From Detectables to Inspectables: Understanding Qualitative Analysis of Audiovisual Data

Krishna Subramanian*
RWTH Aachen University
Aachen, Germany
krishna@cs.rwth-aachen.de

Johannes Maas*
RWTH Aachen University
Aachen, Germany
johannes.maas1@rwth-aachen.de

Jan Borchers
RWTH Aachen University
Aachen, Germany
borchers@cs.rwth-aachen.de

James Hollan
University of California, San Diego
La Jolla, California, USA
hollan@ucsd.edu

Figure 1: Qualitative researchers work with audiovisual recordings to develop insights. The figure shows a video recording of a study participant using her computer. As a researcher, we want to understand why the participant gets distracted during work. To do this, we look for audiovisual segments where the participant is talking to a colleague—these are the *inspectables* (marked with yellow brackets). To find these, we need to look for instances where the participant looks away from the computer to face their colleague, which are the *detectables* (marked with green crosses).

ABSTRACT

Audiovisual recordings of user studies and interviews provide important data in qualitative HCI research. Even when a textual transcription is available, researchers frequently turn to these recordings due to their rich information content. However, the temporal, unstructured nature of audiovisual recordings makes them less efficient to work with than text. Through interviews and a survey, we explored how HCI researchers work with audiovisual recordings. We investigated researchers' transcription and annotation practice, their overall analysis workflow, and the prevalence of direct analysis of audiovisual recordings. We found that a key task was locating and analyzing *inspectables,* interesting segments in recordings. Since locating inspectables can be time consuming, participants look for *detectables,* visual or auditory cues that indicate the presence of an inspectable. Based on our findings, we discuss the potential for automation in locating detectables in qualitative audiovisual analysis.

*Both authors contributed equally to this research.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; *User studies*; • **Information systems** → *Video search.*

KEYWORDS

Video analysis, qualitative research, researchers, information behavior

1 INTRODUCTION

Qualitative research methods are often employed when data cannot be quantified. They are commonly used in human-computer interaction (HCI) [1] and social science [18, 35] research. In HCI, qualitative methods are becoming increasingly prevalent. While 44% of papers from CHI 2014 used qualitative analysis methods [5], a survey of CHI 2019 papers [34] shows that this has increased to 62%. With qualitative research paradigms and software tools being employed more frequently, we need to understand *how* these research practices and accompanying tools are being used.

Coding (also called *annotation*) is an important component of qualitative analysis. Researchers use keywords to categorize data and subsequently use these categories to detect patterns and anomalies. Researchers often code textual transcripts, e.g., of interviews, instead of the audiovisual recordings themselves. An important reason for this is that text is easier to search and navigate than audiovisual recordings [20, 25, 32]. However, audiovisual recordings contain additional rich information about behavior and user interaction not available in text transcripts [12]. Further, creating transcripts is tedious, and researchers often avoid doing it themselves, e.g., by paying student assistants or professionals to create them [28]. Because of these reasons, researchers often need to work directly with the audiovisual recordings. Despite the increasing frequency and importance of qualitative analyses of text-based transcripts and audiovisual recordings, there is a lack of research aimed at understanding researchers' needs and workflows.

To address this, we conducted interviews and a survey to help understand researchers' workflows when qualitatively analyzing audiovisual recordings. Specifically, we addressed the following questions: *How common is direct analysis of audiovisual recordings? Are transcripts involved when audiovisual recordings are available? How do researchers analyze audiovisual recordings? What do they look for? Where do they look?*

As we will detail below, we found that researchers employ a predominantly bottom-up coding approach, in which they determine 'what to look for' only after watching the recordings, often multiple times. Significant time and effort during analysis is dedicated to locating and analyzing the segments in the recordings that researchers deem *interesting*. To locate these interesting segments, researchers search for visual and auditory cues, such as a button click or a particular sound, in the recordings that *signal* the presence of an interesting segment. Towards the end of the analysis, researchers seek information in the recordings, such as interesting user quotations and screenshots, with the purpose of reporting them in a publication or report. We discuss these and other findings,

and present some design recommendations to improve researchers' analysis workflows.

The novel contributions of this paper are:

- Results from interviews and a survey that help to provide an understanding of the prevalence of direct audiovisual analysis in qualitative HCI research, and researchers' corresponding transcription, coding, and navigation practices. Our contribution here includes the vocabulary and concepts of *inspectables, detectables,* and *reportables* which help understand and structure discussions about qualitative audiovisual analysis.
- A discussion of how automation could help researchers deal with the important but laborious task of locating interesting audiovisual segments in recordings, and design recommendations based on our findings.

2 BACKGROUND AND RELATED WORK

In this section, we present some background information about qualitative analysis approaches and researchers' analysis workflows. We then provide an overview of previous research-based and commercial tools that support qualitative analysis of audiovisual data, before discussing existing techniques for audiovisual navigation, summary, and visualization.

2.1 Approaches to qualitative analysis

Common qualitative research methods include grounded-theory based qualitative coding [36] and affinity diagramming [2]. In both methods, researchers identify interesting text or audiovisual snippets, categorize them, and use these categories to detect patterns or trends. While these methods dictate the researchers' overall approach to analysis, individual research fields tend to follow more specific analysis methods. For example, in interaction analysis in HCI, the focus is to understand how humans interact with other humans, objects, and interfaces [22].

2.2 Researchers' analysis workflow

Previous research has investigated how researchers perform qualitative analysis. Melgar et al. proposed a generic process model that describes the various steps in media analysis by researchers [30]; this work is grounded in earlier research on media analysis workflows. The resulting model includes four phases that help us understand researchers' workflows: *exploration* (background research, develop research intent, gather data), *assembly* (select corpus, collect more data if needed, select part of the corpus for analysis), *analysis* (pre-analyze corpus, conduct exploratory data analysis via visualizations, annotate data), and *presentation* (organize data and collect evidence for presentation). Although this model provides an overview of various tasks in media analysis by researchers, it does not investigate the mechanics of interaction employed in these tasks.

Our paper is motivated in particular by Marathe and Toyama's CHI '18 paper that sought to understand how qualitative researchers interact with current software tools [27]. They found that the granularity and type of codes vary across different disciplines, and investigated opportunities to automate certain tasks in qualitative

analysis, e.g., with codebook development and management. Their paper focused on text-based qualitative analysis. We investigate direct analysis of audiovisual recordings.

2.3 Software support for qualitative analysis of audiovisual data

Software applications used for qualitative analysis are referred to as Qualitative Data Analysis (QDA) tools. QDA tools benefit researchers with organization of media files and powerful features to manage codebooks. While there is concern that QDA tools impact the analysis method [13], the core of this concern has largely been refuted [3, 38, 41].

To understand how effective QDA tools are in supporting qualitative analysis, a comprehensive study of QDA tools was conducted as a part of the KWALON experiment [12]. A multimedia data corpus was analyzed by different teams utilizing various mainstream QDA tools. The results showed that due to the wide variety of tools, researchers would have to assess whether a specific tool will suit their workflow [31].

The most widely used commercial QDA tools are ATLAS.ti,[1] NVIVO,[2] MAXQDA,[3] Transana,[4] and HyperRESEARCH.[5] Transana is the only listed tool that prioritizes audiovisual data over text [12]. There are some research-based QDA tools that support audiovisual analysis, such as ELAN (EUDICO Linguistic Annotator) [40], INTERACT,[6] and Observer.[7]

Some research tools tackle specific problems in audiovisual analysis. ChronoViz allows users to view and annotate different types of time-coded data streams, such as videos, sensor data, and field notes, to support more comprehensive analysis [14]. MiMeG[8] (Mixed Media Grid) and DRS [4] help with collaboration and organization of multimedia files respectively. The goal of our research is to understand how HCI researchers use these commercial and research-based tools, in order to identify potential avenues for improvement.

2.4 Techniques to navigate, summarize, and visualize audiovisual content

The finding that navigating time-coded media streams is more tedious than navigating textual data is well established in research [20, 25, 32]. This has spurred several techniques that aim to improve video navigation, e.g., by allowing multiple perspectives of videos [10], media navigation at different levels of granularity [6, 9, 33], and video navigation through direct manipulation of objects in the video [11, 17, 23].

There have also been efforts to summarize audiovisual content in order to reduce viewing or listening time. For audio, this includes eliminating silence or speeding up clips [37] and summarizing spoken text [15]. For video, techniques include providing video thumbnails [24], video skims [8], and keyframe summaries

[16]. In addition to summaries, salient information in audiovisual recordings can be visualized to support 'micro-analysis' [29], i.e., detection of patterns or structure in the recordings. For example, DIVA visualizes annotations in a pseudo-3D representation to help researchers detect patterns and relationships more easily [26]. In music research there are many different visualizations used for analyzing music [21]. Simpler visualizations, such as waveform views and timelines, are more commonplace in QDA tools.

3 HOW RESEARCHERS ANALYZE AUDIOVISUAL RECORDINGS

In this section, we present findings from interviews with researchers who analyze audiovisual recordings using qualitative research methods. We conducted a survey to validate our interview findings, the results of which are discussed in the next section.

3.1 Data collection and method

To recruit participants for our interviews, we reached out to experienced HCI researchers who had conducted audiovisual analyses as well as qualitative researchers we knew personally or through connections. We also recruited participants through mailing lists and by emailing psychology researchers at our university. We recruited ten participants (IP1–IP10, see Table 1) for the interview.

The first three interviews (IP1–IP3) were exploratory, designed to give us a general understanding of their workflows and where the most acute problems lie. These exploratory interviews helped us identify research questions[9] that guided the next five interviews (IP4–IP8). In all interviews, after inquiring about each participant's background and expertise, we had them discuss their intentions during their most recent analysis and walk us through their data collection and analysis. We encouraged participants to keep their recent analysis files and analysis tool ready so that they could walk us through their work. This simultaneously provided us with better understanding of their analysis and gave participants the opportunity to reflect on their analysis process. Then we probed for specific things our participants looked for and discussed how they looked for each. Analyzing data from these eight participants revealed patterns that were strong enough to warrant triangulation via a survey (see Section 4). After the survey, since we wanted to gather specific examples of what researchers look for in audiovisual recordings, we recruited two more participants (IP9 and IP10) for these focused interviews.

Our participants included two Master's students, six PhD students, and two professors; they had varying amounts of experience with qualitative analyses of audiovisual recordings. All participants conducted research in the field of HCI, except for IP6 who worked in psychology.

Due to COVID-19, all interviews were conducted remotely using Jitsi[10] and Skype[11]. We recorded the session audio with the participants' consent, and took notes to inform our analysis.

The second author transcribed the interviews, and identified interesting text snippets and insights from the transcripts and study

---

[1]https://atlasti.com/

[2]https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

[3]https://www.maxqda.com/

[4]https://www.transana.com/

[5]http://www.researchware.com/

[6]http://www.noldus.com

[7]https://www.mangold-international.com/

[8]https://www.surrey.ac.uk/sites/default/files/MiMeG_distinguishingfeaturesFINAL.pdf

[9]Our principal research questions were "What do researchers look for in audiovisual recordings?" and "How do they look for it?"

[10]https://meet.jit.si/

[11]https://www.skype.com/

ID	Field [a]	Status [b]	Duration [c] (minutes)	Research intent [d]	General approach [e]	Tools used
IP1	HCI	PhD	71	Understand UI usage	Exploratory, direct annotation	ELAN
IP2	HCI	Prof	41	Understand UI usage	Exploratory, partial transcription	ELAN
IP3	HCI	PhD	60	Understand UI usage	Exploratory, partial transcription	ELAN
IP4	HCI	MSc	43	Validate a prototype	Rough scheme, tabular notes	Spreadsheet software
IP5	HCI	PhD	28	Find influence of an intervention on a task	Exploratory with POIs, direct annotation	MAXQDA
IP6	Psy	PhD	29	Understand people's relationship with an object	Exploratory, full transcription	NVIVO
IP7	HCI	MSc	30	Validate a prototype	Rough scheme, tabular notes	Notepad, MS Excel
IP8	HCI	Prof	31	Understand usage of a teaching medium	Exploratory, full transcription, direct annotation, and clips	Video player, ELAN
IP9	HCI	PhD	17	Assess a UI	Exploratory with POIs, freeform notes	Paper-based notebook, video player
IP10	HCI	PhD	64	Find usability issues in UI	Exploratory with POIs, full transcription	NVIVO

[a] HCI: Human-Computer Interaction; Psy: Psychology.
[b] PhD: PhD student; Prof: Professor; MSc: Master's student.
[c] Duration of the recordings. Does not include the time taken for the initial introductions and final debriefing/feedback from participants.
[d] We deliberately provide abstract descriptions of participants' research intent to retain their anonymity. For some concrete examples, see Section 3.2.1.
[e] Exploratory with POIs: exploratory, but with points of interest they were looking for; rough scheme: started with a rough scheme that then evolved; tabular notes: noted observations in a table; clips: extracted clips for detailed inspection.

Table 1: Information about interview participants, including their research intent, analysis approach in a recent analysis, and analysis tools.

notes. Some interviews were not conducted in English; the corresponding verbatim quotes reported in this paper are translations. The first and second authors organized the snippets and insights in the style of an affinity diagram[12] [19], in which the categories were developed from the data.

We now present the main findings from our analysis of the interviews organized into three subsections. In Section 3.2, we discuss participants' overall research intent, and how this affected whether they analyze mainly transcripts or audiovisual recordings. In Section 3.3 and Section 3.4, we discuss what information participants looked for in audiovisual recordings, the navigation methods they used to look for this information, and introduce inspectables, detectables, and reportables. We summarize the key findings in Section 3.5.

3.2 How researchers prepare to work with audiovisual recordings

3.2.1 Exploratory workflow and opportunistic navigation. Our participants reported beginning their analysis with broad research questions. Several research questions were *theory-building,* e.g., understanding how users perform a complex task using various programming IDEs, understanding how users react to sensory interventions at regular intervals, or investigating how users use certain

tools for learning. Other research questions were about *validating* a prototype or software application, e.g., by understanding what insights users of a tool would develop, by checking if users could understand the interaction offered by the tool, or by determining the effect a prototype has on learning outcomes.

These research questions served as a starting point for the qualitative analysis. Participants reported to know what events to look for only at an abstract level; none reported using a predefined codebook or annotation scheme before watching or listening to the recordings. This is similar to the *open coding* phase in text-based qualitative analysis, in which researchers determine an initial set of codes only after getting acquainted with the data. To get acquainted with the recordings, participants reported employing an *opportunistic* style of navigation, such as playing videos at a faster speed (IP2, IP5, and IP7), listening to the discourse in the background (IP5), and scrubbing the video to identify salient video segments (IP7). Also, several participants (IP1, IP2, IP3, IP6, and IP8) explicitly mentioned that they approached the analysis with an open mind, trying to *"pursue what the data says"* (IP6), which indicates the prevalence of an inductive analysis approach. The evolution of their codebooks or annotation schemes continues throughout the analysis (IP1, IP4, and IP6), sometimes over multiple iterations (IP6).

3.2.2 What gets annotated—transcripts vs. audiovisual recordings. Our participants' research questions dictated their transcription practice as well as what gets annotated. Transcripts are textual descriptions that capture information in audiovisual recordings.

[12] An overview of the categories that resulted from our affinity diagram is available in the supplements.

While certain kinds of information, such as the verbal discourse and observable user actions, require little to no interpretation, others, such as body language and other non-verbal cues, require more nuanced interpretation and are difficult to describe adequately in transcripts. We found that participants either annotated the audiovisual recordings directly, or they analyzed mainly textual transcripts, using the audiovisual recordings for clarification.

IP1 and IP8 did most of the analysis working directly with the audiovisual recordings. IP8 extracted about 20 video clips from their recordings, so that these clips could be played in a loop via a conventional video player for a more focused analysis. Many participants (IP2, IP3, IP4, IP5, and IP7) performed what we classify as partial transcription of the recordings, and mainly used these transcripts for analysis. However, they also relied upon recordings during the analysis: IP2 and IP3 viewed certain video segments after transcription to observe details, such as a user's emotions, that were missing from the transcript (*"[The] participant did not say anything while they [were] struggling. [...] if we watch the video, we will be able to note down that [...] the participant [...] expressed some frustrations."* – IP2). IP4, IP5, and IP7 did not transcribe the verbal discourse in a conventional sense, but took notes on specific events while watching the videos that were used in the analysis. IP6 prepared full transcripts and used them exclusively during the analysis.

Transcripts are not always limited to verbal information. For example, IP3 included what component of the interface the user was using into the transcript for context (*"[...] we had to see, OK, so he is moving the slider, and then write in brackets behind that that it was the [bottom] slider."* – IP3).

3.3 What researchers look for in audiovisual recordings

Our findings indicate that most participants either worked with audiovisual recordings directly or used them as a supplement when analyzing partial transcripts, e.g., to watch salient events in detail. In this section, we present findings about participants' annotation practice, which typically begins after getting sufficiently acquainted with the videos to determine an initial set of codes or events to look for.

There are broadly two types of information participants looked for in the recordings, which we coin *inspectables* and *reportables*.

3.3.1 Inspectables. During analysis, one main task was to identify content in the recordings or transcripts that needed to be further inspected and reflected upon. We call the audiovisual segments that correspond to this content *inspectables*. The goal is to find these inspectables in the recordings, and analyze them further to derive useful insights. Below are a few examples of inspectables (in *italics*) that are inspired by examples from our interviews.

- To determine whether users understand how a slider interaction works, the researcher wants to inspect *the audiovisual segments after the user interacts with the slider*.
- To determine how users converse with a voice assistant, the researcher wants to inspect *the audiovisual segments in which the user responds to the voice assistant*.

- To determine reasons why a user paused a video, the researcher wants to inspect *the audiovisual segments before and after the video is paused in the recording*.

To analyze inspectables, researchers need to find them in the recordings, which can be time-consuming. We discuss how our participants located these inspectables in Section 3.4.

3.3.2 Reportables. Towards the end of the analysis, some participants looked for information in the recordings or transcripts, which they reported in their publication. We call such content *reportables*. From our interviews, we identified several instances of reportables, such as quotes from participants (IP1, IP2, IP3, and IP5), images or screenshots of the interface or user behavior (IP1, IP5, and IP8), and video clips for creating a supplementary videos accompanying a written publication (IP5). For IP1, IP5, and IP7, identifying reportables was a separate task, disjoint from the main coding task.

3.4 How researchers navigate recordings to seek information

We discussed earlier that to develop an initial set of codes and determine what events to look for, participants used an opportunistic approach. Participants also used the three following *targeted* information-seeking approaches, in which they used specific features to locate the inspectables and reportables in recordings. The first two approaches result in time-based navigation, whereas the last approach requires the user to look for audiovisual cues in the recordings to locate inspectables.

3.4.1 Time codes for backtracking. Many participants recorded time codes[13] alongside their field notes or during transcription. These time codes allow participants to backtrack to the relevant audiovisual segment in the recording. For example, IP2, IP3, and IP5 recorded time codes of relevant audiovisual segments alongside the user quotations they transcribed in case they needed clarification. Some participants captured time codes by annotating the relevant audiovisual segments before analysis, e.g., during transcription.

Although time codes reduce the time required to locate interesting audiovisual segments, they need to be manually recorded either during the study itself or during transcription. Due to the exploratory nature of qualitative research, researchers cannot reliably determine which audiovisual segments are interesting before they begin the analysis. Therefore, backtracking via time codes is not a readily available navigation method in most qualitative research.

3.4.2 Using information about the study structure. Information about the study structure, such as timeline of tasks and experimental conditions, can help researchers locate inspectables. IP5 provided a good example of using a study's structure to help narrow down possible locations where inspectables might occur (*"[The event] happens five times, every three minutes, [for] 15 seconds, right? [I] have basically looked at minute 3—and looked, uh, about half a minute before, what they are talking about, and then a minute later, has something else [i.e. interesting] happened?"* – IP5).

3.4.3 Detectables. Some participants provided examples of using visual or auditory cues in the recording to locate inspectables. The

---

[13] A time code represents a point of time in the recordings. For example, 1:15 represents 1 minute and 15 seconds into the recording.

inspectables represented events that were too hard to find in the recording, but were usually signaled by the presence of another cue that they could more easily detect. We call these cues *detectables*.

To help understand how inspectables and detectables work together, consider the following hypothetical analysis. Imagine that we would like to understand how people use Wikipedia.[14] We want to investigate why users navigate within and across a wiki page. We perform a screen capture of participants using Wikipedia. When watching the recordings, we observe that participants often go back to previous pages, and we become interested in understanding the various reasons why users backtrack.

To make progress with this research question, we want to locate and analyze all video segments before users backtrack—these are our inspectables. We can locate them using certain cues, e.g., we notice that participants always press the browser's back button to backtrack.[15] Thus, *the back button getting pressed* is a detectable that acts as an indicator for the inspectable. We can navigate the recording looking specifically for back button presses, e.g., by quickly going through the video to identify instances where the back button is clicked. Note that we were not aware of this detectable before collecting our data, and thus could not have generated logs to allow for an easier detection of button clicks.

While in the given example the detectable was objective and concrete, we also consider more abstract, subjective cues detectables. For example, if you want to investigate the instances where participants became angry, a suitable detectable might involving analyzing participants' facial expressions. One reason for distinguishing detectables and inspectables is that detectables lend themselves more readily to automatic detection—a topic we will return to in Section 5.

3.5 Key takeaways

In summary, the key findings from our interviews are:

- Audiovisual recordings contain rich information not available in text. Most participants either coded the recordings directly, or used them in addition to transcripts.
- Most participants do not know what to look for before the analysis begins.
- After familiarizing themselves with the recordings, participants' wanted to locate and analyze interesting audiovisual segments called *inspectables.*
- To find inspectables, which can be tedious, participants used visual and auditory cues called *detectables.*
- Towards the end of the analysis, participants also sought information like screenshots and quotes to report in their work. We call such information *reportables.*

4 SURVEY

To triangulate the findings from our interviews, and better understand what researchers need when doing qualitative audiovisual analysis, we conducted an online survey.[16]

4.1 Method

To get responses from a representative sample of HCI researchers, we sent the survey to first authors of CHI '19 papers that reported qualitative analysis of audiovisual recordings.[17]

We excluded two authors who had given feedback on a pilot of our survey to avoid any conflicts or biases. We also sent the survey to a local HCI research lab's mailing list and a Slack[18] group for HCI researchers. The survey was closed after two weeks. 66 respondents took the survey, and the survey took about ten minutes to complete.

4.2 Details of survey respondents

Although we had no mechanism to detect whether the answer came from a CHI '19 author or from our other postings, we suspect that the majority of respondents are CHI authors, as many of them had responded to our email confirming their participation. Respondents from ten different nationalities took the survey, with most from the US (54%) or the UK (19%), and were either PhD students (65%), academic researchers (12%), e.g., postdoctoral researchers and professors, or Master's students (12%). Only 23% of the respondents had conducted fewer than three qualitative analyses with audiovisual data, while 36% had finished 3–5, 23% between 6 and 20, and 18% had conducted over 20 analyses.

We present selected results from the survey that informed our understanding of how researchers analyze audiovisual recordings. Respondents were asked to answer the survey questions in the context of their most recent qualitative analyses of audiovisual data. If they wanted to contribute multiple analyses, they were instructed to take the survey multiple times.

4.3 What gets analyzed? Audio vs. video

24% of respondents used only audio recordings, 61% used both audio and video recordings, and only 3% used solely video recordings (i.e., without audio). Analyses that used only audio tend to follow a similar workflow, e.g., all such analyses utilized a full transcription of the recordings. Analyses that involved both audio and video, on the other hand, did not exhibit a unified workflow. We found two sub-groups of analyses that involved both audio and video:

- Analyses that focused primarily on the audio, in which the video acted as a supplement, e.g., a think-aloud study with a screen recording. Here, the analysis was done primarily on transcripts, and the recordings were used only to clarify ambiguous statements, e.g., "this button is broken."
- Analyses that focused on rich behavior in both audio and video, e.g., a longitudinal study about drivers' GPS usage. Since such studies are interested in nuanced behavior that is

---

[14]Wikipedia is an online, crowd-sourced encyclopedia. https://www.wikipedia.org/
[15] In reality, users may use multiple different ways to backtrack, e.g., also use keyboard shortcuts—an inspectable might have multiple detectables. But in such a case we simply need to find more detectables. For the sake of simplicity, we assume a single detectable in our example.

[16]For the questions and an anonymized data set of responses, please see the supplements.
[17]We used Reinhard's analysis of CHI '19 papers to identify these papers [34].
[18]Slack is an enterprise communication platform. https://slack.com/intl/en-de/.

difficult to capture in transcripts, researchers need to analyze the recordings directly.

4.4 Types of qualitative research

From our interviews, we identified different types of qualitative analyses. Respondents were asked which of the following descriptions better fit their research questions:

- "Theory building. *For example, modeling people's behavior when using a piece of technology, or understanding the meaning of objects in people's lives.*"
- "Validation of an artifact or hypothesis. *For example, testing a software prototype, or testing retention of information with a new learning method.*"
- "Measurement. *For example, the time spent in different locations, or the time spent talking about different topics.*"

Respondents could select multiple answers and give an open-ended text comment. 63% of analyses were categorized as theory building, and 29% as validation. Analyses involving measurement were rare (3%), as expected for qualitative analysis.

Most validation studies had both audio and video recordings available (17 out of 19 responses). On the other hand, 67% of analyses involving only audio were theory building. This is to be expected since most field researchers use interviews, focus groups, and diaries for data collection, and the largely verbal content is typically audio-recorded and transcribed to allow for text-based analysis.

4.5 How exploratory are audiovisual analyses?

In our interviews, we found that most researchers adopt a rather exploratory style of analysis, where they identify what is interesting only after watching the recordings. With our survey, we wanted to see if this finding holds true for a larger sample and, more importantly, assess the *degree* of exploration in typical analyses. To do this, we asked respondents to rate the two following statements on a five-point Likert scale:

(1) *"Before starting the analysis, I knew clearly where in the recordings to look and what concretely to look for."*
(2) *"Only after I watched (some or all of) the recordings did I find concrete events or aspects to focus the analysis on."*

We designed the statements to be contrasting—if a researcher knew where and what to look for (1), they would not need to watch the recordings before finding concrete aspects to focus on (2). Most respondents disagreed with the first statement (51% rated it 1 or 2, 29% rated it 4 or 5), while most agreed with the second statement (63% rated it 4 or 5, 18% rated it 1 or 2), mostly confirming that the statements are contrasting.

The first question, which we anticipated to provoke disagreement, had 29% agree (rated it 4 or 5), whereas the second statement does not reflect this group—only 15% disagreed (rated it 1 or 2). Upon further inspection, out of the 14 respondents who rated the first statement with 4 or 5, 12 used audio and video recordings. It thus appears that there are certain researchers using audio-and-video recordings who have a predetermined notion of what to look for. Interestingly, with their higher rating for the second statement, they agree that they need to watch the recordings before finding

concrete aspects. One explanation for this is that while these researchers might have had predetermined research questions, they nevertheless had to explore the recordings before they can define their *inspectables* (see Section 3.3.1).

We also aimed to quantify how often the coding scheme is modified during the analysis, another measure for how exploratory it is. We asked respondents to rate the following statement: *"I had a coding or classification scheme (from the start or after observing some of the recordings) that was mostly fixed and that I applied in the rest of the analysis."* 65% indicated that their coding scheme was not fixed (rated 1 or 2). Only 24% reported using a fixed scheme (rated 4 or 5).

Overall, these statistics show that more than half of analyses are exploratory, evident from the evolving coding schemes and the researchers' need to immerse in the recordings before knowing what to focus the analysis on.

4.6 Transcription

We asked respondents to rate the following statement on a Likert scale of 1–5: *"All information necessary for the analysis was first extracted from the recordings into a more convenient form for analysis. (For example, into a transcript, as quotes, or as video clips.)"* 58% of the responses indicated a strong agreement (rating of 5) with the statement. Only 15% indicated that they do not extract the information for the analysis (ratings of 1 or 2). While the statement deliberately left open what kind of extraction was used, this suggests that the recordings had to be prepared for analysis.

The most common form of such data preparation is transcription. The survey contained two questions that aimed to identify the role of transcripts in qualitative analysis: what type of transcripts were used (full, partial, or none), and how much they were used during analysis.

72% of respondents reported preparing a full transcription of their recording, 20% prepared a partial transcription (*"only interesting statements"*), and 8% prepared no transcripts. As mentioned earlier, all respondents who had only audio recordings prepared a full transcription.

To understand our respondents' reliance on transcripts versus working with the recordings directly, we identified four levels of transcript use based on our interviews and literature review:

- *"I used only the transcript and did not revisit the recordings at all."*
- *"I used mainly the transcript and revisited the recordings only to clarify in case of ambiguity or errors in the transcript."*
- *"The analysis required information not present in the transcript, so I used both the transcript and the recordings."* (in the following referred to as having an *insufficient transcript*.)
- *"I primarily used the recordings instead of the transcript."*

13% of respondents used transcripts exclusively in the analysis, and 60% used the audiovisual recordings only for corrections. 25% had insufficient transcripts and one respondent (2%) reported having a full transcript but primarily used the recordings in the analysis.

All respondents whose analyses involved only audio recordings either completely relied on transcripts or used recordings only for corrections. However, when both audio and video was available,

the use of transcripts was split: 50% of audio-and-video analyses with a transcript used the recordings only for corrections, whereas 44% considered their transcripts insufficient. The two remaining participants either relied entirely on the transcript or on the recordings.

This confirms our interview finding that in most analyses that involve audiovisual recordings, recordings are either used exclusively, for corrections, or for specific tasks alongside transcripts. Overall, 32% of participants reportedly did not rely solely on a transcript, but required direct analysis.[19]

4.7 Key takeaways

In summary, the following are the key findings we can takeaway from our survey:

- *Text-based analysis is dominant:* 58% of respondents extracted information from their recordings into a more convenient form for analysis. 71% used full transcripts, and 20% used partial transcript.
- *Transcripts are not always adequate:* 32% of respondents could not rely solely on transcripts and needed to work with the recordings directly.
- *Most qualitative analyses that involve audiovisual recordings are exploratory:* 63% are theory building. 53% of respondents did not know what to look for in the videos before analysis, and 65% changed their coding scheme during the analysis.

5 DISCUSSION: AUTOMATION IN FINDING DETECTABLES

From our interviews and survey, we found that a substantial number of qualitative analyses involve direct analysis of audiovisual recordings. In such analyses, researchers face the laborious task of locating and further analyzing interesting segments, the inspectables, in the recording (see Section 3.3.1). Since inspectables can be abstract and hard to find in the recordings, we found that researchers locate them using information about study structure, backtracking, or detectables (see Section 3.4). Of these, information about study structure and backtracking are time-based navigation techniques, and are already well supported by the traditional time-based media navigation interfaces in QDA tools. On the other hand, to find detectables, which are visual or auditory cues, researchers have to manually search the recordings.

To help with this tedious task of searching for detectables, automation via software support could be useful, as suggested by IP3 when prompted to imagine a magical tool that could help them with qualitative analysis:

> "We have done the analysis once completely. And then we go back and see, aha, OK, is that maybe an interesting pattern which we haven't yet thought of. Then [a magical tool] would, so to say, let us extract these […]" – IP3

In this section, we discuss whether such automation is warranted, how effective it can be, and discuss considerations for design.

---

[19]We counted participants with the following criteria as requiring direct analysis: reported having no transcript, having an insufficient transcript, or using primarily the recordings to require direct analysis.

5.1 Call to automation in finding detectables

Since the detectables lend themselves more readily to detection, we propose to use simple techniques like image recognition and sound detection to automatically find potential locations of detectables in audiovisual segments. The goal is not to automate the entire analysis, but only to help researchers with a tedious task, that of locating possible locations of detectables of audiovisual segments.

To show how automation can be used to find detectables, we provide two examples inspired by our interview findings:

- Building on the example from Section 3.4.3, imagine we want to find all instances where users click the browser's back button. Since the button gets darker when it is clicked, we could capture the average brightness of the button's pixels and mark all frames where the brightness is below a certain threshold. This would lead us to all segments where the back button was clicked.
- Imagine that we want to investigate what participants said to trigger a voice assistant. Since the assistant makes a characteristic activation sound before replying, we could automatically search for that activation sound in the recording. After marking all such instances, we could much more easily inspect the segments of the recording leading up to the response.

We now discuss the scope of such automation across various tasks in qualitative analysis and its application to various research questions.

5.1.1 Scope of automation. The most prominent application of automation is to find inspectables via detectables. Additionally, finding reportables, i.e., user quotes or screenshots, may also require audiovisual cues, which can benefit from automation.

Automation can even be useful before the analysis. From our interviews, we found that researchers who perform partial transcription often look for interesting segments in their recordings, which they transcribe for a later in-depth analysis. This process is similar to searching for detectables, and can benefit from automation.

The approach to automation we propose is not dependent on the research question (see Section 4.4). It is instead limited only by whether the detectables in the recording, which are informed by the research questions, can be reliably detected by software. For example, in behavioral analysis, researchers often looks for subtle non-verbal cues in study participants. Such detectables are hard to detect automatically.

5.2 Design considerations for automation

Our call for automation brings up a classic question in qualitative research: Is it desirable to automate qualitative researchers' workflow? Earlier research has called for the use of computation techniques in qualitative analysis [7]. Work from Marathe and Toyama on qualitative analysis of textual content finds that there is potential to automate many parts of qualitative research [27], but that researchers want automation to be transparent, and desire automation only *after* coming up with a codebook and coding some data. Our interview participants expressed similar concerns for automation in analysis of audiovisual content: (*"[…] Sort of a*

good balance between not letting others analyze my data so that [as a consequence] I have no connection to the data, but also removing the redundant parts of the analysis." – IP1).

We now discuss some design considerations for automation based on the insights we obtained in our interviews, survey, and literature review:

5.2.1 Use simple, explainable, and reproducible techniques that focus on detecting primitives. It is important for automation to be transparent and allow users to operate at a concrete level, e.g., by detecting and tracking basic auditory and visual properties like color, position, and audio frequency. Automation could also combine simple operations that target primitives, e.g., to find all red squares that are within a rectangle, similar to existing stream or video algebras [26, 39].

While complex techniques like neural networks may allow us to detect more abstract, complex instances of detectables, they may also cause the researcher to lose trust in the software, especially for subjective tasks like determining whether a person in a recording is frustrated. For this reason, it is preferable to use simple deterministic algorithms that can be easily understood, replicated, and explained to others. In addition to promoting trust in automation, such an approach also results in a more reproducible research. This discussion is related to the ongoing discussion about explainable AI.

5.2.2 Limit the role of automation to finding potential locations of detectables. Automation should take a backseat in qualitative analysis. Its role is to help the researcher find *potential* locations of detectables, not analyze the inspectables. It is important for automation to minimize false negatives even if that increases false positives. False positives can be manually inspected and removed by the researcher. False negatives, however, would require the researcher to look through the entire recording, which is what the automation was supposed to prevent.

5.2.3 Incorporate other data when available. Software that uses automation should incorporate other data collected in user studies, when available, so that automation can use these additional data sources to help locate detectables. For example, assume that, in a study, the researcher has captured field notes along with time codes in a digital format. These notes can then be used as an additional data source for locating detectables, e.g., to highlight video segments corresponding to all occurrences of a keyword in the field notes. ChronoViz is an existing tool that supports visualization of time series data from multiple sources, such as sensors, digital notes, and GPS, to provide a holistic view of the data collected in user studies [14]. We envision future QDA tools to adopt a similar approach, with an added support for automatically finding detectables.

6 LIMITATIONS AND FUTURE WORK

Our work with this research is meant to be a useful first step in understanding and supporting qualitative researchers perform audiovisual analysis. We recommend the reader to take the intended benefits of automation with reservations, and consider the following challenges.

For one, we do not yet know how automation will be used in practice. Will it help researchers develop insights more effectively?

Or will it lead to less involvement from researchers, and thus lower the quality of research findings? It is also conceivable that, in certain analyses, there are no repetitive cues, i.e., detectables, that can be detected reliably by software. For example, in HCI studies that aim to understand human behavior, software detection may fail, or lead to too many false positives, requiring much effort from the researcher to manually filter them. From a development perspective, it is also possible that since detectables may be unique to each analysis, the automation techniques need to be comprehensive, covering several data dimensions.

7 CONCLUSION

In this paper, we studied how qualitative researchers work with audiovisual recordings. We found that 32% of respondents could not work solely with a transcript but relied on audiovisual recordings. An important task is to locate the inspectables to analyze in detail. Finding the abstract inspectables is achieved through time codes, by exploiting study structure, or by looking for concrete detectables that indicate the inspectables. Towards the end of the analysis, researchers look for reportables, which are pieces of information from the recordings that they report in their publication to illustrate their findings.

We also discussed the potential for the search for detectables and presented our design recommendations. We also discussed the potential for the search for detectables and presented our design recommendations. Our vision is that such automation can save qualitative researchers time and effort in locating inspectables, allowing them to focus on *analyzing* them. We hope that our work sparks further research and development to improve the tool support for qualitative analysis of audiovisual data.

REFERENCES

[1] Anne Adams, Peter Lunt, and Paul Cairns. 2008. *A Qualitative Approach to HCI Research.* Cambridge University Press, Burlington, US, 138–157. https://doi.org/10.1017/CBO9780511814570.008
[2] Hugh Beyer and Karen Holtzblatt. 1999. Contextual Design. *Interactions* 6, 1 (Jan. 1999), 32–42. https://doi.org/10.1145/291224.291229
[3] Joy D. Bringer, Lynne Halley Johnston, and Celia H. Brackenridge. 2006. Using Computer-Assisted Qualitative Data Analysis Software to Develop a Grounded Theory Project. *Field Methods* 18, 3 (Aug. 2006), 245–266. https://doi.org/10.1177/1525822X06287602
[4] Patrick Brundell, Paul Tennent, Chris Greenhalgh, Dawn Knight, Andy Crabtree, Claire O'Malley, Shaaron Ainsworth, David Clarke, Ronald Carter, and Svenja Adolphs. 2008. Digital Replay system (DRS): A Tool for Interaction Analysis. In *Proceedings of the International Conference on Learning Sciences (Workshop on Interaction Analysis).* World Academy of Science, Engineering and Technology, Turkey, 12 pages.
[5] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16).* ACM, New York, USA, 981–992. https://doi.org/10.1145/2858036.2858498
[6] Juan Casares, A. Chris Long, Brad A. Myers, Rishi Bhatnagar, Scott M. Stevens, Laura Dabbish, Dan Yocum, and Albert Corbett. 2002. Simplifying Video Editing

Using Metadata. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (London, England) *(DIS '02)*. ACM, New York, USA, 157–166. https://doi.org/10.1145/778712.778737

[7] Nan-Chen Chen, Margaret Drouhard, Rafal Kocielnik, Jina Suh, and Cecilia R. Aragon. 2018. Using Machine Learning to Support Qualitative Coding in Social Science: Shifting the Focus to Ambiguity, In Using Machine Learning to Support Qualitative Coding in Social Science: Shifting the Focus to Ambiguity. *ACM Trans. Interact. Intell. Syst.* 8, 2, Article 9, 20 pages. https://doi.org/10.1145/3185515

[8] Michael G. Christel, Michael A. Smith, C. Roy Taylor, and David B. Winkler. 1998. Evolving Video Skims into Useful Multimedia Abstractions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98)*. ACM Press/Addison-Wesley Publishing Co., USA, 171–178. https://doi.org/10.1145/274644.274670

[9] Christian Corsten, Chat Wacharamanotham, and Jan Borchers. 2013. Fillables: Everyday Vessels as Tangible Controllers with Adjustable Haptics. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. ACM, New York, USA, 2129–2138. https://doi.org/10.1145/2468356.2468732

[10] Miguel Costa, Nuno Correia, and Nuno Guimarães. 2002. Annotations as Multiple Perspectives of Video Content. In *Proceedings of the Tenth ACM International Conference on Multimedia* (Juan-les-Pins, France) *(MULTIMEDIA '02)*. ACM, New York, USA, 283–286. https://doi.org/10.1145/641007.641065

[11] Pierre Dragicevic, Gonzalo Ramos, Jacobo Bibliowitcz, Derek Nowrouzezahrai, Ravin Balakrishnan, and Karan Singh. 2008. Video Browsing by Direct Manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, USA, 237–246. https://doi.org/10.1145/1357054.1357096

[12] Jeanine Evers, Christina Silver, Katja Mruck, and Bart Peeters. 2011. Introduction to the KWALON Experiment: Discussions on Qualitative Data Analysis Software by Developers and Users. In *Forum: Qualitative Social Research*, Vol. 12. FQS, Berlin, Germany, 12 pages. https://doi.org/10.17169/fqs-12.1.1637

[13] Nigel G. Fielding and Raymond M. Lee. 2002. New Patterns in the Adoption and Use of Qualitative Software. *Field Methods* 14, 2 (May 2002), 197–216. https://doi.org/10.1177/1525822X02014002005

[14] Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D. Hollan. 2011. ChronoViz: A System for Supporting Navigation of Time-Coded Data. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA '11)*. ACM, New York, USA, 299–304. https://doi.org/10.1145/1979742.1979706

[15] Sadaoki Furui. 2005. Spontaneous speech recognition and summarization. In *The Second Baltic Conference on Human Language Technologies*. AIP publishing, Melville, US, 39–50.

[16] Andreas Girgensohn, John Boreczky, and Lynn Wilcox. 2001. Keyframe-Based User Interfaces for Digital Video. *Computer* 34, 9 (Sep. 2001), 61–67. https://doi.org/10.1109/2.947093

[17] Dan B. Goldman, Chris Gonterman, Brian Curless, David Salesin, and Steven M. Seitz. 2008. Video Object Annotation, Navigation, and Composition. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology* (Monterey, CA, USA) *(UIST '08)*. ACM, New York, USA, 3–12. https://doi.org/10.1145/1449715.1449719

[18] Nicky Hayes. 2013. *Doing Qualitative Analysis In Psychology*. Psychology Press, Hove, UK. 295 pages.

[19] Karen Holtzblatt, Jessamyn Burns Wendell, and Shelley Wood. 2005. *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Elsevier/Morgan Kaufmann, San Francisco. https://doi.org/10.1016/B978-0-12-354051-5.X5000-9

[20] Wolfgang Hürst and Patrick Stiegeler. 2002. User Interfaces for Browsing and Navigation of Continuous Multimedia Data. In *Proceedings of the Second Nordic Conference on Human-Computer Interaction* (Aarhus, Denmark) *(NordiCHI '02)*. ACM, New York, USA, 267–270. https://doi.org/10.1145/572020.572062

[21] Eric Isaacson. 2005. What You See Is What You Get: on Visualizing Music. In *6th International Conference on Music Information Retrieval Online Proceedings (ISMIR 2005)*. International Society for Music Information Retrieval, London, UK, 389–395. http://ismir2005.ismir.net/proceedings/index.html

[22] Brigitte Jordan and Austin Henderson. 1995. Interaction Analysis: Foundations and Practice. *Journal of the Learning Sciences* 4, 1 (1995), 39–103. https://doi.org/10.1207/s15327809jls0401_2

[23] Thorsten Karrer, Malte Weiss, Eric Lee, and Jan Borchers. 2008. DRAGON: A Direct Manipulation Interface for Frame-Accurate in-Scene Video Navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, USA, 247–250. https://doi.org/10.1145/1357054.1357097

[24] Jongdae Kim, Charles Gray, Paul Asente, and John Collomosse. 2015. Comprehensible Video Thumbnails. *Computer Graphics Forum* 34, 2 (2015), 167–177. https://doi.org/10.1111/cgf.12550

[25] Eric Lee, Henning Kiel, and Jan Borchers. 2006. *Scrolling Through Time: Improving Interfaces for Searching and Navigating Continuous Audio Timelines*. Technical Report AIB-2006-17. RWTH Aachen University. http://sunsite.informatik.rwth-aachen.de/Publications/AIB/2006/2006-17.pdf

[26] Wendy E. Mackay and Michel Beaudouin-Lafon. 1998. DIVA: Exploratory Data Analysis with Multimedia Streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98)*. ACM Press/Addison-Wesley Publishing Co., USA, 416–423. https://doi.org/10.1145/274644.274701

[27] Megh Marathe and Kentaro Toyama. 2018. Semi-Automated Coding for Qualitative Research: A User-Centered Inquiry and Initial Prototypes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, USA, 1–12. https://doi.org/10.1145/3173574.3173922

[28] D. Thomas Markle, Richard E. West, and Peter J. Rich. 2011. Beyond Transcription: Technology, Change, and Refinement of Method. *Forum: Qualitative Social Research* 12, 3 (2011), 21 pages. https://doi.org/10.17169/fqs-12.3.1564

[29] Alan Marsden, Adrian Mackenzie, Adam Lindsay, Harriet Nock, John Coleman, and Greg Kochanski. 2007. Tools for Searching, Annotation and Analysis of Speech, Music, Film and Video—A Survey. *Literary and Linguistic Computing* 22, 4 (July 2007), 469–488. https://doi.org/10.1093/llc/fqm021

[30] Liliana Melgar, Marijn Koolen, Hugo Huurdeman, and Jaap Blom. 2017. A Process Model of Scholarly Media Annotation. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR '17)*. ACM, New York, USA, 305–308. https://doi.org/10.1145/3020165.3022139

[31] Liliana Melgar Estrada and Marijn Koolen. 2018. Audiovisual Media Annotation Using Qualitative Data Analysis Software: A Comparative Analysis. *The Qualitative Report* 23, 13 (March 2018), 40–60. https://nsuworks.nova.edu/tqr/vol23/iss13/4

[32] Liliana María Melgar Estrada. 2015. *From Social Tagging to Polyrepresentation: A Study of Expert Annotating Behavior of Moving Images*. Ph.D. Dissertation. Universidad Carlos III de Madrid. http://hdl.handle.net/10016/22290

[33] Michael Mills, Jonathan Cohen, and Yin Yin Wong. 1992. A Magnifier Tool for Video Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. ACM, New York, USA, 93–98. https://doi.org/10.1145/142750.142764

[34] Jenny Reinhard. 2020. *Participants, Incentives and User Studies: A Survey of CHI 2019*. Bachelor thesis. RWTH Aachen University, Aachen, Germany. https://hci.rwth-aachen.de/publications/reinhard2020a.pdf

[35] Jane Ritchie, Jane Lewis, Carol McNaughton Nicholls, Rachel Ormston, et al. 2013. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. SAGE Publications, London, UK.

[36] Anselm Strauss and Juliet Corbin. 1994. Grounded Theory Methodology. *Handbook of qualitative research* 17 (1994), 273–285.

[37] Simon Tucker and Steve Whittaker. 2005. Novel Techniques for Time-Compressing Speech: an Exploratory Study. In *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1. IEEE, Philadelphia, PA, 477–480. https://doi.org/10.1109/ICASSP.2005.1415154

[38] Jonathan Tummons. 2014. Using Software for Qualitative Data Analysis: Research Outside Paradigmatic Boundaries. In *Studies in Qualitative Methodology*, Martin Hand and Sam Hillyard (Eds.). Vol. 13. Emerald Group Publishing Limited, Bingley, UK, 155–177. https://doi.org/10.1108/S1042-319220140000013010

[39] Ron Weiss, Andrzej Duda, and David K. Gifford. 1995. Composition and Search with a Video Algebra. *IEEE MultiMedia* 2, 1 (1995), 12–25. https://doi.org/10.1109/93.368596

[40] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: A Professional Framework for Multimodality Research. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*. European Language Resources Association (ELRA), Genoa, Italy, 1556–1559. https://archive.mpi.nl/tla/elan

[41] Megan Woods, Trena Paulus, David P. Atkins, and Rob Macklin. 2016. Advancing Qualitative Research Using Qualitative Data Analysis Software (QDAS)? Reviewing Potential Versus Practice in Published Studies using ATLAS.ti and NVivo, 1994–2013. *Social Science Computer Review* 34, 5 (Oct. 2016), 597–617. https://doi.org/10.1177/0894439315596311