Video blogs (vlogs) undergo two stages of production, the actual filming and the subsequent editing, before they are uploaded to YouTube, where they are recontextualized as part of a number of interactive elements making up the vlog environment. Thus, between the production of speech, gestures, and written language by the author of the video and the reception of it by the viewer, there is both a temporal and a spatial gap to be considered. Video bloggers have developed a range of practices that both indicate their awareness of these vastly different contexts and represent attempts to make their interaction with the viewers cohere across them.

The present research shows that vloggers, whose interaction originates in a three-dimensional physical space, use gestures that are adapted to the two-dimensional virtual space of the website. These gestures include pointing gestures whose referent is located in the vlogger’s physical surrounding as well as those whose referent is located on the website outside of the video frame itself. Other practices play with the sequential unfolding of turn-taking that is a characteristic of face-to-face conversation: vloggers respond to written comments by viewers not necessarily in temporal sequence but by simultaneously showing a screenshot of selected written comments and responding to them using spoken language. Similarly, viewers use certain practices to signal how their written comments tie back to anything in either the audio- or the visual stream of the video, including both affordances inherent to the website design and verbal means such as quoting.

This paper describes the practices of meaning making in a complex multimodal asynchronous environment and explains them in terms of their function to enable coherent exchanges.

Keywords: gesture, video blogs.