Multimodal approaches to video analysis of digital learning environments

Carey Jewitt
Institute of Education
London Knowledge Lab
c.jewitt@ioe.ac.uk

Sara Price
Institute of Education
London Knowledge Lab
s.price@ioe.ac.uk

Abstract. Contemporary digital learning environments, such as tangibles, mobile and sensor-based technologies are inherently multimodal, both in terms of representation modalities and interaction modes. Drawing on social semiotics, multimodality offers a valuable approach for analysing video data, as it systematically attends to the interpretation of a wide range of communicational forms (e.g. gaze, posture, action, speech) used for making meaning. This paper explores a multimodal approach to video data in the context of ubiquitous technologies for learning.

Keywords. Multimodality, embodiment, digital learning environments.

1. INTRODUCTION

Our research experience is in the context of digital learning environments. Our current work seeks to understand the role of 'embodiment' in digital interaction for learning and meaning making, with a particular focus on the use of ubiquitous technologies including mobile, sensor and tangible based computing.

New technologies such as these also generate new kinds of research questions and data to be collected and analysed. In particular, such digital learning environments are inherently multimodal in terms of both representation modalities and interaction modes. To gain insight into embodied forms of interaction it is therefore important to take these multimodal aspects of interaction into consideration. Thus, our research also explores multimodal qualitative research methods, to develop and evaluate multimodal research approaches for studying contemporary digital data and environments.

Multimodality is an inter-disciplinary research methodology that has developed over the past decade in relation to the collection, analysis of visual data, video based data, and naturally occurring digital data (e.g. CCTV, visual digital displays, online logs generated through games etc.) (Jewitt, 2009; Kress, 2009). It attends systematically to the social interpretation of a wide range of communicational forms that are used for making meaning. As such it offers an appropriate approach for examining interaction where meaning making involves a variety of modes of interaction

including physical action, gaze gesture, as well as language as communication forms.

Video data is a central part of this work, as it provides a permanent and re-accessible record of complex interactions (both with digital technologies, the physical world and other individuals) over time, enabling repeated viewing and simultaneous viewing and discussion within project teams. It can provide a rich data set for in depth analysis of interaction and meaning making. This paper begins with an overview of multimodal approaches to research, and an outline of key steps in the video analysis process from a multimodal perspective. It then presents the context of our research highlighting a multimodal approach to researching embodiment and digital technologies for learning through video data. Finally, some key research opportunities and challenges are discussed.

2. MULTIMODALITY

Multimodality draws on social semiotic theories of communication, but extends the interpretation of language and its meanings to the whole range of modes of representation and communication employed in a culture (Kress, 2009; van Leeuwen, 2005). Thus, it includes methods for analysing visual, aural, and embodied communication (including gaze, gesture, body posture and position); and attends to spatial modes of communication, and relationships between them.

A multimodal approach is underpinned by three theoretical assumptions. First, it assumes that representation and communication always draw on multiple modes, all of which contribute to meaning.

Its focus of analysis and description is on the full repertoire of meaning-making resources, which are used in different contexts (e.g. action, visual, spoken, gestural, written, three-dimensional etc), and on developing ways that show how these are organized to make meaning. Second, multimodality assumes that all forms of communication (modes) have been shaped through their social, cultural, and historical usage and realize communicative work in distinct ways. All communicational acts are considered to be socially made and meaningful within the social environments in which they have been made. Different modes shape meaning in mode-specific ways, so that meanings are in turn differently realized in different modes. For instance, the spatial extent of a gesture, the range of voice intonation, and the direction and length of a gaze are all part of the resources for making meaning. Third, interaction produces meaning. People orchestrate meaning through their selection and configuration of modes. Multimodality focuses on people's process of meaning making, a process in which people make choices from a network of alternatives: selecting one modal resource (meaning potential) over another (Halliday, 1978). In the context of exploring and analyzing concepts of embodiment and the role of the body in contributing to the meaning making process (a core part of learning), this approach offers the potential to capture complex interaction with digital technologies grounded in physical activity, yet socially mediated through new forms of communication and collaboration, and mediated through new tools. The focus on multiple modes from body posture to gaze to physical action and manipulation as forms of communication is critical in fully considering meaning making in ubiquitous learning environments

3. COLLECTING AND ANALYSING MULTIMODAL DATA

Here we outline some key steps taken in a multimodal approach to research, where systematic attention is paid to meaning and the ways in which people use modes to represent the world and engage in social interaction.

3.1 Collecting and logging data

If the focus is on face-to-face interaction, for instance, in a classroom, data are likely to include a mixture of video recordings, field-notes, materials and texts used during the interaction, participant interviews, and possibly policy documents and other texts related to what has been observed. Video recording of a lesson is viewed along with the field-notes and texts collected from the lesson. A video log or descriptive synopsis is made from observations, and may include sketches of events, video stills, a map of the situated context and trails

of movement, and comments on the teacher and student movement. If the interaction is more confined to examining tabletop interaction in a lab setting, then video recording forms a significant part of the data collection along with post interaction interviews or prompted recall sessions. Video recording often involves multiple camera views, e.g. a bird's eye view of the tabletop as well as a wide-angle view to capture body positioning and movement around the table. With mobile technology interaction, video recording requires a mobile researcher with a camcorder following the action and interaction across space and time. In addition, computer-logged data may be collected and logged for analysis.

3.2 Viewing data

Multimodal analysis involves repeated viewing of the data. We watch video data as a research team to get a range of perspectives. We view video data with both sound and image, and sometimes without sound to focus on action or body posture or gaze. We view with sound only, fast forward, in slow motion – all of which provide different ways of seeing the data. This helps recognize customary acts, patterns of gesture, for example, and routines across the time and space of the interaction. Viewing the data alongside the logs and organizing it in light of the research questions serves to generate criteria for sampling the data, refining and generating new questions, and developing analytical ideas. Computer logs for contexts involving mobile technologies are of particular importance in this process, but raise challenges of combining video and logged data for analysis.

3.3 Sampling data

Using video to collect data inevitably produces rich data, often requiring the syncing of multiple camera views, and with computer-logged data. With a focus on all the modes in play, multimodal transcription and analysis are intensive. Where it is not feasible to analyze all the video data in detail, we sample the data selecting instances (episodes) for detailed analysis. How to select these episodes is a difficult question and one that is intimately guided by the research question. One approach is to focus on those moments in the interaction where the interaction order is disturbed or where a convention is broken, as it is on those occasions where interaction phenomena are made manifest. We also look at sequences of ongoing meaning drawing on notions of turn taking and mapping of the appearance and reappearance of concepts across the full period of interaction.

3.4 Transcribing and analyzing data

In linguistic traditions of transcription conventions are used to express features of speech, such as

intonation, hesitations or pauses, which are not normally expressed in writing. But even if one adopts the most sophisticated set of conventions, the transcriber has to accept there are details that are lost. From a multimodal perspective, the instance of communication is not limited to speech or writing. The 'transductions' (that is the move from one mode to other modes) involved in a multimodal transcript are therefore more varied, requiring us to address gains and losses when moving from gesture, gaze, posture and other embodied modes of communication to image, writing, layout, colour and other graphic modes available in print. Multimodal interaction when working with video can be transcribed in a number of ways. Conventionally time is used to anchor the interaction, rather than talk. Still images, sketches of gestures or maps of spatial arrangements may be included in a transcript. The focus is on showing all modes of communication and how these are interconnected across an interaction (Bezemer, 2012).

4. RESEARCHING EMBODIMENT: A MULTIMODAL APPROACH

Emergent technologies offer a different perspective and approach to interaction from traditional desktop computing. Their very nature offers opportunities for exploiting a wider range of perceptual experiences, and foster more bodily-based physical interaction in new ways. Our research focuses on three kinds of bodily-based interactions to explore concepts of embodiment in digital environments: tangible technologies enable physical objects to be enhanced in new ways through linking to various forms of digital augmentation; mobile technologies can be used to enhance contextually based experience in real world environments fostering new forms of data collection and new ways of thinking about e.g. science; and sensor technologies can be designed to foster kinaesthetic experience through whole body movements.

4.1 Physical manipulation

Tangible technologies, in the form of physical objects embedded with computational power linked to various forms of digital representations, offer new opportunities for learning through hands-on physical manipulation and exploration. These technologies are of interest for embodiment in relation to: how the handling of objects and physical touch makes explicit relevant physical properties of objects and how this might facilitate meaning making (Price et al., 2009; Pontual & Price 2009); how manipulation or gestural interaction with artefacts shape communication and meaning in interaction; how action on objects shapes interpretation of learning concepts. Our work explores these questions through a tangible

tabletop environment designed to support the learning of the physics of light. Children use different coloured (tangible) blocks to explore reflection, refraction and absorption, where digital representations of light behaviour are dynamically displayed on the tabletop surface according to object location, colour and texture. Video data includes a camera view of the whole scene (tabletop and full view of participants), and a focused birds-eye view of hands-on action. A multimodal approach to analysis attends not only to language and movement of objects during interaction, but also to gesture, body posture and gaze; their role in meaning making and communication; and explores moments of transduction. This approach foregrounds the relationship between bodily forms of interaction and meaning making.

4.2 Location, context

Mobile technologies (including GPS and GIS tools) offer opportunities to enhance aspects of the environment, and promote contextual learning e.g. in fieldwork, combining physical experience of the environment with scientific ideas and geospatial concepts. These digital technologies are of interest in the context of the body as they exploit our physical space and perceptual interaction with the environment, and may enhance the physical experience of a space through making contextually relevant information available in-situ (e.g. Rogers and Price, 2008). Applications such as Layar, ARCGis, Google Earth, and learning prototypes such as the mobile application GeoSciTeach offer the potential to change perception and awareness of environmental context. Here embodiment extends the reach of the individual to being part of a larger environment, making individuals context aware through information presentation and digital overlay. Video data in this context is central to accessing interaction patterns, and the role of technology in mediating experience. A multimodal approach focuses on aspects of bodily interaction with the environment and the technology. In particular it looks to examine how the technology determines body orientation, posture or gaze; and how information is communicated between participants, focusing for example, on material sharing of smartphone representations or through mutual gaze at aspects in the environment.

4.3 Whole body/ kinaesthetic interaction

Sensor technologies, wii motes and systems such as Xbox Kinect exploit whole-body interaction and offer the opportunity for exploring whether and how digital technologies can promote kinaesthetic awareness (e.g. Price and Rogers 2004; Sheridan et al., 2009). A multimodal analysis requires video capture of participants' whole body movements, as

well as the digital representation to which this is linked. Again combining video views is needed for analysis. Similar to the approach taken with tangible technologies a multimodal approach attends to bodily movement during interaction, as well as gesture, bodily orientation and gaze, their relationship to the technology and their role in meaning making.

Across all of these activities various strategies are used during video analysis. This may include slowing down video to facilitate tracking of links between e.g. gaze and action; viewing without sound; repeated viewing of critical events.

4. OPPORTUNITIES AND CHALLENGES

One of the strengths of multimodality is its eclectic nature, which offers a beneficial approach to researching complex digital environments that are multimodal: both in terms of representation and interaction. While linguistic theories provided the starting-point for this social semiotic approach, others have expanded this frame of reference to draw on other approaches (e.g. film theory, musicology, game theory, socio-cultural theories). Nevertheless, multimodality as a research method is still at an early stage of development, with much yet to be established, both in terms of theory and in terms of practices of transcription, language of description and analysis, including video data.

One fundamental challenge is identifying the scope and scale of analysis. Too much attention to many different modes may take away from understanding the workings of a particular mode; while too much attention to a single mode runs the risk of 'tying things down' to just one of many ways in which people make meaning. In the context of ubiquitous technologies the need to attend to several modes of interaction including gaze, gesture, body posture and movement as well as speech and action are of central importance. Managing the analysis of such different modes and exploring the transductions between them, suggest the need to undertake microanalysis of sampled sections of video data. This allows detailed analysis of each mode of interaction, its role in communication and learning activity, and how the different modes play out across time in the meaning making process. Once such microanalysis has taken place, patterns of similar kinds of interaction across different episodes of interaction can be explored.

A further challenge when researching learning is accessing data that provides insight into participant reasoning and reflection. In environments where physical action and manipulation are central to the activity talk-aloud protocols would change the nature of interaction, drawing participants'

awareness to talk, and foreground verbal expression. Post-hoc protocols for accessing reflection offer an alternative that enables a more natural real-time interaction in activity. In this work, for example, we are exploring prompted recall, through participant viewing of their own video data; and post activity explanations or demonstrations using the technologies themselves. This might for example, involve children showing what they did on the tangible tabletop and explaining their interpretation of light behaviour.

5. REFERENCES

- Bezemer, J. (2012) How to transcribe multimodal interaction, access: Mode.ioe.ac.uk
- Halliday, M. (1978). *Language as a Social Semiotic*. London: Edward Arnold.
- Kress, G. (2009). *Multimodality: A Social Semiotic Approach to Communication*. London: Routledge Falmer.
- Jewitt, C. (2009) *Routledge Handbook of Multimodal Analysis*, London: Routledge.
- Price, S. & Pontual Falcão, T. (2009) Designing for physical-digital correspondence in tangible learning environments. *Proc 8th Int Conference on Interaction Design and Children*. Italy p. 194-197
- Price, S., Roussos, G., Pontual Falcao, T., & Sheridan, J. G. (2009). Technology and embodiment: relationships and implications for knowledge, creativity and communication. *Beyond Current Horizons: technology, children, schools and families*.
- Price, S. & Rogers, Y. (2004) Let's get physical: the learning benefits of interacting in digitally augmented physical spaces. In (eds) J.D.M. Underwood and J. Gardner, *Computers and Education: Special issue: 21st Century Learning*: 43, 137-151
- Rogers, Y. & Price, S. (2008) The Role of Mobile Devices in Facilitating Collaborative Inquiry in Situ. *Research and Practice in Technology Enhanced Learning*. Asia-Pacific Society for Computers in Education. 3 (3), 209-229
- Sheridan, J.G., Price, S. & Pontual-Falcão, T. (2009) Using Wii Remotes as Tangible Exertion Interfaces for Exploring Action-Representation Relationships. In *Proc. Workshop on Whole Body Interaction, CHI '09*, 4 April, Boston, USA.
- van Leeuwen, T. (2005). *Introducing Social Semiotics*. London: Routledge.