6

Analyzing Videos: Selecting an Approach and Annotating **Gunter Lösel**

- 6.1** Setting the Course for Your Own Project
- 6.2** Analyzing with Videos and Annotations

In the field of artistic research, there is such a wide array of research traditions to choose from that no single approach can be assumed a priori as the most appropriate one. At the outset of a research project using video and annotations, how can researchers find their bearings within the landscape of investigative methodologies without constraining themselves to a strictly defined approach? This chapter illustrates different ways of flagging, arranging, organizing, analyzing, and annotating relevant segments, even when dealing with large amounts of video material.

Analyzing videos is an undertaking that is never entirely free of preconceived notions, and the first step for researchers is to find their bearings within the field of possible research approaches and thus reflect on their own basic assumptions. In 2018 René Tuma presented an overview of the leading research traditions that are of importance within the context of video analysis.

This classification (→ fig. 01) illustrates the historical evolution of these approaches; that is, it examines how they are embedded in existing theories. In keeping with the objectives of our project, we have sought to overcome these disciplinary constraints.

6.1 Setting the Course for Your Own Project

Is my research about manifest or latent patterns? The approaches can be classified according to whether they tend to focus on the surface or the deep structure of the data. Gubrium and Holstein describe this as the interaction between the “whats”—the concrete, describable content—and the “hows”—the contexts that are not directly observable and hence need to be explored (Gubrium and Holstein 1997). Accordingly, the different approaches can also be classified according to whether they primarily examine the “what” or the “how” of the material.

This decision (→ fig. 02) determines, among other things, which and how much data is to be collected. Whereas the search for manifest patterns usually requires a broad database, the focus on latent patterns is usually confined to small yet well-selected sets of data. So, by deciding between “manifest and latent”, you also determine the extent to which your research is broad (breadth) or narrow (depth).

For instance, artistic research often investigates processes that are not conscious and declarative. Instead, it is “tacit knowledge” that must first be rendered observable and describable through specific approaches. Artistic research, therefore, often addresses latent structures (however, unlike other hermeneutic methods, this does not involve the reconstruction of latent structures of meaning, but rather a wealth of intuitive, corporeal, material-related, and other approaches that “somehow” lead to an artistically relevant output). The associated research aims to reveal these invisible structures and incorporated practices. In other instances, though, artistic research also deals with the explicit, declarative knowledge of the artists, which can be explored

01

02

Fig. 01:
Video analysis method proposed by
René Tuma (Source: Tuma 2018, p. 426)

Fig. 02:
“What” and “How” of the analysis
(illustration by author)

in interviews or video-cued recall situations. In such circumstances, it seeks manifest content or “declarative knowledge”.

What Elements of the Video Material Should Be Examined?

A radical response would be: everything that has been recorded and as much as possible should be recorded. This is the approach taken by the visual grounded theory (Mey and Dietrich 2016). The documentary method also applies to the recorded material in its entirety (Burkhard 2018). Here it is a matter of sifting through vast amounts of data because the sought-after patterns (such as the individual style of the author) are not found here and there, but rather throughout the material.

As a rule, though, a selection is made, that is, only selected key sequences are examined. This applies to all hermeneutic methods. Selecting these segments, commonly known as sampling, thus becomes a critical process, which must be well justified from a methodological standpoint.

In extreme cases, only tiny fragments of the video are used (perhaps only individual stills). This applies, for example, to methods derived from image analysis, such as figurative hermeneutics (Eisewicht et al. 2018), in which it is perfectly feasible to work with individual images. But even objective hermeneutics, as envisioned by Ulrich Oevermann, can make do with very little data (Oevermann et al. 1979).

What Context Should I Use for Interpretation?

Researchers can approach an issue or a phenomenon with an abundance of previous knowledge or they can attempt to comprehend it on their own accord. Different research approaches can come into play here:

No context. The manifestation is examined under the premise of being devoid of any context (such as in objective hermeneutics). This makes it possible to approach the sample without bias, and it leads to unconventional interpretations. Consequently, the context has to be “hidden,” as if only one part of an image were examined by covering the rest with paper.

Only context that is accessible to the respective players in the field at a specific time. The goal here is to reconstruct this knowledge from the perspective of the players, not from the

bird's-eye view of the researchers. Accordingly, it is not permitted to use information presented later in the video to interpret earlier states, as this information exceeds the knowledge of the players at the given time. Such an approach corresponds, for instance, to the basic assumptions of ethnomethodology.

All case-specific information, that is, the contexts created by the video itself. For example, phenomena in the video can be explained by information that appears later in the material.

All case-specific information plus any details that are directly related to the creation of the video, such as news announcements, interviews, production materials like screenplays, etc. The context is extended here to allow for a multiperspective analysis.

The entire world knowledge of the researcher (for example, figural hermeneutics, grounded theory).

The entire world knowledge of the individual researcher plus the world knowledge of a group of researchers.

Videography, for example, relies on this approach (Tuma 2018).

The researcher(s) should select the interpretive framework that corresponds to the respective goals of the concrete project.

How Is the Action Behind the Camera Analyzed?

Depending on the approach used, different types of video constitute the basis of the analysis. Videos that are found online or that become available in another form as finished products raise completely different questions than those that are created under the watchful eye of a researcher specifically for the purpose of conducting a study. With previously existing videos, in addition to the content, or the on-camera action, the action behind the camera constitutes a focal point of the analysis because it goes beyond the physical recording of images and encompasses the entire range of postproduction steps, including all decisions on camera angles, framing, light, depth of field, editing, effects, etc.

It is always necessary to establish to what extent the action behind the camera is part of the analysis, and hence to identify, transcribe, and analyze its elements accordingly. With videos produced specifically for research purposes, the action behind the camera reflects the intention of generating knowledge and is thus an integral part of the method. It must be assessed to determine its suitability for achieving the goal of the research and rendered transparent. Depending on the research objective, one or the

other aspect can be given more weight, but both levels should always be included in the analysis (Reichertz 2018).

How Important Is Selective and Subjective Perception?

In striving to achieve the ideal of an “objective camera”, researchers aim to minimize the degree to which they preselect material and seek to capture a wide range of information, maintaining a broad “data focus”. Ideally, the camera endeavors to capture “simply everything”, which of course can never be entirely achieved.

03

Fig. 03:
Action behind the camera as part of
the analysis (illustration by author)

By erecting a 360-degree camera, researchers admittedly come closer to this ideal than was previously possible, but here, too, preliminary decisions are made, for example, on where to place the camera. Instead of an “objective camera”, the work proceeds with a “subjective camera”, which requires that clear decisions be made at the time of shooting, decisions that, out of artistic and analytical considerations, often have to be made very quickly and intuitively. The video data is then, of course, highly selective. Since the assumption of objectivity is subsequently no longer tenable, the active, subjective action behind the camera must be included as part of the analysis. We can conceptualize the different types of audiovisual data as a continuum:

Subjective $\longleftrightarrow$ Objective

Varying degrees of subjectivity require different methods of analysis. Researchers need to make such decisions “before they start filming”.

Top-Down or Bottom-Up?

All research approaches are based on a common understanding that new knowledge is always engendered by the interaction of theoretical concepts and newly acquired data. But there are differences in the sequencing and weighting: What comes first? What is more important?

Although the concepts may vary, a distinction can be made between top-down approaches, which derive practical applications from theoretical concepts that are then tested against the data, and bottom-up approaches, which examine the data with a minimum of preconceived notions in a bid to identify emerging patterns. Most methods alternate between bottom-up and top-down phases (older, related terms for these phases are inductive and deductive).

Top-down approach (→ fig. 04): The theoretical concept, here displayed as a yellow circle with black spot, is used to establish categories that can then be used to detect the corresponding phenomena in the data or, if the phenomena cannot be found, to call into question the concept.

Bottom-up approach (→ fig. 05): The data (observations) come first and constitute the outset. By applying epistemic practices, patterns become recognizable; they “emerge,” or rise to the sur-

face, and are only then theoretically categorized and put into words.

6.2 Analyzing with Videos and Annotations

Once the researchers' own position and methods have been clarified, the data is collected, processed accordingly, and subsequently presented in a form that fits the respective research design. Annotations can be added as live annotations when the video data is recorded (for example, with the Piecemaker2Go tool) and, in any number of steps of analysis, can later be supplemented with more annotations, resulting in what anthropologist Clifford Geertz calls a "thick description" (Geertz 2017). According to our model in the introduction, this corresponds to "enriched data."

Already when formulating annotations, preliminary findings will emerge, which in turn are noted down and provide the material for further steps of analysis. In this respect, the transition from data collection to data analysis is often seamless. In the following sections, we will describe the individual stages as distinct from one another, which, in reality, is rarely the case.

Description of the Data

The first step of data analysis is always a summary description of the data. What types of data are actually available? What is their scope and quality? Quantitative approaches rely on "descriptive statistics" for this purpose, whereas qualitative approaches involve laying out the complete data material and compiling a textual description of the data types, for example, as a table, field notes, or a logbook. The diverse video data is labeled and subjected to an initial review.

Jürgen Raab and Marija Stanisavljevic recommend sifting through all the material first without taking notes. They say that it is only after this overall impression has been achieved that notes should be taken during a second review (Raab and Stanisavljevic 2018). Other methods recommend a logbook or a rough classification in the form of a list or a table. When working with videos, a good filing system for video files and a good file identification system is absolutely essential for screening the material. Descriptions should also include a critical evaluation of the data.

04

05

Fig. 04:
Top-down approach
(illustration by author)

Fig. 05:
Bottom-up approach
(illustration by author)

Creating an overview also compels you to complete the data collection phase at some point. A research project is not an endless journey, but rather a process that moves toward an end, which is why it is important to complete the transition from one phase before proceeding to the next one.

Defining the Unit of Analysis

What is the smallest unit to be examined, that is, the “unit of analysis”? It is usually possible to derive the answer from the theory and the context of the study. How big are the phenomena under study? If we intend to catch small fish, we need a close-meshed net, but if we are looking to catch big fish, we use a wide-meshed net. And if we don’t know what kind of fish we want to catch, we should take samples and decide afterward. For instance, microsociological studies require completely different units of analysis than sociological discourse analyses. Ramey et al. (2016) point out that the units of analysis can change during the research process and argue in favor of a flexible, iterative approach. When conducting video analyses, the unit of investigation can vary greatly depending on the design of the research.

Unit of Investigation for Video Analyses

Very fine	The smallest analyzable units are stills of the video, that is, single images or frames. This is the unit of choice for image-interpreting procedures such as figurative hermeneutics, which are derived from the analysis of still images.
Fine	The units of analysis here are moves, that is, the smallest of meaningful movements. Moves are roughly comparable to words: you can interpret them individually, because each word carries a meaning, but they are all merely elements of an interaction. One example might be a gesture made during an utterance, or an emotional expression that accompanies one. Such phenomena can also occur subliminally in the range of microseconds. In such instances, they only become apparent when watching the video in slow-motion mode.
Middle	Here the unit of analysis is a single interaction element or turn, in which the participants themselves mark the beginning and end of a turn with metapragmatic signals (Goffman 1974). The analysis avoids an “artificial” segmentation and instead follows the “natural” course of events (interaction analysis and conversation analysis).
At the scale of the individual	Here, the focus is on the individual components of action that reflect the strategy, style, or habitus of the respective participant in the interaction. The units of analysis here do not relate to a particular interaction, but rather to a specific individual, and hence they are scattered throughout the material wherever an expression appears “typical” of that person. Such units can be described in terms of their intentionality or functionality, for example, an interaction where A aims to instruct B in a certain way, which can be brought to light in the analysis. A similar idea is the “habitus” concept proposed by Pierre Bourdieu.
Societal	This pertains to social systems and discourses and comprise sequences of communications and actions in the public sphere that make reference to one another but do not necessarily constitute part of the immediate interaction, for example, diverse critiques of a theater production at different times and in different places. The units of analysis in such discourse and viscourse analyses can be of any size and may comprise huge amounts of data.

By displaying the size of the annotation on the track in Research Video, the software provides a direct visualization of the unit of analysis. This makes the tool well suited for determining units of analysis. They are reflected in the size of the annotations on the tracks. In a use case, for example, the acts and scenes of a feature film were dissected using annotations (→ fig. 6).

How Do I Find Relevant Passages in the Material (Sampling)?

After the data has been described, it is usually necessary to identify smaller sequences that are relevant to the analysis. This process, known as sampling, becomes increasingly important as more and more material is selected. There are several recognized methods.

Methods of Sampling

Theoretical sampling	The selection is closely related to the research thesis, and the researcher only selects sequences that are related to the thesis based on theoretical considerations.
Whole-to-part method	Multiple viewers look at the entire data corpus and endeavor to reach a consensus on the relevant passages (Erickson 2006).
Representative sampling	With this option, a “typical” sample is selected from a collection of similar data. The selection is validated intersubjectively by consulting experts.
Selection based on “narrative power”	Derry et al. regard this form of selection as legitimate, in which clips are selected that support the narrative of the research. The story is told with a “disciplined subjectivity” that does not claim to reflect objective perspectives, but is not purely fictional either.
Iterative sampling	Derry et al. do not see sampling as a self-contained step. Instead, they understand it as an iterative process that occurs repeatedly throughout the course of the research, so that one sampling can lead to another sampling.

Raab and Stanisavljevic (2018) propose the following procedure:

- View all material, without taking notes
- Second review with notes
- Select a passage that could be relevant to the research question
- Conduct a detailed analysis of the key segment
- Select additional segments that display minimum contrast
- Select additional segments that display maximum contrast

This method for the formation of minimum/maximum contrasts is described in greater detail in chapter 8.

Indeed, sampling is one of the key practices of qualitative research, even if the selection is by no means as methodologically precise as is occasionally claimed. In the widespread “snow-ball system,” for example, segments are identified simply by the fact that several viewers agree that they are “somehow interesting.” Consequently, selecting the relevant sequences remains to a certain extent subjective and intuitive, no matter how methodically the researchers have worked.

Research Video offers excellent options to split a video into sequences and visualize them. By having the same material sampled by different researchers, a certain intersubjective validation can be quickly and clearly achieved. This makes the tool a sound basis for sampling, regardless of which method is used.

Transcription

The process of analyzing the data also entails transcription as a means of describing observable phenomena. All research approaches include transcription as a necessity element, and it has been pointed out many times that all transcriptions contain preliminary decisions that significantly influence the subsequent course of the research: “To create a transcript is to make consequential choices about which phenomena merit representation” (Ramey 2016, p. 2).

The transcription thus generates a “conceptual lens”, or selective perception, which must be well thought out and justified because it determines what the researchers perceive in the first place and what they might overlook. In addition to language, a wide range of modalities of conduct have been transcribed into videos: gestures, glances, body postures, facial expressions, movement, material objects, etc.

Each research tradition and research thesis determines how fine and how multimodal a transcription should be, for example, whether it should include unconscious or subliminal behavior like throat clearing, pauses, swallowing, hesitation sounds, etc. The transcription also determines how the action behind the camera is to be incorporated into the research.

There are systems and tools for transcription in the respective research traditions, which can be classified as follows (Moritz in Moritz and Corsten 2018, p. 14): “Tools like HIAT, ELAN, and EXMARaLDA are suitable for research that focuses more on spoken language. Field Score, MoviScript, and TraVis are all suitable tools for social scientific analysis. Other software such as Akira, ATLAS.ti, MAXQDA and NVivo bypass transcription entirely and deal with specific methods of evaluation, for instance, within the context of qualitative content analysis.” An overview of the different available tools is provided by Moritz (2014).

Annotated video software like Research Video offers outstanding transcription options in which the respective tracks can be freely filled with content as desired. Both linguistic and non-verbal cues can be entered, in addition to aspects of the action behind the camera. Researchers can also expand these transcription modalities during the analysis.

Data Analysis: Pattern Recognition

Knowledge manifests itself in patterns and contexts. The epistemic (cognition-enabling) function of a video becomes clear when it presents opportunities to see data in a new way, to rearrange it, filter it differently, and recognize patterns that would otherwise remain hidden to the naked eye. The analysis begins with concrete perception and gradually progresses to more complex processing, just as the eye processes perceptual data first, followed by the visual cortex and then other parts of the brain. In artistic research, this is not so much a process of verification or falsification, but rather one of perceptual amplification.

The recognized patterns do not necessarily have to be causal patterns—although these are preferred in experimental research—but may just as well consist of phenomenological patterns, that is, those that allow us to perceive a context differently. You can imagine it as analogous to the edge detection ability of a computer, which allows it to clearly distinguish different surfaces by emphasizing transitions. The computer compares points of

perception with their immediate surroundings and calculates the degree of similarity. This method enables machines to perceive contours and recognize objects. The human brain has a similar mechanism that enhances contrasts where an area ends, so that objects stand out sharply from the background (shape perception).

Similarly, the analysis of video material using the formation of minimum/maximum contrasts can precisely pinpoint such differences and trace their exact progression. This process can also be compared to the use of contrasting agents in biological and medical research. Based on certain practices, the data is “colored” in such a way that patterns are rendered visible (→ fig. 7).

By using tracks as observation categories, such patterns can be immediately made visible in Research Video. If you hide individual tracks, these patterns become even more visible. This not only makes them more vivid, but also makes them comprehensible to other researchers who would like to retrace the process from screening the original video material to drawing the final conclusions. Consequently, the interpretation remains closely bound with the material and can be called into question by other viewers. Do they see the same phenomenon as the researchers? The analysis thus remains available to be challenged and intersubjectively validated.

Recommended Reading

Derry et al. (2010) have developed a comprehensive set of guidelines. They focus on four problems: selection, analysis, technology, and ethics. Another useful source is the paper by Ramey et al. (2016) for a symposium on standards and heuristics of video analysis, which formulates the rules of thumb for tackling key problems.

06

07

Fig. 06:
Sequencing with tracks and annotations

Fig. 07:
Sample project: annotations illustrate the rhythm of the audience's reactions.

The recognized patterns do not include causal patterns—although they do include experimental research—but not phenomenological patterns, thus leading us to perceive a context difference.

not necessarily have to be
these are preferred in
may just as well consist of
that is, those that allow
rently.