

The Future of 'Video' in Video-Based Qualitative Research Is Not 'Dumb' Flat Pixels!

*Exploring Volumetric Performance Capture and Immersive Performative Replay*

McIlvenny, Paul Bruce

*Published in:*  
Qualitative Research

*DOI (link to publication from Publisher):*  
[10.1177/1468794120905460](https://doi.org/10.1177/1468794120905460)

*Publication date:*  
2020

*Document Version*  
Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
McIlvenny, P. B. (2020). The Future of 'Video' in Video-Based Qualitative Research Is Not 'Dumb' Flat Pixels! Exploring Volumetric Performance Capture and Immersive Performative Replay. *Qualitative Research*, 20(6), 800-818. <https://doi.org/10.1177/1468794120905460>

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

**Take down policy**

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

## **Title**

The future of ‘video’ in video-based qualitative research is not ‘dumb’ flat pixels! Exploring volumetric performance capture and immersive performative replay

## **Author**

Paul McIlvenny, Department of Culture and Learning, Aalborg University, Kroghstræde 3, Aalborg 9210, Denmark

## **Abstract**

Qualitative research that focuses on social interaction and talk has been increasingly based, for good reason, on collections of audiovisual recordings in which 2D flat-screen video and mono/stereo audio are the dominant recording media. This paper argues that the future of ‘video’ in video-based qualitative studies will move away from ‘dumb’ flat pixels in a 2D screen. Instead, volumetric performance capture and immersive performative replay relies on a procedural camera/spectator-independent representation of a dynamic real or virtual volumetric space over time. It affords analytical practices of re-enactment – shadowing or redoing modes of seeing/listening as an active spectation for “another next first time” – which play on the tense relationships between live performance, observability, spectatorship and documentation. Three examples illustrate how naturally occurring social interaction and settings can be captured volumetrically and re-enacted immersively in virtual reality (VR), and what this means for data integrity, evidential adequacy and qualitative analysis.

## **Keywords**

Video research, Big Video, ethnomethodological conversation analysis, volumetric, virtual reality, immersive qualitative analytics

## **Author biography**

Paul McIlvenny is professor in the Department of Culture and Learning at Aalborg University. He is research leader of the Centre for Discourses in Transition (C-DiT) research

group. His current concern is to develop immersive visualisation technologies (XR) to investigate the relationship between mobilities, embodiments, mediated social interaction and discourse in complex environments.

---

## Introduction

Qualitative research that focuses empirically on practical action, social interaction and talk in social practices has been increasingly anchored, for good reason, in working with collections of audiovisual recordings of natural occasions. This paper will focus on the methodological approach of ethnomethodological conversation analysis (EMCA), but many points are relevant to a broader video-based qualitative research agenda (Flick 2018). For both, 2D flat-screen video and mono/stereo audio recordings have been, and still are, the dominant recording media for data collection. Taking a different tack, this paper makes the case that there is a future for ‘video’ in video-based qualitative research that in some ways is *not* video per se. Ultimately, video-based qualitative studies will migrate away from ‘dumb’ flat pixels in a 2D screen towards volumetric performance capture and immersive performative replay in virtual reality (VR).<sup>1</sup> It is true that certain traits of the use of video in video-based research will remain important, such as replay and freeze frame, but recent and emerging technological possibilities require creative rethinking and bold experimentation that will have significant methodological consequences.

As a result of practices and technologies of virtualisation, we are on the cusp of a *scenographic* turn that entails a rethinking of space, volume and movement. There is a need for a more scenographic imagination (Hahn, 2018; McKinney and Palmer, 2017) – for instance, to draw from, but not be bound by, immersive theatre and computer gaming, and thus to discover and multiply an innovative apparatus of observation with fresh ways to ‘sense-with-a-camera’ and ‘sense-with-a-microphone’. Unlike ‘video’, volumetric performance capture relies on a procedural camera/spectator-*independent* representation of a dynamic real or virtual volumetric space over time. A complete, reliable workflow for live volumetric performance capture in everyday physical settings is not yet commercially available nor computationally feasible today, but this paper maps out its possibilities and limits, primarily for data collection and staging the data, through consideration of current examples of novel *immersive qualitative analytics* (IQA) software and new types of time-

based data collection that go beyond (and yet remain within the orbit of) video. Three examples illustrate how naturally occurring social interaction and settings can be captured volumetrically over time and what this means for data integrity, evidential adequacy and qualitative analysis.<sup>2</sup>

### **The prevalence, practices and problems of 2D video and mono/stereo audio**

Since the 1990s, 2D digital video has been stored as an encoded digital file that a computer media player streams and renders according to a codec (coder-decoder algorithm) as pixels on a flat screen. However, the unreflective adoption of these complex, often opaque, apparatuses of capture, storage and replay that co-constitute 'objects of observation' has led to 'video' becoming a black-box. For example, in relation to the early days of using video recordings to study social interaction, Erickson (2004: 202) noted that there were "differences in scholars' routine looking and listening practices which resulted from the differing affordances of different kinds of audiovisual reviewing equipment." Hirschauer (2006: 418) argues that "instead of the selectivity of an observer, recordings yield to the selectivity of their media.... The video-camera as well only produces excellent documents when we grant to the selected detail in the camera's display what we would not grant to any human participant: that it is preserving what has 'actually' happened." Moreover, EMCA and media scholars have critiqued the dominant use of (2D) video recordings because of, for instance, a) the inadequacy of video for analysing the situated accomplishment of certain phenomena/orders of everyday action, b) the limitations of video to 'capture' reality, c) the intrusiveness of video recording in everyday settings, d) the screen essentialism and planocentrism of flat video, and e) the false universalism of one-dimensional audio (Liegl and Schindler, 2013; Livingstone, 1987; Lowood, 2016; Schröter, 2014).

A number of scholars in EMCA have given practical advice about collecting data using video cameras, as well as how to view and analyse the resulting footage (Büscher, 2005; Goodwin, 1993; Heath et al., 2010; Knoblauch et al., 2006).<sup>3</sup> Additionally, some studies have topicalised camerawork and the use of video as a set of practices in the workplace or in leisure pursuits (Broth et al., 2014). There are well known problems (and solutions) when using video cameras/audio recorders to collect data in natural settings, but the sociotechnical dimension of the camera/microphone as an apparatus or agency of observation is often

neglected (Barad, 1998: 94; Gallagher, 2019). From the earliest days, some ethnomethodologists, such as Harold Garfinkel and David Sudnow, were aware that “you will have a hard time reproducing the actual spatial relations that members see because the camera does not operate like the eye does” (Hill and Crittenden, 1968: 54). Sudnow claimed that a wide-angle lens would solve the problem, but such lenses tend to flatten space and push objects away from the viewer. In his discussion of the recording practices he developed when using the early portable video recorders, Goodwin (1981: 43) noted “any camera position or framing of participants involves a choice from a set of alternatives and any of the alternatives not selected would have produced a different record of the event.”

From a more critical stance, Branigan (2006: 8) questions the notion of a camera’s ‘point of view’ and the resulting ‘frame’. In interrogating the camera as metaphor, Branigan notes that there are at least fifteen different uses of ‘frame’ as a metaphor in film studies: “It does not simplify matters to say that a camera merely ‘frames’ an object” (8). Branigan, however, is primarily referring to the traditional ‘frame’ of the 2D video camera. What is to be made of passive 360° cameras, which since 2016 have become commercially viable and easily available in the consumer market for adoption by qualitative researchers (Davidsen and McIlvenny, 2016; Gómez Cruz 2017; McIlvenny and Davidsen, 2017)? One could argue that the continuous footage that results from such a multi-lens camera, after the separate wide-angle lenses are algorithmically stitched together into one ‘equirectangular’ frame or ‘cube map’, allows a viewer to see a flat 2D visual representation of the *totality* of a scene from a single location but in all directions at once, i.e. ‘frameless’ without a *subjective contour* (Branigan, 2006: 103-105). The most common view one has onto this representation is usually enframed by the traditional rectangular ‘viewport’, which changes, much as Branigan suggests, as if one ‘turns’ one’s head to look elsewhere in the scene.<sup>4</sup> Thus, the phenomenological experience is similar to that of a singular subject gazing out into the world (through the multiple ‘lens-eye’s centred at the camera’s fixed location).

### **A concern for space and volume in ethnomethodology**

There are a few precursors to the scenographic turn that raise praxeological questions about body, space and perception. Maurice Merleau-Ponty, in his seminal work entitled *Phenomenology of Perception* (first translated into English in 1962), explored the perception

of the visual field, space and verticality, including the reported experiences of wearing inverting lenses or “goggles”. Drawing on a deliberate misreading of Merleau-Ponty, Harold Garfinkel posed what he called ‘tutorial problems’ for his students using the technologies of inverting lenses and auditory side-tone delay (Garfinkel, 2002). In a significant methodological shift, Garfinkel concluded that what could be learnt from wearing the inverting lenses was the achievement of *bodies of practices*: “These bodies have eyes that are skills; eyes that are skills in the ways that eyes do looking’s work. Where seeing is something more, other and different than formal analytically describable positioning the orbs to assure certain retinal registration of a perceptual field, let alone a visual field” (210). One of Garfinkel’s students, David Sudnow, envisaged that if he were to record a man going for a walk, he would need a photographic “camera mounted to his head” (Hill and Crittenden, 1968: 54). In fact, this camera was to record the scene as it appeared to the man, as if present on all sides (cp. 360° camera). Following on from Merleau-Ponty, Todes (2001) gives us a way to talk about what Garfinkel presciently referred to as “bodies of practices”, namely body-direction in the practical visual spatiotemporal field. Todes writes that our left and right “appear to be merely two different sides of our body. But when we are active they generate an apparent spatial field for all objects, and themselves acquire a place in this field. They do this by bearing our capacity to *turn in place*; that is, to turn around without moving or passing, and therefore to turn, in a certain sense, in an instant. Turning in this in-stant (literally, standing-in), or the sense of our ability to do so, apparently gives us at once our simultaneous spatial circum-stance (where what stands around us is), and our own position or place as the in-stance (that which stands inside) of that circum-stance” (49). With passive 360° cameras, we catch a glimpse of how we might begin to recover the lived practice of circum-stance and the artful production of a spatial field.

In his introduction to ethnomethodology, Livingstone (1987) raises the example of discovering the witnessable social order of pedestrian crossings “available to members as situated practices of looking-and-telling” (1). He notes that a sociologist might put a film camera overhead to find disengaged structural patterns in their crossings, but, he argues, to understand how pedestrians cross “we must, *metaphorically*, move the camera to eye level” (22, my emphasis). Here he is careful *not* to assert that actual cameras are comparable to eyes

or that they should be mounted at eye level, but that any form of documentation should take into account the crossers’ bodies and eyes in the spatial practices of crossing.

The visual has most often taken primacy over sound, but there are occasional hints that there are also specific issues with sound, space and volume.<sup>5</sup> As part of their argument about hearability and the ringing of a telephone as a hearable summons, Garfinkel and Wieder (1992) acknowledge as an aside that the directionality of sound is significant in terms of the “listened-for direction of the ring” (197). They refer to David Sudnow’s concern with “sounded doings” and note that “the direction from which a sound is heard is a detail with which the listened to sound is recognized and identified as a sounded doing, that is, the *sound-of-the-coherent-object*; the coherence-of-details-developingly-listened-to-and-listened-for” (197). In Bjelić (2019)’s excellent retelling of Garfinkel & Wieder’s argument about ‘hearability’ in relation to CA’s focus on ‘hearship’, he notes that “CA reduces members’ *spherical* hearing as a phenomenal background to the logic of conversation, because for CA, the context of a conversation is internal to its recordable features” (my emphasis). This is the only mention in his paper of ‘spherical hearing’ (e.g. the hearing of spatial sound or spatial audio surrounding the listener) and it is not elaborated. In Mengis et al. (2018)’s exploration of the ways in which camera technique shapes what can be ‘seen’ in the resulting recordings – the praxeology of camerawork – they largely ignore sound and audio, but they do note that their study “indicates that the audio aspect of video recording also contributes to our understandings of space, and this dimension should also be systematically explored” (310).

### **Beyond ‘dumb’ flat moving image pixels?**

Even though we can recover some sense of space from a 2D perspectival image, and there is a kinetic depth effect – a perceptual cue about space, that arises exclusively from the movement of a camera (Branigan, 2006: 9) – there has been very little reflection on either the dominance of the monoplane image or the flat quality of the pixelated digital image in qualitative research, and very few studies that draw upon and/or record stereoscopic video (and spatial audio) for analytical purposes (McIlvenny, 2019a; McIlvenny, 2019b). To get a better understanding of how we can move beyond 2D video, Schröter (2014) proposes that there are four series of optical knowledge: geometrical optics, wave optics, physiological optics and virtual optics. For Schröter, there is “a blind spot in the existing historical studies



chronicling optical or visual media” (3). They have not adequately accounted for what he provisionally calls “the history of the technological transplane image.” Unfortunately, he argues, “planocentrism privileges the plane that supposedly is readily comprehensible ‘at a single glance’, i.e. as a presence within the awareness of an observer who is present him- or herself” (35). He defines the ‘spatial image’ as “the comprehensive category comprising both transplane images and those that have a three-dimensional material support (like types of sculpture or, for example, globes)” (38).<sup>6</sup> Galloway (2014) argues that there was a bifurcation in the mid-nineteenth century between i) a cinematic concern with a point of view of a singular experience of a gazing subject/lens standing in one central location, and ii) an anticinematic mode in which the eye was virtualised into a “metastatic virtual camera able to view an object from any point of view whatsoever” (66). Consequent on the extensive exploration of the first branch of the bifurcation, arguably leading to the predominance of 2D video in qualitative research, this paper revisits that bifurcation to see what can be gleaned from the second branch. The second proliferated the number of points of view dispersed within a space to conceive of the multiple points of view as temporally synchronous, but was limited by the technology of the day and the phenomenological paradigm. With the increased powers of spatial computing and new forms of virtual reality technology (VR) it is an appropriate time to return to this bifurcation.

### **Pioneers in volumetric performance capture and replay**

Schröter (2014) and Galloway (2014) have highlighted an alternative history of volumetric media in the nineteenth century in which point of view has no meaning. In the 1860s, François Willème’s ‘photosculpture’ technique, using a ring of discrete but synchronised photographic cameras around a fixed subject, permitting the “virtualization of the eye” (Galloway, 2014: 66). His work, and that of Braune and Fischer in the 1890s (Zielinski, 2006: 245-248), added dimensionality by disrupting “the singular experiences of a central gazing subject (or lens eye), however much it may be complicated by montage or the use of two or three concurrent cameras” (Galloway, 2014: 63).

Over one hundred years later, the pioneers in volumetric performance capture include motion capture, volumetric film, Machinima, mixed reality and VR games. It was in 3D video games that a new movement called Machinima developed in the 1990s to remix well-known games



and create dramatic narratives using the games and their assets as a backdrop (Lowood and Nitsche, 2011; Ng, 2013).<sup>7</sup> Greenhalgh et al. (2000) documented one of the first attempts to build a software solution for capturing and replaying collaborative virtual environments in innovative ways. The Digital Replay System further developed the highly flexible ways a user could navigate a complex corpus of recordings/data streams in a collaborative environment (Benford and Giannachi, 2011; Crabtree et al., 2015; Greenhalgh et al., 2007; Murgia et al., 2008). More recently, computer scientists and performance art theorists have combined their efforts to develop systems such as *CloudPad*, with which anyone can regenerate a hybrid replay of the mixed reality performance (Giannachi et al., 2012). Steptoe and Steed (2012) suggest a reference architecture for replay that “collates multiple components of a user’s nonverbal and verbal behavior in single log file, thereby preserving the temporal relationships between cues” (Steptoe and Steed, 2012: 388). Lastly, in theatre performance studies, Delbridge and Tompkins (2012: 62-63) argue that motion capture technology has led to a shift in the understanding of the documentation of an actor’s body in terms of (noncinematic) *frameless* movement in a 3D *capture volume* that the system can ‘register’.<sup>8</sup>

Although a complete, reliable workflow for live volumetric performance capture in everyday physical settings is not yet feasible at present, we can begin by mapping out its possibilities and limits through consideration of examples of novel software and new types of time-based data collection that go beyond video.

### **Towards volumetric capture by staging and inhabiting video data**

Increasing the complexity of recordings, both 2D and 360°, of natural settings has encouraged scholars to explore how to better represent space and volume in social interaction and practices, which in turn results in a reconsideration of 2D and passive 360° video recordings and their limitations for analysis (McIlvenny, 2019b). The complex, time-based data sets that result are hard to visualise using traditional solutions. An alternative approach within *immersive qualitative analytics* (IQA) is to *stage* video by reconstructing the site and the scenes of social interaction over time in an interactive and immersive 3D representation or model in VR, and to *inhabit* video by exploring complex spatial video and audio recordings of a single scene through a tangible interface in VR. Examples of novel software

tools to stage, inhabit and analyse passive 360° videos include *SQUIVE* (Staging Qualitative Immersive Virtualisation Engine) and *AVA360VR* (Annotate, Visualise, Analyse 360 videos in Virtual Reality).<sup>9</sup> Because volumetric video capture is in its infancy, the *SQUIVE* staging software engine is a foretaste of what it would be like to inhabit a complex site of social interaction *as if* it was recorded volumetrically.

### ***Example 1***

*SQUIVE* has been applied to the staging of a specific complex site of social and material conduct which was recorded over five days with several 2D and 360° cameras and traditional and ambisonic microphones in demanding conditions with respect to lighting and sound. The event was an award-winning interactive exoskeleton robot-human theatre performance entitled *Inferno* at a municipal theatre, including four public performances and some additional experiments on the role of music and dance on the receptiveness of the subjects to wearing the performative exoskeletons.<sup>10</sup> Using the 360° video recordings as a memory aid, *SQUIVE* was used to construct a bespoke interactive 3D model. Avatars, lighting, furniture and props were styled to reflect the ambience and actants in the relevant spaces in which the performances took place. With a head-mounted display (HMD), a user can walk around the simulated theatre and encounter the different rooms and connecting spaces that the participants interacted in and moved through. A user can interact with the cameras that were used to record scenes and they can hear spatial sound in the space as it was recorded. Footage from the physical cameras can be selected, played, cloned and clipped (see Figure 1).

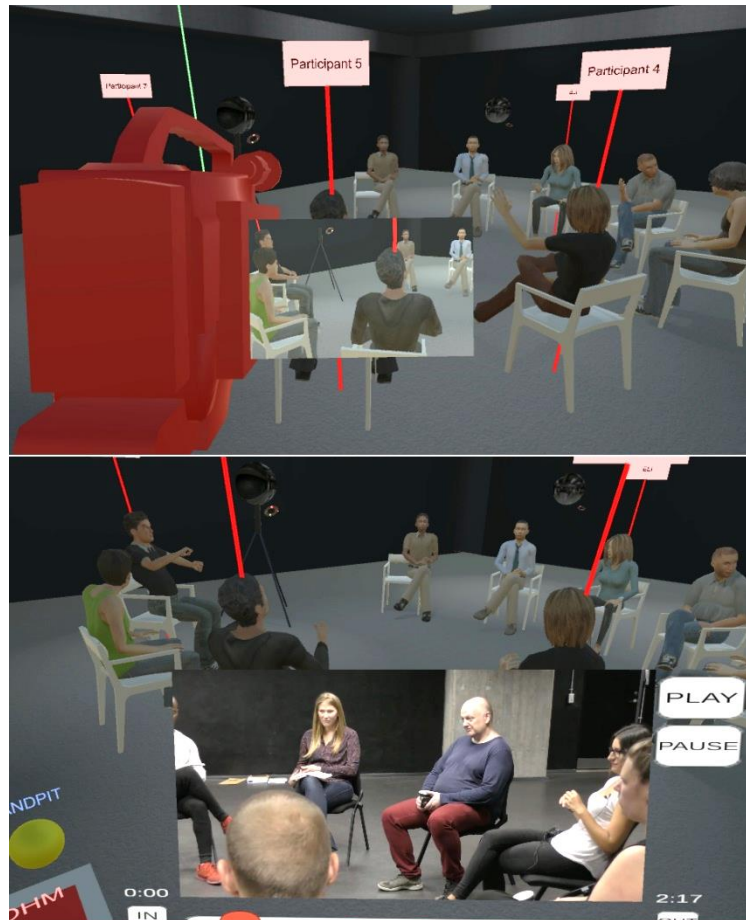


Figure 1 – Virtual cameras, avatars and real footage of a focus group staged in *SQUIVE*

A five-minute video clip can be found online which shows a 2D screen capture of the viewport of a user in *SQUIVE* on a walking tour through the theatre site showing time-slices of the cohort of experimental subjects engaged in different activities across four rooms, e.g. briefing and dressing, contact dance preparation, exoskeleton performance, and focus group.<sup>11</sup> Animated avatars substitute for each participant and the reconstructed 3D scene is populated with virtual cameras that substitute for the actual cameras that recorded the event. Each virtual camera can be selected to re-view the reconstructed scene from the perspective of the actual 2D and 360° camera footage. Moreover, a user can launch *AVA360VR* (McIlvenny, 2019b) from any virtual 360° camera, so that the user can annotate the clip recorded by any physical camera or microphone in the scene concurrently during the same time-slice.

The reconstruction in *SQUIVE* stands on its own as an anonymised, interactive *volumetric* archive of the event *and* its documentation. In its current state, the user can alternate between

*navigating* the staged virtual volumetric reconstruction and *re-viewing* the physical camera footage from actual locations in the space of the event. Therefore, *SQUIVE* is a useful antidote to mono-planar recordings and representations, since one is continually questioning the evidential adequacy of each mode of staging and inhabiting. Of course, analytical care must be taken in order *not* to make strong claims about conduct from observation and interaction with the staged models alone. The virtual models are a tool to organise qualitative enquiry; they are *not* a substitute for ethnography, skilled camerawork and originary sources.

### **Two examples of volumetric performance capture in virtual reality**

As I have argued above, there are problems with treating 2D video recordings as veridical *reproductions* of past social events, especially with regard to access to scenographic orders and spatial experience. Volumetric performance capture, on the other hand, affords analytical practices of re-enactment, shadowing or redoing modes of seeing/listening as an active spectation for “another next first time” (re-documenting). The two examples that follow illustrate how naturally occurring social interaction in virtual settings can be captured volumetrically. In each case, analytically relevant aspects of the spatio-temporal flow and performance of the event are recorded. They are immersively replayable *as if* live from any spectatorial position. Of course, the recordings presented in the following examples are from screen captures of the computationally rendered scene on a computer display ‘viewport’, but this is for convenience given the limitations of traditional academic publishing genres. The ‘data’ is really the volumetric performance capture that *can be* re-enacted. Auslander (2018) contends that re-enactments of documentations are always productions, not reproductions, in a new context. They are *not* historical documents that give access to objective experiences of the original performance. Auslander suggests that ‘reactivation’ might be the better choice for accounting for the relationship in which “each reactivation discloses the original, but discloses it under different circumstances” (85). The ‘data’ in the examples below only lives phenomenologically *when* reactivated, so the unique traces of performance are derivative of the *capacity* for immersive, performative replay.

### *Example 2*

On the other end of the continuum from Example 1, ‘data’ was collected from computer gaming environments that can capture and re-enact a player’s 3-D avatar game play in VR for another player to re-inhabit. Example 2 focuses on an instructed action sequence in a social VR game (*Mindshow*), in which players record, edit and share staged scenes with avatar puppets that are controlled by the player.<sup>12</sup> Shared scenes can be reinhabited and modified by other players, thus supporting a remix culture. An experienced player, John (JOH), has captured themselves giving instructions how to act out a 3D animated character/avatar (circled in shot 1 in Figure 2) in a scene in anticipation of a new player in the future wishing to learn how to do the same.<sup>13</sup> In 30 seconds, John performs a short, public tutorial entitled “How to Make a Mindshow 1” in the direction of a virtual placeholder (circled in shot 3 in Figure 2, but also visible in shots 4 and 9-16) that signifies the position of the player when they start/inhabit the re-enactment. The scene was ‘recammed’ – re-filmed as a 2D video during a re-enactment – from multiple angles. Sixteen unique re-enactments (shots numbered 1-16 in Figure 2) were undertaken by the author for “another next first time” (Garfinkel, 1996: 10).<sup>14</sup> For the purposes of documentation, a 2D composite video was made comprising the sixteen mainly static, handheld virtual camera angles, including three performances of laic analysis (shots 14-16), recording the viewport of a mobile spectator within the re-enactment of the ‘original’ scene (see Figure 2).<sup>15</sup> Following Mengis et al. (2018)’s typology of camera views, the sixteen shots were indicative of the Panoramic View (3, 4, 9, 10, 11, 12, 14, 15, 16), the American-Objective View (1, 2, 5, 6), the Roving Point-of-View (mobile virtual cam, 7, 8, 13), and the Infra-Subjective View (All, HMD).<sup>16</sup> As this is volumetric capture, all of these views are represented *uniquely, simultaneously, and multiple times*, in the composite video of the same tutorial scene with John.



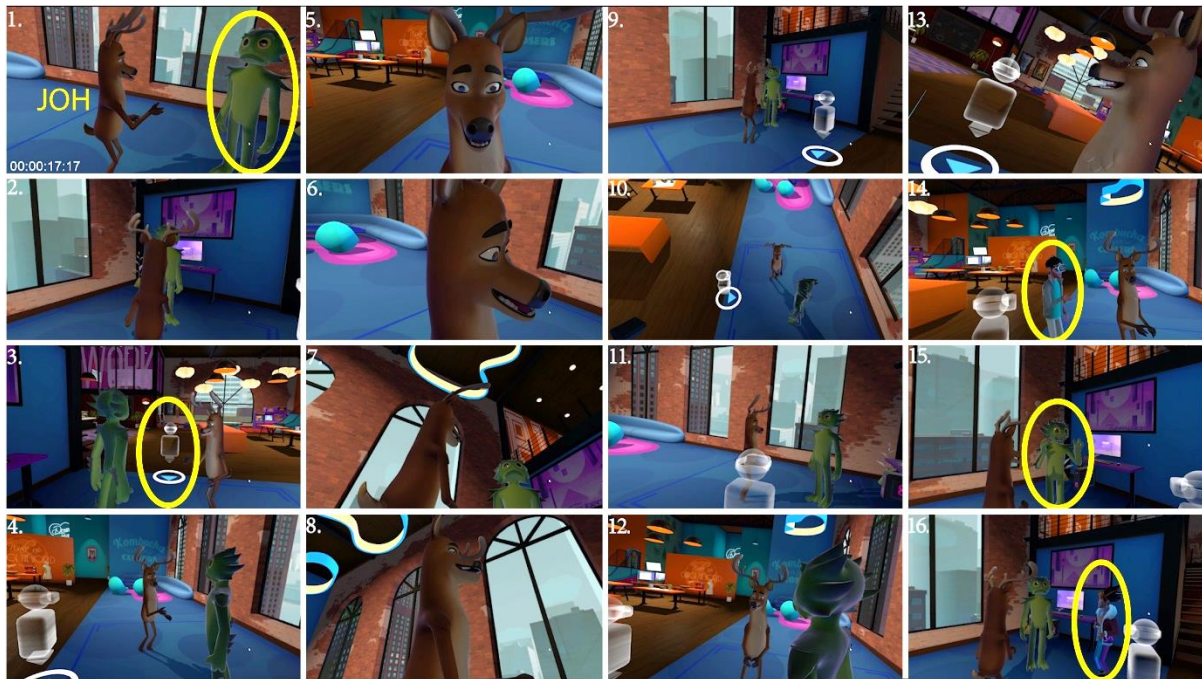


Figure 2 – Composite of sixteen Mindshow replays

A 30-second video clip is available online that shows the complete instruction sequence with sixteen synchronised unique replays.<sup>17</sup> The audio track is from the original voice track recorded by John. A speech-centred, script-based transcript is included with the online video clip that characterises the speech of John in relation to the gross movements of his head/HMD and hands/controllers.

This example demonstrates how an instructed action unfolds in a virtual environment in anticipation of a re-enacted viewing in the future. In this case, unlike in YouTube instructional videos, it is not the video itself which is the accountable documentation of the instructed action, it is the volumetric performance capture (combined with the temporality of the audio recording of voice) which is the documentable. Although the performance is designed for a future player occupying a specific location (indicated in shot 3 in Figure 2), that so-positioned-player can move their head during the re-enactment, and one can also inhabit the scene as a spectator from any position anyway. The performance accomplishes its viewability in a typical facing-formation (an ‘F-formation’) constituted by a circular vis-à-vis arrangement of bodies delineating an ‘o-space’ (Kendon, 1990; Kendon, 2010), in which the instructor instructs a future *Mindshow* player in front of them, a virtual arrangement which includes the static object/avatar. This is a familiar triadic face-to-face social encounter.

### Example 3

The third example illustrates another instructed action sequence in a different social VR game (*VREAL*), in which a player, known as “Alliehugs”, deploys another VR tool (*Tilt Brush*) to instruct a future player how to use the game to capture volumetrically their own gameplay, and replay someone else’s, in a supported third-party VR game.<sup>18</sup>



Figure 3— Composite of VREAL replays

Unlike *Mindshow*, which is self-contained, *VREAL* cleverly builds a capture/replay layer on top of other selected VR games, i.e. it allows the player to capture their normal use of a VR game and for anyone to re-enact what was captured from any perspective in the 3D scene that was recorded.<sup>19</sup> A composite video was made by the author comprising multiple replays recorded from a variety of unique (and arbitrary) virtual camera angles rendered to the viewport of the player, including a) an original player HMD viewport, b) a mobile spectator POV viewport, c) a static aerial cam, and d) a static behind-a-poster cam (see Figure 3). A three-minute video clip is available online that shows the beginning of the tutorial by Alliehugs.<sup>20</sup> The composite video comprises two videos made up of four replays edited together in sync with the original voice track. The clip begins after 1:21, and at this timecode cam1 shows a POV and cam2 shows a HMD viewport. A speech-centred, script-based transcript is attached to the online video clip that represents the speech and movements of



Alliehugs in relation to her gestures and the visual props in the virtual environment. Given there are fewer ‘recams’ than in Example 2, shifts between virtual cameras are noted explicitly.

Although similar in some respects to Example 2, this example demonstrates more sophistication in its mode of volumetric documentation of actual game play. In analytical terms, we get to re-view a richer instructional sequence – e.g. there is a broader toolset available, repurposed from the VR game that VREAL piggybacks on. This instructional sequence is designed for a future player in that it accomplishes a spherical co-viewing. The performance accomplishes its viewability in the virtual volume through the instructor’s staging of information screens in a circle encompassing the virtual participants. Also, the instructee-to-be is explicitly invited to occupy the space in a negotiable, virtual over-the-shoulder or side-by-side arrangement that co-constitutes the F-formation of their focused virtual social encounter.

## **Conclusion**

As discussed above, it is not yet possible to fully capture a naturally occurring physical scene volumetrically with the necessary degree of fidelity and reliability required for qualitative research. Nevertheless, the first example above illustrated how a volumetric performance capture could be envisaged via a 3D reconstruction of the site as an interactive archive, which provided a spatial interface to the actual recordings made of the event. The second and third examples took place purely in virtual reality, in which it is possible to capture the performance volumetrically and replay it in VR. It is important to note that even though the latter may appear to be an inferior (computationally rendered) life-world, which has little bearing on the complexity and richness of our everyday sociomaterial worlds, that is beside the point. Despite their limitations, such cases still demonstrate that members treat them as having a scenic (and sonic) intelligibility; they are virtually real (Shields, 2003).

This paper has focused on new forms of volumetric data collection, as well as the staging and inhabiting of such data for the purposes of qualitative research. In such cases, the technologies of capture and replay are congruous to a degree with the technologies and techniques of qualitative data analysis, and yet there are differences (Flick 2018). If one does have the technology and the skills to make a volumetric performance capture, then a strategy

to begin analysis might be, for instance, to re-enact the volumetric performance capture and find positions and angles on the witnessability, observability and visibility of the phenomena. One could, for example, take a panoramic stance to capture the whole scene from the outside, or roam around the scene as a mobile spectator or take the embodied position of each participant, i.e. their head movements, eye movements and body movements. One could ‘recam’ those views that are significant. With such a strategy, there are clear advantages and affordances that volumetric performance capture and immersive performative replay has over traditional 2D video recordings. For example, there is (a) frameless capture, (b) the freedom to move the camera during a continuous replay shot, (c) the freedom to review the scene from any position and angle, and (d) the opportunity to hear the directionality of sounded doings. Some positive use cases unique to the approach considered in this paper include:

1. A means for other researchers to re-view the audio or video ‘data’, depending on ethical constraints, and find alternative accounts based on a re-seeing or a re-listening<sup>21</sup>;
2. A means to resolve occlusions in the 2D video image, e.g. a gesture or shift in eye gaze is occluded by another person or object in the scene;
3. A means to resolve views of the scene that are not viable with a specific configuration of passive 2D and 360° footage;
4. A better purchase on the scenic and sonic, namely volumetric or spherical, intelligibility of bodies and objects in practical activities;

Some researchers might argue, for instance, that the *virtualisation* of video as a technology of re-viewing is unnecessary or a distortion of a participants’ perspective, which ordinary video escapes. However, this stance is vulnerable because 2D ‘video’ (and ‘audio’) itself has always relied on a *virtualisation* of detail inside the ‘frame’ that can be recovered by replay, yet which is not available to the participants. Additionally, claims that monophonic audio is exempt from the epistemological problems faced by ‘subjective’ video data are often anchored in an uninterrogated falsely universalist notion that sound is uniform and equally accessible to all. Suchman and Trigg (1991: 78, my additions) note that “video-based interaction analysis affords a powerful corrective to our tendency to see [hear] in a scene what we expect to see [hear].” It instils a healthy scepticism about the validity of observations that were made without the possibility to check the record more than once. What is the

significance of this observation when there is no videotape and no definitive ‘record’? I contend that it leads to a renewed scepticism about the validity of observations that were made based on ‘the record’ when there is no possibility to check it (against) the records-to-come.

It is important for qualitative research to open up a dialogue about a new set of metaphors, assemblages, optics and acoustics, and critically examine their ‘phase cancellations’, ‘dead zones’ and ‘blind spots’. As Liegl and Schindler (2013: 262) have argued with respect to video in sociological research, specific vis-abilities emerge, in which “a nexus of knowledge and vision comes into being. In order to comprehend or trace these practices sociologically, one must acquire the practical knowledge of the respective participant in order to be able to see the same situation as the participants see.” We must always interrogate, following Hirschauer (2006: 422), how any new mode of recording transfers “everyday incidents from their native contexts into the context of sociological argumentation by exceeding and falling below the participants’ knowledge”, how they textually reify “a singular event, emancipating ‘data’ from the participants’ control” and establish “a stable empirical referent within sociological discourse”. Within an earlier paradigm of data collection, Ashmore et al. (2004: 355) raise the spectre of ‘tape fetishism’, in which (audio)tapes are treated as having “the power to transport the listener directly to the original event.” We can speculate, following Ashmore and Reed (2000), whether or not a new ‘stronger’ analytical object under different circumstances is engendered through volumetric performance capture, something ontologically closer to the ‘event’ that disrupts the epistemic authority of the ‘Tape’, and, therefore, to which the ‘Tape’ would relate, much as the Transcript does today to 2D video ‘Tape’. Is there a danger of volumetric performance capture fetishism? Auslander (2018: 16-17) inoculates against this by shifting our attention away from the documentation to the experience of the re-activation as performative itself. Although documentation is unequivocally something ‘other than performance’, we can still ask what makes documentation “so inexorably linked to it that it is hard to separate it from it” (Giannachi et al., 2012: 171)? Following Auslander (2009: 85), “each reactivation discloses the original, but discloses it under different circumstances.” Arguably, what volumetric performance capture affords is not *yet* a witnessing, not *yet* a making visible of the work of assembling visible social fields as practical courses of action, not *yet* a finding of the scenic intelligibility

of a course of action, finding in its course the continuous ‘framing’ of the capture. That work and achievement of intelligibility is yet to be undertaken ‘after’ the originary event when a proliferation of simultaneous continuous ‘shots’ can be performatively accomplished. Of course, we must be cautious here; this is not a witnessing in the course of action in the original scene, i.e. the ghostly re-witnessing has no impact on the re-enacted actions.

### **Acknowledgements**

I would like to thank my colleague Jacob Davidsen for our mighty collaboration on Big Video, in which many of the strands and ideas of this article were brewing. Also, much gratitude to Nicklas Haagh Christensen, who was the lead programmer and developer for BigSoftVideo. He suffered long discussions with me on core technical concepts and coding relevant to this article. Colleagues provided useful commentary and criticism in a ‘silent pre-data session’ focusing on the examples presented in this article. The Video Research Lab (VILA) at Aalborg University provided technical equipment and support staff for software development.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **ORCID**

Paul McIlvenny <https://orcid.org/0000-0003-2327-2124>

### **References**

- Ashmore M, MacMillan K and Brown SD (2004) It's a Scream: Professional Hearing and Tape Fetishism. *Journal of Pragmatics* 36(2): 349-374.
- Ashmore M and Reed D (2000) Innocence and Nostalgia in Conversation Analysis: The Dynamic Relations of Tape and Transcript. *Forum: Qualitative Social Research*, 1(3).

Pre-print version of a published article: McIlvenny, Paul (2020). The Future of 'Video' in Video-Based Qualitative Research Is Not 'Dumb' Flat Pixels! Exploring Volumetric Performance Capture and Immersive Performative Replay. *Qualitative Research*. <https://doi.org/10.1177/1468794120905460>.

Auslander P (2009) Reactivation: Performance, Mediatization and the Present Moment. In: Chatzichristodoulou M, Jefferies J and Zerihan R (eds) *Interfaces of Performance*. Farnham: Ashgate, 81-93.

Auslander P (2018) *Reactivations: Essays on Performance and Its Documentation*, Ann Arbor: University of Michigan Press.

Barad K (1998) Getting Real: Technoscientific Practices and the Materialisation of Reality. *differences* 10(2): 87-128.

Bates C (ed) (2015) *Video Methods: Social Science Research in Motion*. Abingdon: Routledge.

Benford S and Giannachi G (2011) *Performing Mixed Reality*, Cambridge, Mass.: MIT Press.

Bjelić D (2019) 'Hearability' Versus 'Hearership': Comparing Garfinkel's and Schegloff's Accounts of the Summoning Phone. *Human Studies*. DOI: 10.1007/s10746-019-09506-6.

Branigan E (2006) *Projecting a Camera: Language-Games in Film Theory*, London: Routledge.

Broth M, Laurier E and Mondada L (eds) (2014) *Studies of Video Practices: Video at Work*. Abingdon: Routledge.

Büscher M (2005) Social Life Under the Microscope? *Sociological Research Online*, 10(1).

Coulter J and Parsons ED (1990) Praxiology of Perception: Visual Orientations and Practical Action. *Inquiry: An Interdisciplinary Journal of Philosophy* 33(3): 251-272.

Crabtree A, Tennent P, Brundell P, et al. (2015) Digital Records and the Digital Replay System. In: Halfpenny P and Proctor R (eds) *Innovations in Digital Research Methods*. London: Sage, 193-220.

Crary J (1990) *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*, Cambridge, Mass.: MIT Press.

Davidson J and McIlvenny P (2016) Guest blog post: Jacob Davidson and Paul McIlvenny on Experiments with Big Video. In: *ROLSI journal guest blog*. Available at:

Pre-print version of a published article: McIlvenny, Paul (2020). The Future of 'Video' in Video-Based Qualitative Research Is Not 'Dumb' Flat Pixels! Exploring Volumetric Performance Capture and Immersive Performative Replay. *Qualitative Research*. <https://doi.org/10.1177/1468794120905460>.

<https://rolsi.net/guest-blogs/guest-blog-jacob-davidsen-and-paul-mcilvenny-on-experiments-with-big-video/>

Delbridge M and Tompkins J (2012) Reproduction, Mediation, and Experience: Virtual Reality, Motion Capture and Early Modern Theatre. In: Roihankorpi R and Paavolainen T (eds) *SPACE–EVENT–AGENCY–EXPERIENCE*. Tampere: Centre for Practise as Research in Theatre, 56-65.

Erickson F (2004) Origins: A Brief Intellectual and Technological History of the Emergence of Multimodal Discourse Analysis. In: LeVine P and Scollon R (eds) *Discourse and Technology: Multimodal Discourse Analysis*. Washington, DC: Georgetown University Press, 196-207.

Flick, E (ed) (2018) *The Sage Handbook of Qualitative Data Collection*. London: Sage.

Gallagher M (2019) Voice Audio Methods. *Qualitative Research*. DOI: 10.1177/1468794119867549.

Galloway A (2014) Polygraphic Photography and the Origins of 3-D Animation. In: Beckman K (ed) *Animating Film Theory*. Durham, NC: Duke University Press, 54-67.

Garfinkel H (1996) Ethnomethodology's Program. *Social Psychology Quarterly* 59(1): 5-21.

Garfinkel H (2002) *Ethnomethodology's Program: Working Out Durkheim's Aphorism*, Lanham, MD: Rowman & Littlefield Publishers.

Garfinkel H and Wieder DL (1992) Two Incommensurable, Asymmetrically Alternate Technologies of Social Analysis. In: Watson G and Seiler RM (eds) *Text in Context*. London: Sage, 175-206.

Giannachi G, Lowood H, Worthey G, et al. (2012) Documenting Mixed Reality Performance: The Case of CloudPad. *Digital Creativity* 23(3-4): 159-175.

Gómez Cruz E (2017) Immersive Reflexivity: Using 360° Cameras in Ethnographic Fieldwork. In: Gómez Cruz E, Sumartojo S and Pink S (eds) *Refiguring Techniques in Digital Visual Research*. Cham: Palgrave Macmillan, 25-38.

Goodwin C (1981) *Conversational Organisation: Interaction between Speakers and Hearers*, New York: Academic Press.

- Pre-print version of a published article: McIlvenny, Paul (2020). The Future of ‘Video’ in Video-Based Qualitative Research Is Not ‘Dumb’ Flat Pixels! Exploring Volumetric Performance Capture and Immersive Performative Replay. *Qualitative Research*. <https://doi.org/10.1177/1468794120905460>.
- Goodwin C (1993) Recording Interaction in Natural Settings. *Pragmatics* 3(2): 181-209.
- Greenhalgh C, French A, Tennent P, et al. (2007) *From ReplayTool to Digital Replay System*. In: 3rd International Conference on e-Social Science.
- Greenhalgh C, Purbrick J, Benford S, et al. (2000) *Temporal Links: Recording and Replaying Virtual Environments*. In: MULTIMEDIA '00: Proceedings of the eighth ACM international conference on Multimedia, Marina del Rey, California, USA, 67-74: ACM.
- Hahn R (2018) *Beyond Scenography*, London: Routledge.
- Harris AM (2016) *Video as Method: Understanding Qualitative Research*, Oxford: Oxford University Press.
- Haw K and Hadfield M (2011) *Video in Social Science Research: Functions and Forms*, Abingdon: Routledge.
- Heath C, Hindmarsh J and Luff P (2010) *Video in Qualitative Research: Analysing Social Interaction in Everyday Life*, London: Sage.
- Hill RJ and Crittenden KS (eds) (1968) *Proceedings of the Purdue Symposium on Ethnomethodology*. Purdue, IN: Institute for the Study of Social Change, Department of Sociology, Purdue University.
- Hirschauer S (2006) Putting Things into Words. Ethnographic Description and the Silence of the Social. *Human Studies* 29(4): 413.
- Jayyusi L (1988) Toward a Socio-Logic of the Film Text. *Semiotica* 68(3/4): 271-296.
- Jochum E, Demers L-P, Vorn B, Vlachos, E, McIlvenny, P and Raudaskoski, P (2018) Becoming Cyborg: Interdisciplinary Approaches for Exoskeleton Research. In: *Proceedings of EVA 18: Politics of the Machines – Art and After conference*, Copenhagen: Electronic Workshops in Computing (British Computer Society).
- Kendon A (1990) *Conducting Interaction: Patterns of Behavior in Focused Encounters*, Cambridge: Cambridge University Press.
- Kendon A (2010) Spacing and Orientation in Copresent Interaction. In: Esposito A, Campbell N, Vogel C, et al. (eds) *Development of Multimodal Interfaces: Active Listening and*



Pre-print version of a published article: McIlvenny, Paul (2020). The Future of ‘Video’ in Video-Based Qualitative Research Is Not ‘Dumb’ Flat Pixels! Exploring Volumetric Performance Capture and Immersive Performative Replay. *Qualitative Research*. <https://doi.org/10.1177/1468794120905460>.

*Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 200*. Berlin: Springer, 1-15.

Knoblauch H, Schnettler B, Raab J, et al. (eds) (2006) *Video Analysis: Methodology and Methods*. Oxford: Peter Lang.

Knoblauch H, Tuma R and Schnettler B (2015) *Videography: Introduction to Interpretive Videoanalysis of Social Situations*, Berlin: Peter Lang.

Laurier E and Reeves S (2014) Cameras in Video Games: Comparing Play in *Counter-Strike* and *Doctor Who Adventures*. In: Broth M, Laurier E and Mondada L (eds) *Studies of Video Practices: Video at Work*. Abingdon: Routledge, 181-207.

Liegl M and Schindler L (2013) Media Assemblages, Ethnographic Vis-Ability and the Enactment of Video in Sociological Research. *Distinktion: Journal of Social Theory* 14(3): 254-270.

Livingstone E (1987) *Making Sense of Ethnomethodology*, London: Routledge & Kegan Paul.

Lowood H (2016) It Is What It Is, Not What It Was. *Refractory: A Journal of Entertainment Media* 27.

Lowood H and Nitsche M (eds) (2011) *The Machinima Reader*. Cambridge, Mass.: MIT Press.

Macbeth D (1999) Glances, Trances, and Their Relevance for a Visual Sociology. In: Jalbert PL (ed) *Media Studies: Ethnomethodological Approaches*. Lanham, Maryland: University Press of America, 135-170.

McKinney J and Palmer S (eds) (2017) *Scenography Expanded: An Introduction to Contemporary Performance Design*. London: Bloomsbury.

McIlvenny P (2019a) How Did the Mobility Scooter Cross the Road? Coordinating with Co-movers and Other Movers in Traffic. *Language & Communication* 65: 105-130.

McIlvenny P (2019b) Inhabiting Spatial Video and Audio Data: Towards a Scenographic Turn in the Analysis of Social Interaction. *Social Interaction: Video-Based Studies of Human Sociality* 2(1).

- Pre-print version of a published article: McIlvenny, Paul (2020). The Future of ‘Video’ in Video-Based Qualitative Research Is Not ‘Dumb’ Flat Pixels! Exploring Volumetric Performance Capture and Immersive Performative Replay. *Qualitative Research*. <https://doi.org/10.1177/1468794120905460>.
- McIlvenny P and Davidsen J (2017) A Big Video Manifesto: Re-sensing Video and Audio. *Nordicom Information* 39(2): 15-21.
- Mengis J, Nicolini D and Gorli M (2018) The Video Production of Space: How Different Recording Practices Matter. *Organizational Research Methods* 21(2): 288-315.
- Murgia A, Wolff R, Steptoe W, et al. (2008) *A Tool for Replay and Analysis of Gaze-Enhanced Multiparty Sessions Captured in Immersive Collaborative Environments*. In: Proceedings of the 2008 12th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications, 252-258: IEEE Computer Society.
- Ng J (ed) (2013) *Understanding Machinima: Essays on Filmmaking in Virtual Worlds*. London: Bloomsbury.
- Schröter J (2014) *3D: History, Theory and Aesthetics of the Transplane Image*, London: Bloomsbury.
- Shields R (2003) *The Virtual*, London: Routledge.
- Steptoe W and Steed A (2012) Multimodal Data Capture and Analysis of Interaction in Immersive Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments* 21(4): 388-405.
- Stevanovic M, Himberg T, Niinisalo M, et al. (2017) Sequentiality, Mutual Visibility, and Behavioral Matching: Body Sway and Pitch Register During Joint Decision Making. *Research on Language and Social Interaction* 50(1): 33-53.
- Suchman L and Trigg R (1991) Understanding Practice: Video as a Medium for Reflection and Design. In: Greenbaum J and Kyng M (eds) *Design at Work*. Hillsdale, NJ: Lawrence Erlbaum, 65-90.
- Todes S (2001) *Body and World*, Cambridge, Mass.: MIT Press.
- Zielinski S (2006) *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means*, Cambridge, MA: MIT Press.

## Notes

---

- <sup>1</sup> In themselves, ‘dumb’ pixels do not contain volumetric information about the scene; they only represent RGB colour values within a uniform 2D grid based on the complex mediations of lenses, sensors, codecs, etc. that use geometrical optics to map a 3D light field onto a 2D planar surface that preserves linear perspective (Crary, 1990).
- <sup>2</sup> For publication purposes, indicative screen shots of the examples are presented as images. Additionally, web links are given to illustrative videos (uploaded to a public, online archive) that were recorded and edited by the author. Each webpage with an embedded video clip also contains a text transcript. Nevertheless, it is challenging to give readers of a journal article access to volumetric performance capture ‘data’, especially given its volumetric form and its re-enactability.
- <sup>3</sup> Video as a method has become central in qualitative research more broadly (see, for example, Bates, 2015; Harris, 2016; Haw and Hadfield, 2011; Knoblauch et al., 2015).
- <sup>4</sup> Cf. the ‘continuous shot’ – in a ‘frame’ seen through the ‘viewfinder’ – as a practical accomplishment (Macbeth, 1999).
- <sup>5</sup> Coulter and Parsons (1990) and Jayyusi (1988)’s otherwise excellent elaborations of the praxeology of perception and scenic intelligibility are largely silent about the scenic intelligibility of the life-world in terms of sound.
- <sup>6</sup> One could also argue that ‘transplane sounds’, e.g. ambisonic, are not materially three-dimensional but they provide more information on space or the spatial structures of sounds than sounds projected in mono or stereo formats.
- <sup>7</sup> Laurier and Reeves (2014) analyse how multiple virtual game-cameras in 3D video games on flat screens are assembled in practical game play as an interactional resource for the game player(s). The footage of actual game play (and the players) was, however, only recorded with 2D video cameras and 2D screen capture.
- <sup>8</sup> Some scholars have explored motion capture to advance qualitative research in an experimental CA tradition, e.g. Stevanovic et al. (2017) on ‘body sway’ correlated to physical measurements from motion capture suits, but the resulting data is abstracted and

graphically presented in a quantitative fashion. Capture data is *not* replayed to ‘see’ the phenomenon in the lived body.

<sup>9</sup> *AVA360VR* was developed by Nicklas Haagh Christensen, Stefan Tanderup, Artúr Barnabas Kovács, Jacob Davidsen and Paul McIlvenny in the BigSoftVideo team at Aalborg University (McIlvenny, 2019b). *SQUIVE* was developed by Paul McIlvenny.

<sup>10</sup> *Inferno* is designed, choreographed, composed and realized by Louis-Philippe Demers and Bill Vorn (<http://prix2016.aec.at/prixwinner/19611/>). The data shown here was collected by the research group EXACT (Exoskeletons, Art, Choreography and Training) (Jochum et al., 2018). Consent was granted by all participants for the full recordings to be used for research purposes.

<sup>11</sup> <https://youtu.be/YRkrWKf7Imw>

<sup>12</sup> *Mindshow* has been available on the *STEAM* gaming store (<https://store.steampowered.com/app/382000/Mindshow/>) since August 2017 in a free ‘early access’ version. The recordings for this paper were made in August 2019.

<sup>13</sup> The *Mindshow* source code is 3GB and the captured assets necessary for replaying this example are only 84KB.

<sup>14</sup> The number sixteen is arbitrary. It was chosen simply because a grid of four-by-four viewports would fit snugly in one rectangular video frame. One could ask how many cameras are needed to capture a dynamic interactional scene so that it is fully volumetric? In VR, the answer is *none*, because it can be captured volumetrically *without* cameras. Moreover, if one can re-enact the scene and re-view it from an infinite number of unique virtual camera trajectories, then there is no need to predetermine how many is enough for analytical purposes.

<sup>15</sup> Arguably, although I have maintained that the future of video is *not* video per se, the practical implementation of VR technology at present is driven by the miniaturisation of high resolution, small screens, one for each eye, so one can feel perceptually immersed in the visual and volumetric presentation of the virtual world as those screens get very close to the eyes in an HMD. Ultimately, and ironically, VR is still flat screen-based. Nevertheless, this is besides the main point, namely that the screens are *not* reproducing

the view of a fixed point-of-view camera in the original scene; instead, they are able to provide a dynamic (stereoscopic) view *from anywhere, simultaneously*.

<sup>16</sup> Some views are hard to fit into the typology, e.g. 7 and 8 were taken while lying on the floor looking up. In addition, all the views are, arguably, Infra-Subjective because they were recorded from the screen capture of the ‘subjective’ viewport of the author’s HMD.

<sup>17</sup> <https://youtu.be/qcMJrOarpSI>

<sup>18</sup> *VREAL* was available as a free ‘early access’ release on the *STEAM* store (<https://store.steampowered.com/app/440120/Vreal/>) from June 2018 to July 2019. Since then the company has shut down its operations, so *VREAL*, and its archive of game play captures, is no longer publicly available. The recordings for this paper were made in July 2019. Even though this game capture was publicly available, consent was given by the employee behind this avatar for the recordings to be used for research purposes.

<sup>19</sup> The *VREAL* source code is roughly 4.7GB and the capture assets necessary for replaying the scene in this example are 4.5GB.

<sup>20</sup> <https://youtu.be/GkJFCqFuNcs>

<sup>21</sup> The approach proposed in this paper raises deeper ethical issues that are not considered further here. These include how to gain *informed* consent *before* undertaking volumetric performance capture, and how to deal with the *inadvertent* or *inappropriate* capture of conduct that is ancillary to the main purpose of the data collection.