(In press, 2015). In A. Georgopoulou & T. Spilloti (Eds.), *The Routledge handbook of language and digital communication.* London: Routledge.

NEW FRONTIERS IN INTERACTIVE MULTIMODAL COMMUNICATION

SUSAN C. HERRING

INTRODUCTION

This essay describes two emergent phenomena related to multimodality in digital communications. The first phenomenon is *interactive multimodal platforms*—Web 2.0 platforms that support a convergence of channels or "modes" (text, audio, video, images) for user-to-user communication. The second is *robot-mediated communication*—human-human communication in which at least one party is telepresent through voice, video, and motion in physical space via a remotely controlled robot. At first blush, these two phenomena may appear unrelated—websites are on the Internet, whereas robots are physical, mechanical objects; web interaction is persistent and often asynchronous, whereas robot-mediated interaction takes place in real time and does not leave a verbal trace; and so forth. At the same time, both technologies mediate human-to-human communication, support social as well as task-related interaction, and involve multiple modes or channels. More generally, both can be situated under the broad rubric of multimodal computer-mediated communication (CMC), as represented in Figure 1.

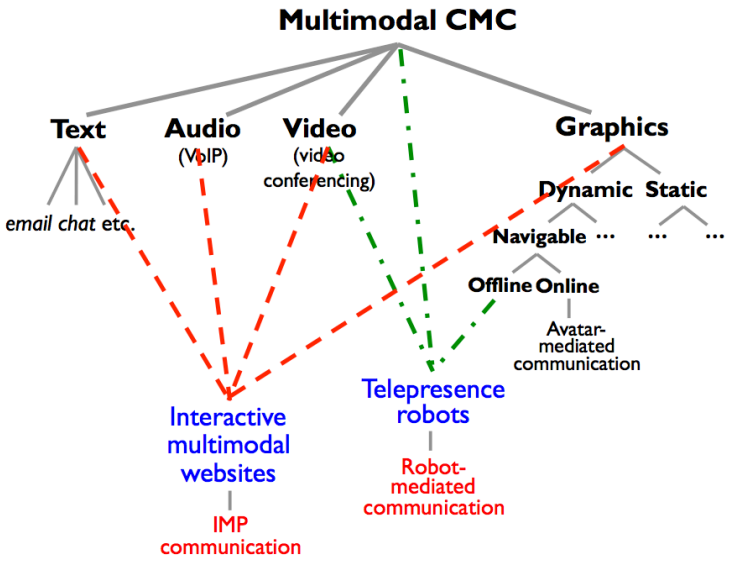

Figure 1. Multimodal computer-mediated communication

In what follows, I discuss each phenomenon in turn. I also identify research opportunities and challenges raised by the two phenomena for scholars of multimodal discourse, and conclude by considering their future outlook.

INTERACTIVE MULTIMODAL PLATFORMS

Interactive multimodal platforms (henceforth, IMPs) allow social media users to comment on multimodal content via multiple channels on a single website—and even within a single thread or conversation. An IMP minimally involves text plus one other mode (audio, video, and/or graphics); the modes may be synchronous or asynchronous. While IMPs are Web 2.0 sites, in that they are web-based platforms that incorporate user-generated content and social interaction, not all Web 2.0 sites are IMPs: Sites on which messages are mainly textual (excluding multimedia attachments), such as Wikipedia and Twitter, are not IMPs in their current form. One of the first IMPs was YouTube, which allowed users to comment on a shared video asynchronously, via either text or video. Facebook became an IMP when it added video chat to its suite of textual communication options. Another example is the multiplayer online game World of Warcraft, which for several years incorporated synchronous audio chat (Voice-over-IP) in addition to text chat. The messaging service WhatsApp is arguably an example of an IMP on a mobile device: In addition to text messaging, it enables smart phone users to exchange images, video, and audio media messages in a single "conversation."

Multimodal commenting environments raise theoretical and practical questions about why and how people communicate in a given mode. To what extent does the choice of text, audio, video, and/or images affect the nature of users' communication? Which is most efficient, most positive in tone, most social? What impressions do messages in each mode make on their recipients? Does communication in one mode influence communication in other modes? From a practical perspective, can knowledge of mode differences be leveraged to engineer more pro-social outcomes through the design of multimodal web environments?

IMP research is starting to be conducted, and its early findings suggest that mode choice makes a communicative difference. In her study of World of Warcraft, Newon (2011) found that voice chat was dominated by a few individuals, whereas text chat favored more democratic participation. Sindoni (2014) researched a one-on-one communication environment similar to Skype, observing that interlocutors were more self-conscious in video chat than in written exchanges. Moreover, seeing themselves in the feedback image "produce[d] psychological effects influencing the verbal and nonverbal features of the online exchange" (p. 333). Relatedly, in discussions on Voicethread.com, a website that supports asynchronous commenting in text, audio, and video, Herring and Demarest (2011) found that audio and video comments were more self-conscious and ego-focused than text comments, as well as being more positive in tone. As regards tone or sentiment, Bourlai and Herring (2014) found that emotions expressed in animated GIFs in Tumblr posts were more positive than emotions expressed in text comments. Text, in addition to being more negative, was more sarcastic.

Although the latter two studies are based on limited data and their results should therefore be considered preliminary, the finding that textual communication was more negative than communication in non-textual modes in both Voicethread and Tumblr is intriguing. Following early CMC theorists such as Daft and Lengel (1984), Bourlai and Herring (2014) propose that the relative paucity of paralinguistic and social cues in textual CMC

creates a distancing effect between interlocutors; it also lends itself to ambiguity, a prerequisite for sarcasm. Thus IMP research can help evaluate claims of technological determinism, which in its strong form holds that features of a technological medium determine user behavior through that medium.

ROBOT-MEDIATED COMMUNICATION

Telepresence robotics is a sophisticated form of robotic remote control in which a human operator has a sense of being on location. Telepresence robots are distinct from autonomous robots, which depend on pre-programmed artificial intelligence, in that the behavior of telepresence robots is controlled in real time by human operators. They are mainly used to facilitate geographically-distributed communication, e.g., for teleworkers, academics, and medical professionals, and in security/high-risk operations. Because of the embodiment and enhanced control they offer, particularly in terms of mobility, telepresence robots provide a richer sense of "being there" than online videoconferencing technologies such as Skype (e.g., Rae, Mutlu, & Takayama 2014).[1]

Telepresence robots are giving rise to a new form of CMC: robot-mediated communication (henceforth, RMC), human-human communication mediated by one or more telepresence robots. How to classify RMC in relation to CMC is not yet clear. On one hand, RMC is a type of videoconferencing supplemented by movement. On the other hand, it can be considered a type of avatar-mediated communication (like graphical avatars in virtual worlds such as Second Life) in which the user's avatar is a robot that moves around in physical space. The telepresence robot could also be considered a mediating technology in and of itself—a mode, on a par with text, audio, and video. These conceptual relationships are represented schematically in Figure 1 above.

RMC constitutes a potentially rich domain of analysis for scholars of discourse and social interaction. How is interaction management accomplished when human-human communication is mediated by a robot avatar? How does the limited mobility and range of visibility of persons piloting robots affect their ability to attract attention, gain and hold the conversational floor, and time turn-taking appropriately? What is the social and hierarchical status of robot-mediated communicators: Are they taken seriously in positions of leadership? Do they receive politeness and deference the same as if they were physically present, and to what extent does this vary by gender—theirs and that of their interlocutors? How do others refer to the person-in-the-robot (as 'you,' 's/he,' 'it?') and what social and/or technical factors condition variation in reference?

RMC has only been investigated by one research group so far, to the best of my knowledge, and its studies are based on laboratory experiments. One study, for example, found that subjects rated "leaders" interacting through taller robots as more persuasive than "leaders" interacting through shorter robots (Rae, Takayama & Mutlu 2013a). Another found that persons communicating through a telepresence robot were trusted more than those communicating through a tablet computer (Rae et al. 2013b). While such experimental findings are valuable, they do not necessarily transfer into naturalistic contexts of RMC use.

FUTURE OUTLOOK

As the trend towards media convergence continues, more IMPs will emerge. Along with the new communicative possibilities that they open up, IMPs are a rich source of data for multimodal discourse analysis. However, most discourse analysis methods (including the approach to computer-mediated discourse analysis put forth in Herring 2004) were devised for spoken or written/typed language, but not for nonverbal communication in video, graphics, music, etc. The challenges are compounded by the practical reality that different modes or channels of communication often co-occur (and co-construct meaning) on the same platform, in the same interaction, and even in the same message. An approach needs to be developed that analyzes disparate modes in relation to one another, ideally with a common set of research questions, methods, and so forth, to permit meaningful comparisons across modes and across platforms. In-depth studies of individual IMPs are also needed. The findings of such research raise the practical possibility of engineering web platforms to optimize the use of modes that produce specific outcomes—for example, platforms with audio and video commenting to reduce negativity in political forums, which tend to be contentious and polarized.

As for RMC, it is still in its infancy—telepresence robots are just starting to be commercially produced and employed. RMC will probably never completely supplant videoconferencing, however. For one thing, it would hardly make sense for *all* participants in remote gatherings to use robots; if there is no reason to be in a particular physical location, virtual presence is simpler and more economical. But telepresence robots are already coming into greater use for remote participation in classrooms and conferences, instruction, and collaborative physical tasks (see, e.g., Herring 2013), and their use produces RMC.

Researching RMC is arguably even more challenging than researching IMPs. There are ethical issues associated with collecting data from naturally-occurring robot-mediated interactions, which could be misconstrued as mobile surveillance. Unlike web communication, RMC is not self-archiving; the researcher needs to devise methods for recording, transcribing, and presenting information not normally found in CMC, such as movement, proximity, and gaze direction. To address these challenges, RMC analysts might usefully borrow from the ethnographer's methodological toolkit. And as with IMP research, RMC research has practical implications: It can inform the design of robots that better support natural interaction and communication.

Finally, these emergent technologies have the potential to alter the way we understand multimodal CMC. Multimodality in CMC is not new, of course. But IMPs are new in that they enable the use of multiple channels in the same conversation. And RMC is new in that it extends video avatars offline, enabling physical movement and navigation (and potentially, gestures) as part of mediated interaction. They both enrich multimodal CMC, according to Information Richness Theory as laid out by Daft and Lengel (1984): They add channels of communication. But Daft and Lengel assume that the more channels, the more the communication will resemble face-to-face communication (the richest, ideal mode). In contrast, both IMP-mediated and robot-mediated communication give rise to forms of communication that are distinct from face-to-face communication, with their own unique

affordances, which are both less than and more than face-to-face interactions. In this sense, IMPs and RMC provide fertile ground for theorizing about language in digital interaction, as well as for empirical investigation.

ENDNOTE

1. Examples of telepresence robots that are currently commercially available in the U.S. include the Beam (Suitable Technologies), the VGo (VGo Robotics), and the Double (Double Robotics).

REFERENCES

Bourlai, E & Herring, SC 2014, 'Multimodal communication on Tumblr: "I have so many feels!"', *Proceedings of WebSci 2014*, ACM, New York, available at: http://www.sheridanprinting.com/14-websci4chRV610jmp/docs/p171.pdf

Daft, R & Lengel, RH 1984, 'Information richness: a new approach to managerial behavior and organizational design', in BM Staw & LL Cummings (eds.), *Research in organizational behavior*, JAI Press, Greenwich, CT,, pp. 191-233.

Herring, SC 2004, 'Computer-mediated discourse analysis: an approach to researching online behavior', in SA Barab, R Kling & JH Gray (eds,), *Designing for virtual communities in the service of learning*, Cambridge University Press, New York, pp. 338-376.

Herring, SC 2013, 'Telepresence robots for academics', *Proceedings of ASIST 2013,* November 1-6, 2013, Montreal, Canada, prepublication version available at: http://ella.slis.indiana.edu/~herring/AM13.telepresence.robots.pdf

Herring, SC & Demarest, B 2011, 'Mode choice in multimodal comment threads: effects on participation and language use', Unpublished manuscript.

Newon, L 2011, 'Multimodal creativity and identities of expertise in the digital ecology of a World of Warcraft guild', in C Thurlow & K Mroczek (eds.), *Digital discourse: language in the new media,* Oxford University Press, pp. 203-231.

Rae, I, Takayama, L & Mutlu, B 2013a, 'The influence of height in robot-mediated communication', *Proceedings of the 8th ACM/IEEE international conference on human-robot interaction (HRI '13),* IEEE Press, Piscataway, NJ, pp. 1–8.

Rae, I, Takayama, L & Mutlu, B. 2013b, 'In-body experiences: embodiment, control, and trust in robot-mediated communication', *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '13)*, ACM, New York, pp. 1921–1930.

Rae, I, Takayama, L & Mutlu, B. 2014, 'Bodies in motion: mobility, presence, and task awareness in telepresence', *Proceedings of CHI 2014,* ACM, New York, pp. 2153-2162.

Sindoni, MG 2014, 'Through the looking glass: a social semiotic and linguistic perspective on the study of video chats', *Text & Talk,* vol. 34, no. 3, pp. 325-347.