International Society of the Learning Sciences

MultiMod: A platform for qualitative analysis of multimodal learning analytics

Chris Proctor, David Mawer
chrisp@buffalo.edu, dmawer@buffalo.edu
University at Buffalo (SUNY)

Abstract: Multimodal analytics are increasingly important for research on learning, equity, and justice in our digitally-mediated world. However, challenges to using multimodal analytics in research include selection, analysis, technology, and ethics. We present MultiMod, a platform for aggregating, selecting, analyzing, and presenting multimodal texts. A case study of multimodal sensemaking, intersubjectivity, and collaboration in Minecraft will illustrate how MultiMod supports our research team's qualitative analysis and allows annotated simulation as a mode of sharing research findings.

Introduction

The promise of multimodal learning analytics (Blikstein & Worsley, 2016) to inform research on formal and informal learning continues to grow. Our daily activity is increasingly mediated by overlapping digital interfaces (e.g. social media, web platforms for learning and work, cell phones, videoconferencing, access cards, security cameras) which produce high-granularity logs of our behavior. These logs are used extensively for advertising and surveillance, occasionally for "quantified self" introspection, and rarely for public-interest research. Because multimodal analytics are so instrumental in our interactions with governments, corporations, and social worlds, research on and with multimodal analytics can help us understand the relationship between learning, equity, and social justice.

There are multiple challenges to the effective use of multimodal analytics in research on learning. The challenges to using video in education research identified by Derry et al. (2010)–*selection, analysis, technology, and ethics*–are even more salient for multimodal analytics. Consider, for example, remote learning through a videoconferencing platform such as Zoom. The problem of *selection* is deepened when there are multiple streams of analytics to consider (e.g. each user's video, audio, and chat log) , when some streams record synthetic activity which would not occur without the interface (e.g. raising hands, sharing screens), and when some streams are downstream analyses of other streams (e.g. estimations of gaze, focus, or facial expression based on video). The problem of *analysis* is particularly acute for multimodal analytics. A principled, theoretically-grounded approach will likely require human qualitative analysis. Which representations of multifaceted data are suitable for the researcher's interpretive eye, and how, during qualitative analysis, do we distinguish data from its representation? The problem of *technology* is also substantial: although we use digital platforms every day, the logs they generate are seldom made available to end-users. Aggregating, processing, presenting, and analyzing log data requires specialized tooling. Finally, the problem of *ethics* pervades any work with multimodal analytics. Although we routinely consent to invasions of privacy and abuse of our personal data when we use social media or credit cards, or when we enter a school or a coffee shop, truly informed consent for research is difficult because we generally have so little awareness of the significance and uses of our personal data (Zuboff, 2019). Positionality also presents an increased ethical challenge when the skills to collect, analyze, and represent multimodal data are so unevenly distributed.

In this demo we present MultiMod, a platform for aggregating, selecting, analyzing, and presenting multimodal analytics, which addresses some of these challenges. MultiMod offers an interface similar to qualitative analysis tools such as NVivo or Atlas.TI, however the texts which can be analyzed go beyond audio and video. MultiMod allows multiple temporal streams such as audio and video to be synthesized with spatial data and other signals into an interactive simulation. This novel form of representing data provides a platform for observers to engage with research data as if they were present in the activity itself. Eisner (1997) highlights the promises of providing conditions for capturing *authenticity* in the use of alternate forms of data representations; we have found that the present tool indeed captures this authenticity, occasioning a phenomenological sense of "being there."

This demo will present a case study of how MultiMod has supported qualitative analysis of collaboration in Minecraft (Proctor & Muller, 2022). The demo will demonstrate how MultiMod is currently being used within our research team, and will showcase our initial findings of how players recruit multiple modalities of communication to achieve intersubjectivity and collaboration (Meier, Spada, & Rummel, 2007). This paper offers an overview of MultiMod's design goals and the affordances developed to meet them, as well

as a technical description of the system. We close with a discussion on implications for future technology integration including the use of virtual and augmented reality in qualitative research. We envision MultiMod being used in future research involving multimodal sensemaking, embodiment and placemaking, from the scale of dyadic collaboration to civic life and activism at the scale of a city.

Design overview

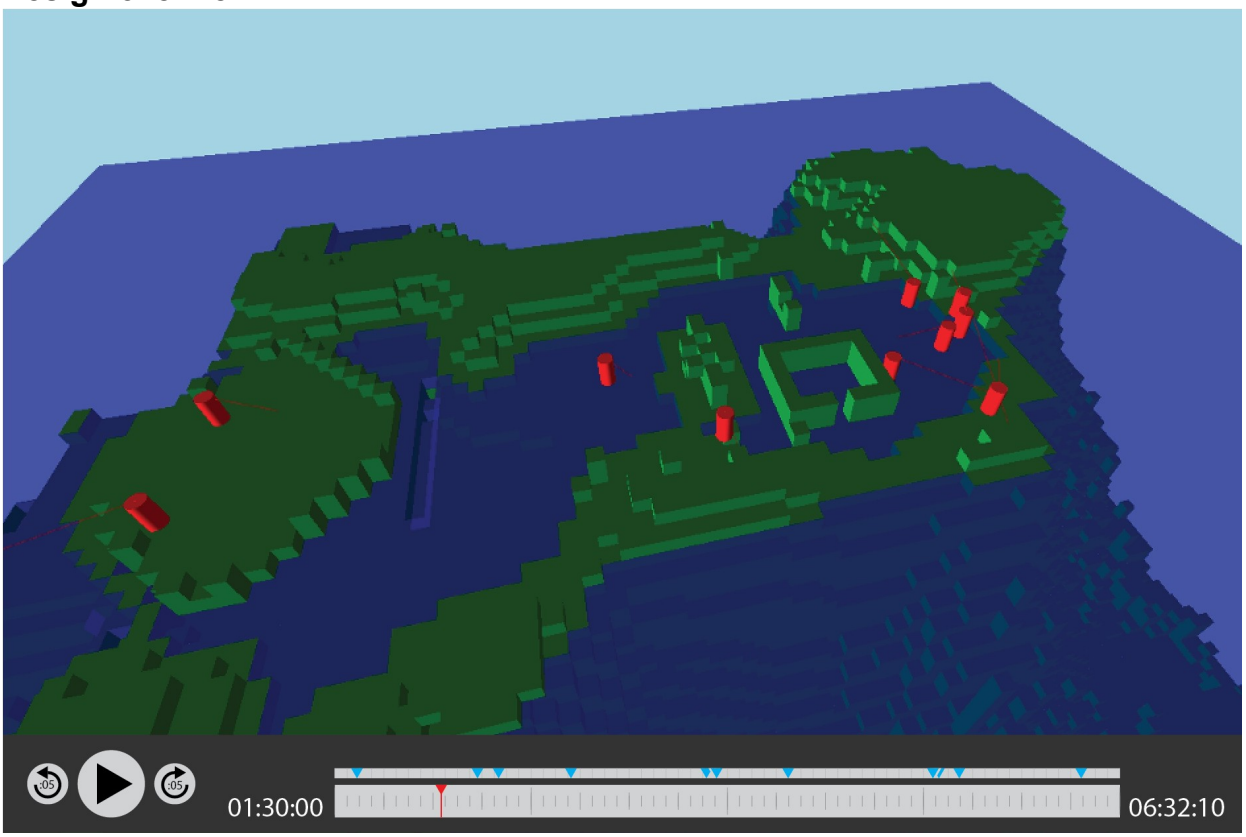

Figure 1. Screenshot of MultiMod. Users (represented by red cylinders) are collaboratively designing a community space on an island in Minecraft.

We had four design goals for MultiMod:

- ***Aggregate*** multiple analytics streams into a simulation which can be experienced and analyzed.
- ***Select*** spatial and temporal bounds, as well as data layers representing phenomena of interest.
- ***Analyze*** the phenomena of interest using qualitative coding, supported by upstream automated preprocessing of constructs relevant to the conceptual framework.
- ***Present*** an annotated simulation which evokes a sense of authenticity and presence.

These goals are enacted through the design of MultiMod's interface, shown in Figure 1. The interface consists of two panes: the view and the timeline. The view presents an animated, three-dimensional representation of users, interactions, and environments, as well as overlays showing analytical products such as joint visual attention or the traces of players' movement. Figure 1 shows users (the red cylinders) as well as each user's gaze vector (the red lines), which connects their eye position to the block they are looking at. The researcher can pan, rotate, and zoom the scene, and can view the scene on a computer screen or in virtual reality. The timeline displays layers which are time-indexed but which do not have a spatial dimension, for example, line graphs and time stamped qualitative codes. The play head indicates the current time, and it can be dragged to scrub across time or to begin replay at a specific point. During replay, temporal layers such as audio are played if they are included in the instance of MultiMod. Finally, MultiMod functions as a qualitative coding tool for multimodal learning analytics. Existing tools such as Atlas.TI allow qualitative coding of temporal artifacts such as audio and video recordings; MultiMod functions in a similar way but allows coding of the multilayered analytics described in the previous paragraphs.

We wanted MultiMod to provide an interactive "window" into the reality which produced the multimodal analytics, allowing for an immersive experience during analysis and subsequent presentation of the research findings. Taking up Eisner's (1997) call for capturing *authenticity* in data representations, we wished for MultiMod to provide a phenomenological sense of immersion in the data, a feeling of being present to the phenomena under study.

System description

MultiMod is a standalone html page (typically around 25 megabytes) with all dependencies and assets bundled together, making it easy to share and host on a server. The runtime is implemented using the vue.js, d3.js, and three.js libraries. Several optimizations from the field of computer graphics were used to improve the file size and runtime performance.

A particular instance of MultiMod is produced using Proctor & Muller's (2022) framework for integrating multimodal data streams. The parameters for each instance, specified in a declarative configuration file, include: the bounding box in the x, y, and z dimensions; the start and end timestamps; and a list of layers to include along with their own parameters. Each layer requires a description of initial state and a list of ops. An op consists of a timestamp and a description of the forward and the backward transition. This makes it possible to move forward through time by sequentially applying ops' forward transitions, or to move backward in time by sequentially applying ops' backward transitions. The following paragraphs describe the data sources and preprocessing required to prepare initial state and ops for each layer. These layers are specific to our use case of analyzing interaction in Minecraft, but many other data sources could be integrated by specifying new kinds of layers and preprocessing data accordingly.

The terrain layer (shown in green and blue in Figure 1) represents land and water in the Minecraft world. During our workshop (Proctor & Muller, 2022), we logged about 250,000 events in which a player placed or removed a block; each such event is represented by an op. We also saved a copy of the original world state in Minecraft's native .mca file format. In order to generate the initial state for an instance of MultiMod, we read the original voxel data for the specified bounding box from the saved .mca file, and then apply all the ops prior to the specified start timestamp. The ops between the start and end timestamp are provided to the terrain layer.

The player location layer represents the locations of players (marked with red cylinders in Figure 1). We logged over six million player move events; each is translated into an op specifying a change from one location to another. These same events are used to produce the player gaze layer, shown with red lines in Figure 1. At each player move event, we also log the direction of the player's gaze and the target block, if one exists, by projecting the gaze vector from a player's in-game eye location until it hits a block or reaches a distance threshold. The target block tells us where the player's focus was at that moment. While playing Minecraft, the target block is the one under a small crosshair at the middle of the screen, and it is given a subtle visual highlight.

While the terrain layer, the player location layer, and the player gaze layer represent observations which would have been visible to workshop participants, additional analysis layers can also be added. For example, the trace layer represents the past and future locations of players throughout the segment by highlighting the locations in their trajectories. The joint visual attention (JVA) layer represents three-dimensional areas in which two or more players establish joint visual attention during the segment. (The extension of Schneider and Pea's (2013) algorithm for larger groups and three-dimensional environments is the subject of a forthcoming publication.)

Discussion

This work demonstrates an immersive tool for analyzing and representing data collected from collaborative Minecraft sessions. We feel that the multimodal representational elements and interactivity afforded by MultiMod provide an effective platform for gathering novel insights around collaborative Minecraft gameplay. Further, we feel that particular affordances of interactive simulators which are exemplified in MultiMod reveal exciting opportunities for creating new forms of data analysis and presenting arguments in a variety of other research contexts.

First, the unique form of representing data in MultiMod provides a platform for observers to engage with research data as if they were present in the activity itself. Eisner (1997) highlighted the promises of providing conditions for capturing authenticity in the use of alternate forms of data representations; we believe that MultiMod indeed captures this authenticity, occasioning a phenomenological sense of "being there" while viewing these Minecraft sessions in both non-immersive and immersive virtual reality (Freina & Ott, 2015). By

way of the camera and rendered three-dimensional world, users of MultiMod are thrust into the environment in which data was originally collected and are immersed in a multimodal representation of this collected data.

Through the use of MultiMod, we have found evidence that modalities of communication contribute to effective collaboration according to nine dimensions as defined by Meier, Spada, and Rummel's (2007) rating scheme. Specifically, emerging findings point to players coordinating block manipulation, gaze, and spatial positioning, and discourse to engage in collaborative activity. Here, players recruit combinations of these modalities in distinct ways to collaborate, and this collaborative process can be measured by these multiple dimensions of collaboration. As a defining example, in several instances we have found that block building is *contagious* during collaborative activity within Minecraft—when players witness the act of building by a fellow player within close proximity (as measured by gaze, relative position, and block manipulation events) they often respond by engaging in block building themselves. Resultantly, players interacting within these collaborative building acts demonstrate effective technical coordination, sustaining mutual understanding, and task division, three dimensions of collaboration as characterized by Meier, Spada, and Rummel's (2007) framework.

We believe this work presents opportunities for research beyond its immediate scope. We envision that future annotated simulations which incorporate multimodal analytics might support researchers in understanding findings and generating new insights into research. In this way, qualitative coding in similar simulators might assume new forms that draw on the spatiotemporal nature of the simulator and the multimodal data presented within. Here, qualitative codes might one day take the form of "four dimensional" annotations that both describe phenomena in spatial and temporal terms and are visualized similarly. In similar simulations, codes might be anchored to spatial elements as well as temporal points, presented to the researcher in both space and time. Moreover, using multimodal data streams as input for this type of tool could afford triangulation processes among modalities. For example, a researcher or AI might deliberately apply codes to a particular spatially represented entity within a simulation each time an associated voice is identified in an audio layer. Ultimately, we see a rich future for annotated simulations similar to MultiMod, where novel data analysis and presentation forms emerge from the unique modes of interaction afforded by these tools.

References

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238. https://doi.org/10.18608/jla.2016.32.11

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J. L., Sherin, M. G., & Sherin, B. L. (2010). Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology, and Ethics. *Journal of the Learning Sciences*, *19*(1), 3–53. https://doi.org/10.1080/10508400903452884

Eisner, E. W. (1997). The Promise and Perils of Alternative Forms of Data Representation. *Educational Researcher*, *26*(6), 4–10. JSTOR. https://doi.org/10.2307/1176961

Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. *Elearning & Software for Education* 1, 133-141.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. International Journal of Computer-Supported Collaborative Learning, *2*(1), 63-86. https://doi.org/10.1007/s11412-006-9005-x

Proctor, C., & Muller, D. A. C. (2022). Joint visual attention and collaboration in Minecraft. *Proceedings of 15th International Conference on Computer Supported Collaborative Learning*.

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning, 8*(4), 375–397.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.