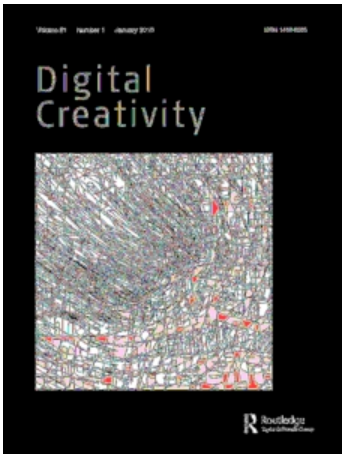

Digital Creativity
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t714576173

Mixed reality interaction: audience responses to robots and virtual characters

Kathy Cleland[a]
[a] University of Sydney, Australia

Online publication date: 26 May 2010

To cite this Article Cleland, Kathy(2010) 'Mixed reality interaction: audience responses to robots and virtual characters', Digital Creativity, 21: 1, 30 — 38
To link to this Article: DOI: 10.1080/14626261003654608
URL: http://dx.doi.org/10.1080/14626261003654608

Digital Creativity
2010, Vol. 21, No. 1, pp. 30–38

Mixed reality interaction: audience responses to robots and virtual characters

Kathy Cleland

University of Sydney, Australia

kathy.cleland@sydney.edu.au

Abstract

This article explores the way audiences respond to screen-based (virtual) and embodied (robotic) entities in the mixed reality terrain of the gallery space. While it would seem that physical three-dimensional objects in a gallery space, especially self-moving objects such as robots, have a distinct advantage in the reality stakes over screen images, the author suggests that there is no hard and fast distinction between how audiences respond to robotic entities and to screen-based virtual characters. It is the ability of an artwork to respond to and 'dialogue' with its audience—to 'look back' and 'talk back'—that is the key factor in making it an engaging and believable social partner. Artists discussed include Mari Velonaki, Stelarc, Ruairi Glynn, Karolina Sobecka and Golan Levin.

Keywords: mixed reality, phenomenology, robotics, avatars, virtual reality, virtual worlds, art, media art, new media art, audience interaction, audience response, mirror neurons

1 Introduction

Our experience of the world is increasingly becoming a mixed reality experience, a complex blend of the real and the digital. In digital culture we have seen a shift from a virtual reality paradigm to a mixed reality paradigm. The virtual reality paradigm of the 1990s was characterised by a split between virtual computer-generated worlds and the everyday physical world around us. Virtual reality existed as a world parallel to the real world; a world where new and fantastic terrains could be created and experienced, while the real world was left behind. In the twenty-first century, while virtual fantasy worlds still play a strong role in the popular imaginary, the separation between virtual environments and physical world has been significantly eroded. Mixed reality and augmented reality applications, where virtual images and information are merged or overlaid on the physical world and vice versa, are becoming growing technological, commercial and social trends (Milgram and Kishino 1994, Asuma 1997, Haller *et al.* 2007, Crabtree and Rodden 2008, Kabisch 2008). We are now living in a mixed reality paradigm where the real and the virtual, the natural and the artificial blend and intermingle in complex ways.

ISSN 1462-6268 © 2010 Taylor & Francis
DOI: 10.1080/14626261003654608
http://www.informaworld.com

Digital Creativity, Vol. 21, No. 1

In these new mixed reality environments, humans interact with screen-based virtual characters and computers interact with human audiences not just through the screen interface but 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 (Holz *et al.* 2009). Screen images can be housed in sculptural objects or projected on three-dimensional surfaces. With recent advances in 3-D image technologies, for example, James Cameron's *Avatar*, virtual screen images are likely to become increasingly life-like and three-dimensional. All of these techniques add to the embodied presence of screen-based or projected virtual entities and blur distinctions between 2-D screen reality and 3-D physical reality.

2 Mixed reality interactions

How can we theorise and think about this new mixed reality paradigm? What is the 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?

Mari Velonaki and her colleagues at the Social Robotics Unit at the University of Sydney argue that the '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' (Velonaki *et al.* 2008, pp. 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' (Penny 1997). Velonaki *et al.* also 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.

This embodied multi-modal interaction is exemplified in Velonaki's robotic installation *Fish-Bird: Circle C—Movement B* (a collaboration with David Rye, Steve Scheding and Stefan Williams), where two robotic wheelchairs interact with each other and with audience members by moving around the gallery space and communicating via text messages printed out on slips of paper (see Figure 1). The experience of interacting with the wheelchairs is a very physical one. In the gallery space, as one of the wheelchairs moves towards me, I move away—we dance back and forth. One of the robots wheels itself into a corner and another audience member follows it, wanting to check that it is okay. Visitors discuss the behaviour of the wheelchairs, speculating on their motivations and 'states of mind'. The text messages provide clues, but so does the 'body language' of the two robots as they move around the gallery. The experience is visceral and playful—physically and emotionally engaging.

How does this experience compare with a purely screen-based experience such as that of Stelarc's virtual *Prosthetic Head*? Based on scans of the artist, Stelarc's disembodied virtual alter ego is a three-dimensional computer graphic animation that engages gallery visitors in conversation (see Figure 2). Although the projected head is 'contained' on the two-dimensional gallery wall, the giant scale of the animated projection creates a strong sense of presence. In the intimacy of the darkened gallery space, audience members type questions on a computer keyboard and the head answers using a computer-generated voice. The conversational exchanges range from the intellectual to the playful and intimate.

Emergence: Art and Artificial Life, the exhibition accompanying the 2009 Digital Arts and Culture Conference at the Beall Center for Art + Technology, also provides some interesting and

Digital Creativity, Vol. 21, No. 1

Figure 1. Exhibition image of *Fish-Bird: Circle C—Movement B* by Mari Velonaki (with David Rye, Steve Scheding and Stefan Williams), 2005, from the *Mirror States* exhibition, Campbelltown Arts Centre, Sydney. © 2008 Kathy Cleland.

Figure 2. Exhibition image of *Prosthetic Head* by Stelarc, 2003, from the *Face to Face* exhibition at Hazelhurst Regional Gallery, Sydney. © 2009 Silversalt Photography. Reproduced with permission of Silversalt Photography and d/Lux/MediaArts.

instructive examples of audience interaction with robotic and screen-based entities. In *Performative Ecologies*, Ruairi Glynn's robotic entities swing themselves around to orient themselves to face the audience, capturing the gaze of curious audience members with their camera eyes before starting a quirky dance, with their robotic tails emitting dramatic flashing light displays (see Figure 3). Glynn has programmed the four robots to compete with each other to attract the audience's attention; however, audience members also vie with each other, calling to the robots to get them

Digital Creativity, Vol. 21, No. 1

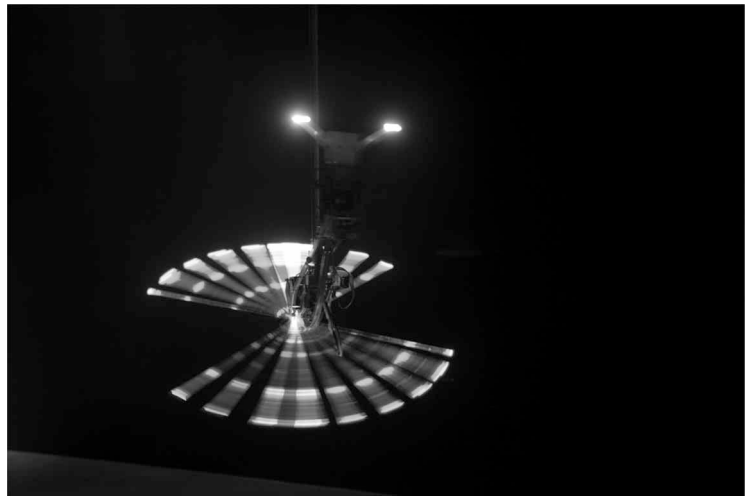

Figure 3. Exhibition image of *Performative Ecologies* by Ruairi Glynn, 2009, from the *Emergence-Art and Artificial Life* exhibition, Beall Center for Art + Technology, University of California Irvine. © 2010 David Familian and the Beall Center for Art + Technology. Reproduced with permission of David Familian and the Beall Center for Art + Technology.

to look at them and perform for them: 'Hello!' 'Over here!' 'Look at me!' Although the robots are fixed on metal supports, they can swing around 360 degrees and perform a range of kinetic movements, shaking, vibrating and dancing. The delight of the audience at the work was apparent at the exhibition, with audience members exclaiming and laughing in amusement, murmuring words of encouragement and moving around the robots trying to elicit their movement.

In contrast to Glynn's three-dimensional robots, Karolina Sobecka's *Sniff* features an abstract two-dimensional animated virtual dog (see Figure 4). The dog is projected on the gallery wall and its white silhouette stalks along the wall, responding as visitors enter the gallery space. As individuals walk from one end of the wall to the other, the virtual dog walks along with them, following their movements inquisitively and looking out at them from its screen space. Just as with Glynn's three-dimensional robots, audience members try to elicit responses from the dog by approaching it and making gestures. The dog's reactions and movement depend on the behaviour of the gallery participants. If you hold out an open palm the dog will sniff and tentatively approach you. If you raise your hand aggressively, it will growl. Even though neither Sobecka's virtual dog nor Glynn's robots can hear them, it was very common for visitors to talk to them, greeting them and encouraging them to interact.

As can be seen from the examples above, the distinction between audience responses to embodied physical entities and to 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), the opposition between the virtual (unreal) and the physical (real) is a shaky one in our new mixed reality paradigm.

In a phenomenological sense, Don Ihde argues that 'both RL and VR are part of the lifeworld and VR is thus both 'real' as a positive presence and a part of RL' (2002, p. 15). A phenomenological approach is also a key part of Paul Dourish's notion of 'embodied interaction' in human computer interaction (HCI) and his investigation of the way computational systems are embedded into environments and social contexts. As Dourish comments, 'Physically, our experiences cannot

Digital Creativity, Vol. 21, No. 1

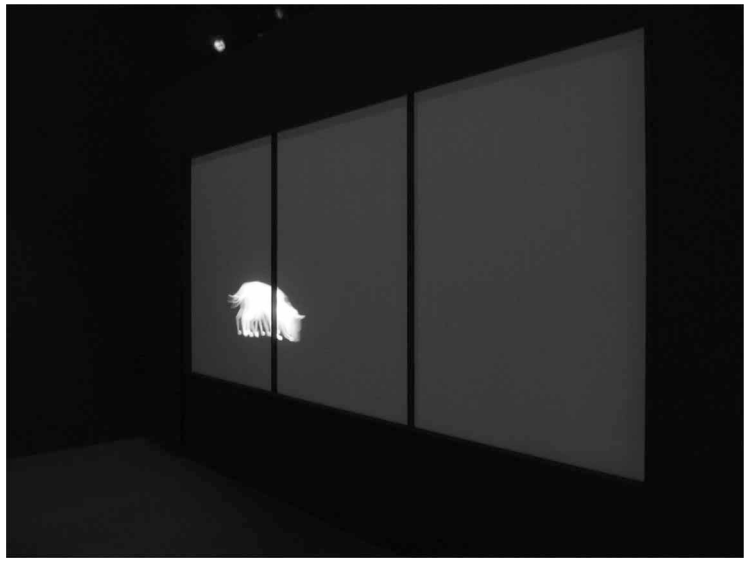

Figure 4. Exhibition image of *Sniff* by Karolina Sobecka with software developed by Jim George, 2009, from the *Emergence–Art and Artificial Life* exhibition, Beall Center for Art + Technology, University of California Irvine. ©2010 David Familian and the Beall Center for Art + Technology Reproduced with permission of David Familian and the Beall Center for Art + Technology.

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' (2001, p. 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' (2006, p. 3). Everything we experience—whether it's walking around in the world or playing a video game—is experienced by the human body and the human senses. 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 (1996, p. 1).

The perceptual experiences and affective responses generated by media images can feel just as real as those generated by the physical world. This is the essence of the mixed reality experiential paradigm—both physical entities and media images produce real experiences. 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' (1999, p. 19).

In *The Media Equation* (1996) Byron Reeves and Clifford Nass investigate this phenomenon, arguing 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. 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 the 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. 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 (Ramachandran 2001, Gallese 2003, Rizzolatti and Craighero 2004, Jabbi *et al.* 2006). The same mirror neuron responses and corresponding physical motor responses are also triggered by screen-based images, actions and emotions; pornography is a key example of this phenomenon (Ponseti *et al.* 2006, Mouras *et al.* 2008).

So, if we can respond just as strongly to screen-based entities as we do to physically embodied robots, what other factors are at play in creating art works that generate strong social and emotional responses in 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' (McCarthy and Wright 2004, p. 68).

This 'answerable engagement with a responsive other' is also a key component of Walter Benjamin's 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 (1986, p. 188).

With the 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 space 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'.

Art works that mirror the gaze and/or the physical movements of audience members also create very effective and engaging social interactions A good example of this is Golan Levin's

Digital Creativity, Vol. 21, No. 1

Figure 5. *Double-Taker (Snout)* by Golan Levin et al., 2008. ©2008 Golan Levin. Reproduced with permission of Golan Levin. Link: http://www.flong.com/projects/snout.

three-dimensional *Double-Taker (Snout)* (see Figure 5), a giant cartoon-like googly eye on the end of a 2.5 metre articulated robotic snout. Situated on the top of a building, *Snout* reacts to pedestrians 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 body language of Levin's robotic *Snout* also plays a big role in triggering a physical mirroring response in viewers. Gesture and body movement comprise a language that is more visceral and universal than spoken or written language. 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.

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 personality and emotions, then we will treat them as if they do. 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 living object, then it becomes functionally sentient or alive and that is how audiences will treat it. The audience also plays a key role in helping to make them participants through their own psychological projections and emotional responses. As Reeves and Nass point out: 'Social and natural responses come from people, not from media themselves' (1996, p. 252).

Drawing on their own knowledge of physical movement and physiological responses, as well as social, psychological and emotional states, audience members 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. Discussing his own influential early robotic artwork *Petit Mal*, 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 (Penny 1997).

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 appropriate and believable.

When they don't and our expectations are not met, the result can be one of frustration and annoyance. While some interesting conversations occur between audience members and Stelarc's *Prosthetic Head*, often the head's conversational responses are socially inappropriate or fall short of audience's expectations of a typical human conversation. 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.

A similar problem occurs to a lesser extent with Sobecka's *Sniff*. Even in its abstract form, audience members still have ingrained expectations about how a virtual dog should behave based on their own experience of interacting with real dogs. Because the dog's behaviour draws on a limited range of responses and has a tendency towards unhappy and angry barking (even when audiences are trying to be soothing or friendly) the result is that audiences can feel a bit put off and rejected. A few soft whimpers or happy barks in the dog's repertoire would help to build a stronger social bond with audiences.

In all cases, the more sensing technologies that an art work can incorporate (sight, sound, motion, etc.) and the more sophisticated its programming, the greater the range of dialogic responses that are possible. Stelarc has plans to incorporate vision-sensing technologies so that his *Prosthetic Head* can see and comment on what visitors are wearing and possibly recognise facial expressions. Similarly, if sound-sensing technologies were used in Sobecka's *Sniff*, the dog could be programmed to recognise tones of voice and pitch and respond accordingly. However, training and programming artificial entities to interpret what they see, hear and sense is a very complex task.

3 Conclusion

While it would seem that physical three-dimensional objects in a gallery space, especially self-moving objects such as robots, have a distinct advantage in the reality stakes over screen images, as we have seen there is no hard and fast distinction between how audiences respond to robotic entities and to screen-based virtual characters. Rather, it is the ability of an artwork to respond to and 'dialogue' with its audience—to 'look back' and 'talk back'—that is the key factor in making it an engaging and believable social partner. Utilising appropriate mirroring body language, gaze behaviour and voice or text-based interaction are all powerful strategies in creating socially engaging encounters. 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. The 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 reality entity that incorporates both screen-based and physically embodied components.

References

Asuma, R.T., 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6 (4), 355–385.

Benjamin W., 1986. On Some Motifs in Baudelaire. *In*: H. Arendt ed., H. Zohn trans. *Illuminations*. New York: Schocken Books.

Bolter, J.D. and Grusin, R.A., 1999. *Remediation: understanding new media*. Cambridge, MA: MIT Press.

Crabtree, A. and Rodden, T., 2008. Hybrid ecologies: understanding cooperative interaction in emerging physical-digital environments. *Personal Ubiquitous Computing*, 12, 481–493.

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

Gallese, V., 2003. The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology*, 36 (4), 171.

Haller, M., Billinghurst, M. and Thomas, B., eds., 2007. *Emerging technologies of augmented reality: interfaces and design*. Hershey, PA: Idea Group.

Digital Creativity, Vol. 21, No. 1

Hansen, M.B.N., 2006. *Bodies in code: interfaces with new media*. New York: Routledge.

Hayles, N.K., 1996. Embodied virtuality, or how to put bodies back into the picture. *In*: M.A. Moser and D. MacLeod, eds. *Immersed in technology: art and virtual environments*. Cambridge, MA: MIT Press, 1–28.

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 (1), 83–89.

Ihde, D., 2002. *Bodies in technology.* Minneapolis, MN: University of Minnesota Press.

Jabbi, M., Swart, M., and Keysers, C., 2006. Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage*, 34 (4), 1744–1753.

Kabisch, E., 2008. Datascape: a synthesis of digital and embodied worlds. *Space and Culture*, 11 (3), 222–238.

McCarthy, J. and Wright, P., 2004. *Technology as experience*. Cambridge, MA: MIT Press.

Mouras, H., *et al.*, 2008. Activation of mirror-neuron system by erotic video clips predicts degree of induced erection: an fMRI study. *NeuroImage*, 42 (3), 1142–1150.

Milgram, P. and Kishino, F., 1994. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems, E77-D* (12).

Penny, S., 1997. Embodied cultural agents: at the intersection of art, robotics and cognitive science. *AAAI socially intelligent agents symposium*. MIT [online]. Available from: http://ace.uci.edu/penny/texts/embodied.html

Ponseti, J., *et al.*, 2006. A functional endophenotype for sexual orientation in humans. *Neuroimage*, 33 (3), 825–833.

Ramachandran, V.S., 2001. Mirror neurons and imitation learning as the driving force behind 'the great leap forward' in human evolution [online]. *Edge*. Available from: http://www.edge.org/3rd_culture/ramachandran/ramachandran_p1.html [Accessed 5 November 2007].

Reeves, B. and Nass, C.I., 1996. *The media equation: how people treat computers, television, and new media like real people and places*. Stanford, CA: CSLI Publications.

Rizzolatti, G. and Craighero, L., 2004. The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.

Velonaki, M., *et al.*, 2008. Physicality and synthetic reality. *In Proceedings of the ISEA2008: 14th international symposium on electronic art*, Singapore, 513–515 [online]. Available from: http://www.isea2008singapore.org/ISEA2008

Kathy Cleland is a curator, writer and lecturer specialising in new media art and digital culture. She is Director of the Digital Cultures Program at The University of Sydney. Her curatorial projects include the *Cyber Cultures* exhibition series which toured to over twenty venues in Australia and New Zealand from 2000 to 2003; the *Mirror States* exhibition (2008) at MIC Toi Rerehiko, Auckland, NZ, and Campbelltown Arts Centre, Sydney; and *Face to Face: Portraiture in a Digital Age* for d/Lux/MediaArts, a digital portraiture touring exhibition in Australia (2008–2011).