts the average satisfaction scores for conditions with "low composite realism," pairing low sensitivity ("threshold") with low synchronicity ("asynchronous"); "high composite realism," pairing "high sensitivity ("absolute") and high synchronicity ("synchronous"); and "mixed composite realism," pairing mixed levels of realism (e.g., high and low). The orange curve similarly depicts the average scores of animations varying their combinations of "temporal synchronicity" and "shared physical realism."

To investigate these interactions, we created a "composite realism" variable to indicate the overall level of veridicality of each animation effect: first, we computed the average satisfaction scores for each of the 12 effects; second, we grouped effects depending on the levels of veridicality (all-low, mixed, and all-high) for the two parameters of interest, and then we averaged the score of each group. For example, for the Motion Sensitivity and Temporal Synchronicity pairing, the composite satisfaction score was computed by grouping the satisfaction scores of conditions with low sensitivity and low synchronicity (regardless of their level of shared physical realism), of those with high sensitivity and low synchronicity, and of those with mixed levels of sensitivity and synchronicity. For the Temporal Synchronicity and Shared Physical Realism pairing, we followed the same procedure, and we grouped the effects with medium shared physical realism within the class of effects with mixed levels of temporal synchronicity and shared physical realism –see blue line in Figure 3 for the grouped Motion Sensitivity/Temporal Synchronicity effects, and the orange line for the grouped Temporal Synchronicity/Shared Physical Realism effects. We observed that effects that shared a similar level of realism (either all-high or all-low) outperformed those that had mixed veridicality. In other words, "trading in" a degree of realism decreased user satisfaction. "Adding in" a degree of realism to the low-realism "base case" (e.g. making a Threshold animation Synchronous, rather than

Asynchronous) *also* decreased user satisfaction. This non-monotonicity is a characteristic of an "uncanny valley."

We believe that this phenomenon is akin to the notion of the "*perceptual magnet effect*" [9] to embodied interaction. A mid-level of veridicality leads the users into a state of "uncertainty" (i.e. an uncanny valley): they perceive contrasting stimuli – one veridical characteristic of the user's jump, either timing or magnitude, seems to have a role in enhancing realism, while the other characteristic suggests the opposite. The mapping between the user's body (locus of interaction) movement and the effect (focus of interaction) is, instead, much clearer at the two extremes of realism. An intermediate state seems to confuse the user.

DESIGN IMPLICATIONS

The contribution of this paper is to highlight a risk that may unveil itself if designers additively combine design factors that can contribute to "realism," because they do not combine linearly nor monotonically. This phenomenon has the following implications for the design of interactive, embodied experiences: (1) Different perceptual stimuli interact with each other, e.g., the realism in how the users' movement is represented (Body Awareness & Skills) impacts users' expectations of physical rules not directly related with their body movement (Naïve Physics). Designers should be aware that manipulating one factor affecting "realism" may demand changes in other factors (e.g., avoid using a visualization with realistic simulated physics unless their sensors can accurately detect the timing or magnitude of user motions). (2) Coherence may be more important than realism for supporting the usability of an embodied experience. Conditions with "low composite realism" outperformed those with "mixed composite realism." Designers should consider if there is a minimal degree that is "real enough" for their application, rather than assuming maximal realism leads to maximal usability.

CONCLUSION

In this paper, we showed how the concept of the "Uncanny Valley" can be extended to Embodied Interaction. Increasing realism does not always correspond to a better embodied user experience. Intermediate levels of realism may have a significant negative impact on the way users perceive some pairings of body movements and effects, especially when the perceptual stimuli lead the user to conflicting ideas of how their body movements (locus of interaction) control the effects shown on the virtual worlds (focus of the interaction).

This work represents only a first pass at exploring the link between realism, familiarity, and embodied interaction design, but it reveals that designers need to tread cautiously when combining different methods of affecting interaction "realism," lest their design fall into the "uncanny valley" of embodied interaction.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under NSF INSPIRE 1248052: Studying and Promoting Quantitative and Spatial Reasoning with Complex Visual Data across School, Museum, and Web-Media Contexts. We would like to thank Vikrem Singh Bhagi for his help with the data collection, and Natalie Jorion for her help with the statistical analysis.

REFERENCES

1. Cafaro, F., Lyons, L., Kang, R., Radinsky, J., Roberts, J., and Vogt, K. Framed Guessability: Using Embodied Allegories to Increase User Agreement on Gesture Sets. In *Proceedings of TEI'14*, ACM (2013), 197–204.

2. Cafaro, F., Panella, A., Lyons, L., Roberts, J., and Radinsky, J. I see you there!: developing identity-preserving embodied interaction for museum exhibits. In *Proceedings of CHI'13*, ACM (2013), 1911–1920.

3. Dourish, P. *Where the action is: the foundations of embodied interaction*. MIT Press, Cambridge, MA, 2001.

4. Jacob, R. and Girouard, A. Reality-based interaction: a framework for post-WIMP interfaces. In *Proceedings of CHI'08*, (2008), 201–210.

5. Jentsch, E. On the Psychology of the Uncanny. *De Psychiatrisch-Neurologische Wochenschrift 8*, 22-23 (1906).

6. MacDorman, K.F., Green, R.D., Ho, C.-C., and Koch, C.T. Too Real for Comfort? Uncanny Responses to Computer Generated Faces. *Comput. Hum. Behav. 25*, 3 (2009), 695–710.

7. MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S. J. & Ishiguro, H. Assessing human likeness by eye contact in an android testbed. *XXVII Annual Meeting of the Cognitive Science Society*, (2005).

8. Maciel, A., Nedel, L.P., Jorge, V. a. M., Ibiapina, J.M.T., and Silva, L.F.M.S. Reality cues-based interaction using whole-body awareness. In *Proceedings of SAC '10*, ACM (2010), 1224.

9. Moore, R.K. A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific reports 2*, (2012), 864.

10. Mori, M. Bukimi no tani (The Uncanny Valley). *Energy*, (1970).

11. Seyama, J. and Nagayama, R. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments 16*, 4 (2007), 337–351.

12. Song, P., Goh, W.B., Hutama, W., Fu, C.-W., and Liu, X. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of CHI'12*, ACM (2012), 1297.

13. Wobbrock, J.O., Morris, M.R., and Wilson, A.D. User-defined gestures for surface computing. In *Proceedings of CHI'09*, ACM (2009), 1083–1092.