

UC Irvine

Embodiment and Performativity

Title

Intimate Encounters: the Mixed Reality Paradigm and Audience Responses

Permalink

<https://escholarship.org/uc/item/1964v9cb>

Author

Cleland, Kathy

Publication Date

2009-12-12

Peer reviewed

Intimate Encounters: the Mixed Reality Paradigm and Audience Responses

Dr Kathy Cleland
Digital Cultures Program
University of Sydney
AUSTRALIA
+61 2 9351 4721

kathy.cleland@usyd.edu.au

ABSTRACT

Our lived experience is that of a mixed reality. The online and the offline—the real and the virtual—are becoming increasingly blurred and enmeshed. Humans take on the virtual form of avatars to interact in cybernetic virtual environments. Computers become ‘social actors’ interacting face-to-face with human audiences through the screen interface or by leaving the screen to interact as physically embodied robotic entities. This paper investigates the phenomenological nature of our embodied and lived experiences with both screen-based and physically embodied entities and explores the way sensorial and emotional affects are distributed between the physical and the virtual. Examples are drawn from a range of new media art projects focusing on the audience experience of different screen-based (virtual) and embodied (robotic) entities and the mixed reality terrains they inhabit with their human audiences.

Categories and Subject Descriptors

A.0 Conference Proceedings. H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities. J.5 [Arts and Humanities]: Fine arts, Performing arts.

General Terms

Human Factors, Theory.

Keywords

Mixed reality, phenomenology, robotics, avatars, virtual reality, virtual worlds, art, audience response, new media art, mirror neurons.

1. INTRODUCTION

Our experience of the world is increasingly becoming an experience of a mixed reality, a complex blend of the real and the mediated. In the 21st century, the proliferation of images and media forms is leading to an increasingly diverse and complex media ecology where different image and media types co-exist and converge through the meta-medium of digital technology.

The online and the offline—the real and the virtual—are becoming increasingly blurred and enmeshed. Humans take on the virtual form of avatars to interact in cybernetic virtual environments. Computers become “social actors” interacting face-to-face with human audiences through the screen interface or by

leaving the screen to interact as physically embodied robotic entities. The distinction between screen-based virtual agents and robots is also starting to blur as agent systems combine elements of both virtual (screen-based) and physical embodiment. [13]

How can we theorise and think about this new mixed reality paradigm and the mixed reality selves and others that inhabit it? Ron Burnett describes the combination of real and virtual elements in a shared hybrid image environment as a “middle space” [4: xx], and this hybrid reality is also expressed in the terms “augmented reality” and “mixed reality” used in computer science and increasingly by the arts and entertainment industries.¹ Computer scientist Paul Milgram suggests the idea of a “virtuality continuum” to describe the different levels of blended or mixed reality between the real environment of the physical world and the virtual environments generated entirely by computers. [19] The intermingling and interpenetration of physical and virtual environments in mixed reality and ubiquitous computing environments has been explored in a number of recent papers, for example Andy Crabtree’s and Tom Rodden’s ‘Hybrid ecologies: understanding cooperative interaction in emerging physical-digital environments’ (2008) and Eric Kabisch’s ‘Datascape: a synthesis of digital and embodied worlds.’ [17]

In this paper I am particularly interested in looking at our lived experience of these mixed reality and hybrid spaces and the entities that inhabit them. How do these interactions *feel*? How do we *experience* them? What are the similarities and differences in our interactions with virtual screen spaces and personas and with physically embodied entities and robots? How do audiences perceive and respond (physically, intellectually and emotionally) to virtual screen personas and how do they perceive and respond to physically-embodied three dimensional entities such as robots that share our physical space?

2. AUDIENCE RESPONSES TO SCREEN-BASED AND ROBOTIC ENTITIES

When one of Mari Velonaki’s *Fish-Bird* robots follows me around a gallery space my whole body responds. As the robot moves towards me, I move away—we dance back and forth. The

¹ The terms “augmented reality” [1, 10] and “mixed reality” [19] are both used to describe the blending of real and virtual (digitally created) environments and experiences.

experience is visceral and playful—physically and emotionally engaging. The robot stops and prints out a message for me—I bend down to pick it up...it says, “you give me life”...then I wait for the next message.

Can a virtual screen experience evoke a similarly strong emotional reaction?

In the virtual world *Second Life*, my onscreen avatar Bella Bouchard can move around freely, but my physical body remains immobile behind a computer screen. This would appear to be a very different experience from the highly engaged physical interaction I have experienced with the *Fish-Bird* robots. What effects, if any, do the experiences of my virtual avatar body have on my physical body? Can the delegated physical agency I have through my screen avatar have an affect on my physical body on the other side of the screen?

Visiting Gazira Babeli’s 2007 exhibition *Collateral Damage* in *Second Life*, my onscreen avatar is possessed by one of her performance codes as it sits on a virtual chair in her work *Avatar on Canvas*. As soon as my avatar body makes contact with the chair, it becomes grotesquely deformed, its limbs stretching out of alignment and becoming grossly extended and distorted. The effects of this performance code continue even after my avatar leaves the exhibition. I can’t control my body. I feel mortifyingly embarrassed by my avatar’s distorted body and movements and by my inability to control my own body and movements. I don’t want to be seen like this. I am surprised by the strength of my feelings of discomfort and embarrassment. I feel the same way as I would if a practical joke was played on me in real life. While my physical body may be safely distant from the experiences of my onscreen avatar identity, nevertheless it reacts with physical sensations, my palms sweat, I feel my cheeks grow warm and my breathing rate intensifies. The online experience is not limited to intellectual and psychological engagement; rather, sensation and affect are distributed between the virtual and the physical.

What, if anything, differentiates these two experiences, the one an interaction with a physical object and the other a purely screen-based encounter?

Velonaki and her colleagues at the Social Robotics Unit at the University of Sydney cite recent research that suggests that multi-modal interaction, “encompassing many senses, can have a synergistic effect in increasing the ‘believability’ of interaction” between humans and machines. They also argue that “physical embodiment of a virtual agent contributes strongly to engaging interactions between a human and a ‘character’ because the character physically inhabits the same space as the human, with all the implications that this co-inhabitation brings” [27: 514-515].

This perspective is also a long-held position of media artist and theorist Simon Penny who comments: “I am particularly interested in interaction which takes place in the space of the body, in which kinesthetic intelligences, rather than ‘literary-imagistic’ intelligences play a major part.” [22]

Clearly, multi-modal interaction incorporating senses beyond the audio-visual is easier to achieve in an interaction with an embodied robotic presence than with a screen image. This would appear to give physical objects in a gallery, especially self-moving objects such as robots, a distinct advantage in the reality stakes over screen images. The ability of an art object to move into an audience’s physical space and interact with them there triggers a

psycho-physiological response to that movement as we assess whether that object presents a threat (triggering a flight or fight response), or an opportunity for a more positive type of interaction. Do we want to run away or do we want to engage?

However, while my visceral ‘kinesthetic’ hunch is strongly in agreement with Penny and Velonaki et al., the distinction between audience responses to embodied physical entities and virtual screen based entities is not at all clear cut.

Although the ‘virtual’ realities of screen-based representations are typically seen as being of a different order of reality than the material physicality of so-called ‘real life’ (RL), this opposition between the virtual (unreal) and the physical (real) is a shaky one in our new mixed reality paradigm. Don Ihde argues that in a phenomenological sense “both RL and VR are part of the lifeworld and VR is thus both “real” as a positive presence and a part of RL.” [15: 15] A phenomenological approach is also a key part of Paul Dourish’s notion of “embodied interaction” in HCI and his investigation of the way computational systems are embedded into environments and social contexts. As Dourish comments, “Physically, our experiences cannot be separated from the reality of our bodily presence in the world; and socially, too, the same relationship holds because our nature as social beings is based on the ways in which we act and interact, in real time, all the time.” [8: 18]

New media theorist Mark Hansen also stresses the importance of the human body as the key interface in the “interpenetration of physical and virtual spaces.” [11: 3] Everything we experience—whether it’s swimming in the ocean or playing a video game—is experienced by the human body and the human sensorium. Even with virtual reality technologies, where the myth of disembodied experience is at its highest, this experience is still necessarily mediated by (and constituted by) the physical body. As N. Katherine Hayles comments: “Cyberspace, we are often told, is a disembodied medium. ... In a sense, [this is] correct; the body remains in front of the screen rather than within it. In another sense, however, [this is] deeply misleading, for [it] obscure[s] the crucial role that the body plays in constructing cyberspace. In fact, we are never disembodied. ... Far from being left behind when we enter cyberspace, our bodies are no less actively involved in the construction of virtuality than in the construction of real life.” [12: 1]

The perceptual experiences and affective responses generated by media images can feel just as real as those generated by the physical world. Jay David Bolter and Richard Grusin argue that: “Media have the same claim to reality as more tangible cultural artefacts; photographs, films, and computer applications are as real as airplanes and buildings.” [3: 19]

In *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places* (1996) Byron Reeves and Clifford Nass argue that there is no essential or functional difference in how the brain responds to the ‘real’ physical world, and how it responds to media images and artificial entities. According to Reeves and Nass our “old brains” have not yet caught up with our new media technologies and they do not have the sophistication to distinguish between a real physical object in the world and a media image or robotic simulation of that same object. This means that people tend to respond in essentially the same way to screen images of a person or a virtual computer persona as they would to a real person. Even though we

may be consciously aware that screen images and simulated entities are not real, nevertheless, we have an ingrained unconscious tendency to treat them as if they were.

Reeves' and Nass's experiments demonstrate that people's responses to media images and computers show essentially the same physiological and behavioural patterns as real world responses, even though the participants clearly know that what they are responding to is not real. Ingrained physiological responses (such as reacting to sudden movement and sound) and social responses (such as a tendency to be polite) are carried over from the physical world into our interaction with screen images and artificial characters. Images that move on a screen (especially in the audience's peripheral vision) trigger similar responses to those of objects in the physical world. Faces that get bigger (i.e. appearing to move closer to the viewer) or that look directly at the viewer also generate instinctive physiological responses.

Recent research into the phenomenon of mirror neurons also suggests a neuroscientific basis for this physical and emotional response to screen images and artificial entities. Experiments show that areas of the brain collectively known as the 'mirror neuron system' respond not only when individuals perform an action themselves but also when they watch someone else perform that action. Watching someone pick up an object triggers a similar response to actually picking up the object yourself. Screen-based actions and experiences also trigger mirror neuron responses and corresponding physical motor responses; pornography is a key example here. [21, 23] Similarly, watching someone cry, being hit, or expressing emotion, can trigger empathic mirror neuron responses so that those actions and emotions are experienced by the person watching. [9, 26, 16]

So, if we can respond just as strongly to screen-based entities as we do to physically embodied robots, what other techniques can artists utilise to increase a sense of social and emotional investment for audiences? I would argue that it is not whether the gallery entity is screen-based or robotic that is of key importance but the way it responds to and interacts with audience members. Real-time interactivity and responsiveness are key factors in achieving a compelling sense of social engagement and reciprocal agency. The ability of a gallery entity to respond and to 'answer back,' and its ability to command a response, are crucial here, evoking Mikhail Bakhtin's concept of dialogism, an "answerable engagement with a responsive other." [18: 68]

This "answerable engagement with a responsive other" is also a key component for Walter Benjamin in his discussion of auratic presence and agency. In "On Some Motifs in Baudelaire" Walter Benjamin explicitly identifies the reversibility of the gaze in the intersubjective seer/seen relationship (with its returned look) as a key feature of auratic presence: "...looking at someone carries the implicit expectation that our look will be returned by the object of our gaze. Where this expectation is met (which, in the case of the thought processes, can apply equally to the look of the eye of the mind and to a glance pure and simple), there is an experience of the aura to the fullest extent. ... The person we look at, or who feels he is being looked at, looks at us in turn. To perceive the aura of an object we look at means to invest it with the ability to look at us in return" [2 188].

With the various use of microphones and cameras, as well as touch and motion sensors, interactive gallery entities can now sense and 'look back' at audiences. Gallery objects that were

previously deaf, dumb and blind can now see and hear their human interlocutors. They know where they are in the gallery and what they are doing or looking at so they can respond with 'intelligence,' becoming responsive social partners in a shared physical and social space. This increasing awareness and agency on the part of the gallery entity also creates an emergent subjectivity and 'aura.'

As in the Turing Test, it is the perceived intelligence or awareness of the entity that is most important here. If the interactive entity's actions and responses are indistinguishable from those of a human or another living object then it becomes functionally human or alive and that is how audiences will treat it. Reeves and Nass argue that it is human nature to treat media entities in a social way and to treat them at 'face value' – if entities appear to be intelligent and to have emotions then we will treat them as if they do. However, it is us the audience who make them participants by our psychological and physiological reactions. As Reeves and Nass point out: "Social and natural responses come from people, not from media themselves." [25: 252]

Drawing on their own knowledge of physical movement and physiological responses as well as social, psychological and emotional states, audience members typically project complex life-like and human-like motivations on to their interactive gallery partners. It is these interpretive responses of audience members in reaction to the behaviour of the interactive gallery object that generates the emotional and social depth of the interactive encounter. Simon Penny comments, "...viewers (necessarily) interpret the behavior of the robot in terms of their own life experience. In order to understand it, they bring to it their experience of dogs, cats, babies and other mobile interacting entities. The machine is ascribed complexities which it does not possess. This observation emphasises the culturally situated nature of the interaction. The vast amount of what is construed to be the 'knowledge' of the robot' is in fact located in the cultural environment, is projected upon the robot by the viewer and is in no way contained in the robot." [22]

Of course, this tendency to treat screen images and robots as social partners means that we have a corresponding tendency to expect them to react in ways that are socially and naturally appropriate and believable. When they don't, and our expectations are not met, the result can be one of frustration, disappointment and annoyance. Using ultra-realistic life-like human representations in virtual and robotic entities does not necessarily provide more satisfying dialogic encounters for audiences, indeed they may even be counter-productive due to their tendency to create higher audience expectations and to trigger feelings of uncanniness. As discussed in a previous paper, the attempt to create very realistic human representations in robots and virtual characters such as Stelarc's interactive *Prosthetic Head* (2002) can fall into the trap of Masahiro Mori's uncanny valley, which postulates that as representations become almost (but not quite) human they generate feelings of disquiet and uncanniness. [5, 20] There seems to be an innate human tendency to become spooked by robots and digital animations that look not quite human; when the illusion fails or breaks down the effect can be distinctly eerie. Stelarc's *Prosthetic Head*, a 3D computer graphic animation based on scans of the artist's own head, was created to look as much like Stelarc as possible and audience members converse with the head by typing questions that the head answers using a computer generated voice. While some interesting conversations

occur between audience members and head, the combination of the non-human computer voice and the head's digital animation create an unintentionally eerie effect. The head's computer animated movements appear off-puttingly machine-like and inhuman; its smile and facial expressions are just a little 'off' and its eyes disconcertingly lifeless. Similarly, when the head's conversational responses are socially inappropriate or fall short of what audiences would expect of a typical human conversation, the illusion is broken. The more human-looking the entity is, the more we expect of it. When humanoid entities don't react appropriately or don't understand things we would normally expect humans to understand, the illusion is shattered, and this inevitably leads to disappointment and disaffection.

Creatively limiting audience expectation has proved to be a useful audience strategy here. Very simple systems and representations, both screen-based and robotic, can be surprisingly effective in engaging audiences in 'dialogic' encounters. The use of camera or motion-based sensors and surveillance systems to track and respond to audience behaviour and movement is one of the simplest and most effective means of creating art works that show an awareness of audience members and act as responsive social partners. Mirroring the gaze or movement of audience members is a simple and effective technique.

In his gallery installation *Bruce (The Watcher)* (2003), Sean Kerr uses giant black dots on a monitor to represent eyes that follow the movements of audience members as they enter and walk around the gallery. Bruce also talks back to his audience with a computer generated voice (Bruce is one of the Apple voice presets), using phrases from a database made up of comments written in the gallery visitor book, for example, "You are always interesting and a bit challenging" or "I don't understand your genre." Kerr's virtual entities are also frequently embodied physically in the gallery space, housed in cardboard boxes or other physical objects.

Kerr's artworks typically incorporate screen-based computer images with physically embodied gallery installation components. In *Klunk, Clomp, Aaugh! – Friends Reunited* (2008), the 'friends' are variously represented by cartoon-like dots on a computer screen representing eyes and other non-human objects that are triggered by motion sensors and automated computer programs. In one version of the work (exhibited as part of the *Mirror States* exhibition in 2008 at Campbelltown Arts Centre, Sydney, Australia), sensors trigger a giant inflatable plastic finger to inflate and deflate as audience members approach cartoon-like eyes displayed on two monitors positioned in a corner of the gallery. In another version of the work (exhibited at the Moving Image Centre in Auckland, NZ), the 'friends' include a robotic bucket that skitters across the floor and a life-size wooden finger that is raised and lowered (like its giant inflatable cousin) to give audiences the finger. Another friend, represented by two screen eyes that peer out of a darkened room through a partially closed door, sings a plaintive computer voiced version of the Beatles song 'Help.' The physicality of the gallery installation components along with the screen-based 'eyes' creates a distributed gallery personality that occupies the space of the gallery, however it is the 'eyes' and computer voice component that generates the strongest sense of a personality for audiences to engage with.

Also moving from the screen-based to the physical is Golan Levin's three-dimensional *Double-Taker (Snout)* (2008), a giant

googly eye on the end of a 2.5 metre robotic snout. Situated on the top of a building 'Snout' reacts to passers by, mirroring their movements and orienting its gaze to theirs as it rears up and moves from side to side as if to get a better view of them. The importance of the gaze and the returned look is of key importance here but the physicality and body language of Levin's robotic 'Snout' also plays a big role in triggering a visceral mirroring response in viewers. It is common to see audiences moving their heads and bodies from side to side in response to Snout's sideways movements creating an engaging dance-like interaction.

The importance of voice, gaze and body language is clear in all of these works. Providing gallery entities (either screen-based or physically embodied) with sensor based awareness and 'intelligent' computer programming so that they can demonstrate appropriate and engaging behaviours makes them active social participants that can initiate interactions and respond to audiences in socially engaging encounters.

3. CONCLUSION

All of the works described in this paper create highly engaging encounters across the mixed reality paradigm for their audience participants. All experiences, physical and screen-based (online and offline), incorporate sensual, emotional and intellectual components. Actions in the virtual have impacts in the real world and vice versa and agency, sensation and affect are distributed throughout the mixed reality sensorium. The ability of an artwork to respond to and 'dialogue' with its audience—to 'look back' and 'talk back'—is a key factor in making it an engaging and believable social partner. While incorporating more modalities and senses can help to make the audience experience perceptually richer and more tangible, it is clear that psychological, emotional and even physical engagement can be equally intense with both (virtual) screen-based and (physically embodied) robotic entities. When the screen image 'looks back' and responds to its audience, it attains a compelling sense of presence and agency. The multi-modality of an entity's sensory capabilities and its ability to act and respond is thus a more important indicator of presence and agency than whether it is a three dimensional object in the gallery or a screen image, or a mixed reality entity that incorporates both screen-based and physically embodied components.

4. REFERENCES

- [1] Asuma, R.T. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 355-385.
- [2] Benjamin, W. and Arendt, H. 1986. *Illuminations*. Schocken Books, New York.
- [3] Bolter, J.D. and Grusin, R.A. 1999. *Remediation: understanding new media*. MIT Press, Cambridge, Mass.
- [4] Burnett, R. 2004. *How images think*. MIT Press, Cambridge, Mass.
- [5] Cleland, K. 2004. Talk to Me: getting personal with interactive art. In *Interaction: systems, practice and theory*, R. Gibson and E. Edmonds Eds., UTS, Sydney.
- [6] Cleland, K. 2009. Prosthetic Identities and Virtual Cyborgs. *Second Nature*, in press.

- [7] Crabtree, A. and Rodden, T. 2008. Hybrid ecologies: understanding cooperative interaction in emerging physical-digital environments. *Personal Ubiquitous Computing* 12, 481-493.
- [8] Dourish, P. 2001. *Where the action is: the foundations of embodied interaction*. MIT Press, Cambridge, Mass.
- [9] Gallese, V. 2003. The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36, 171.
- [10] Haller, M., Billingham, M. and Thomas, B. 2007. *Emerging technologies of augmented reality: interfaces and design* Idea Group Pub., Hershey PA, xiii, 399.
- [11] Hansen, M.B.N. 2006. *Bodies in code: interfaces with new media*. Routledge, New York.
- [12] Hayles, N.K. 1996. Embodied virtuality, or how to put bodies back into the picture. In *Immersed in Technology: Art and Virtual Environments*, M.A. Moser and D. MacLeod Eds. MIT Press, Cambridge, MA, 1-28.
- [13] Holz, T., Dragone, M. and O'Hare, G.M.P. 2009. Where Robots and Virtual Agents Meet: A Survey of Social Interaction across Milgram's Reality-Virtuality Continuum. *International Journal of Social Robotics* 1, 83-89.
- [14] Huizinga, J. 1955. *Homo ludens: a study of the play-element in culture*. Beacon Press, Boston.
- [15] Ihde, D. 2002. *Bodies in technology*. University of Minnesota Press, Minneapolis.
- [16] Jabbi, M., Swart, M. and Keysers, C. 2006. Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage* 34, 1744-1753.
- [17] Kabisch, E. 2008. Datascape: a synthesis of digital and embodied worlds. *Space and Culture* 11.
- [18] McCarthy, J. and Wright, P. 2004. *Technology as experience*. MIT Press, Cambridge, Mass.
- [19] Milgram, P. and Kishino, F. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems* E77-D.
- [20] Mori, M. 1970. Bukimi no tani (The uncanny valley). *Energy* 7, 33-35.
- [21] Mouras, H., Stoléru, S., Moulier, V., Pélégriani-Issac, M., Rouxel, R., Grandjean, B., Glutron, D. and Bittoun, J. 2008. Activation of mirror-neuron system by erotic video clips predicts degree of induced erection: an fMRI study. *Neuroimage* 42, 1142-1150.
- [22] Penny, S. 1997. Embodied Cultural Agents: at the intersection of Art, Robotics and Cognitive Science. In *Proceedings of the AAAI Socially Intelligent Agents Symposium*, MIT, November 8-10 1997.
- [23] Ponseti, J., Bosinski, H.A., Wolff, S., Peller, M., Jansen, O., H.M., M., Büchel, C. and H.R., S. 2006. A functional endophenotype for sexual orientation in humans. *Neuroimage* 33, 825-833.
- [24] Ramachandran, V.S. 2001. Mirror neurons and imitation learning as the driving force behind "the great leap forward" in human evolution. In *Edge*. http://www.edge.org/3rd_culture/ramachandran/ramachandran_p1.html
- [25] Reeves, B. and Nass, C.I. 1996. *The media equation: how people treat computers, television, and new media like real people and places*. CSLI Publications, Cambridge University Press, Stanford, Calif., New York.
- [26] Rizzolatti, G. and Craighero, L. 2004. The Mirror-Neuron System. *Annual Review of Neuroscience* 27, 169-192.
- [27] Velonaki, M., Scheduling, S., Brown, I. and Rye, D. 2008. Physicality and Synthetic Reality. In *ISEA2008: 14th INTERNATIONAL SYMPOSIUM ON ELECTRONIC ART*, Singapore, 513-515.