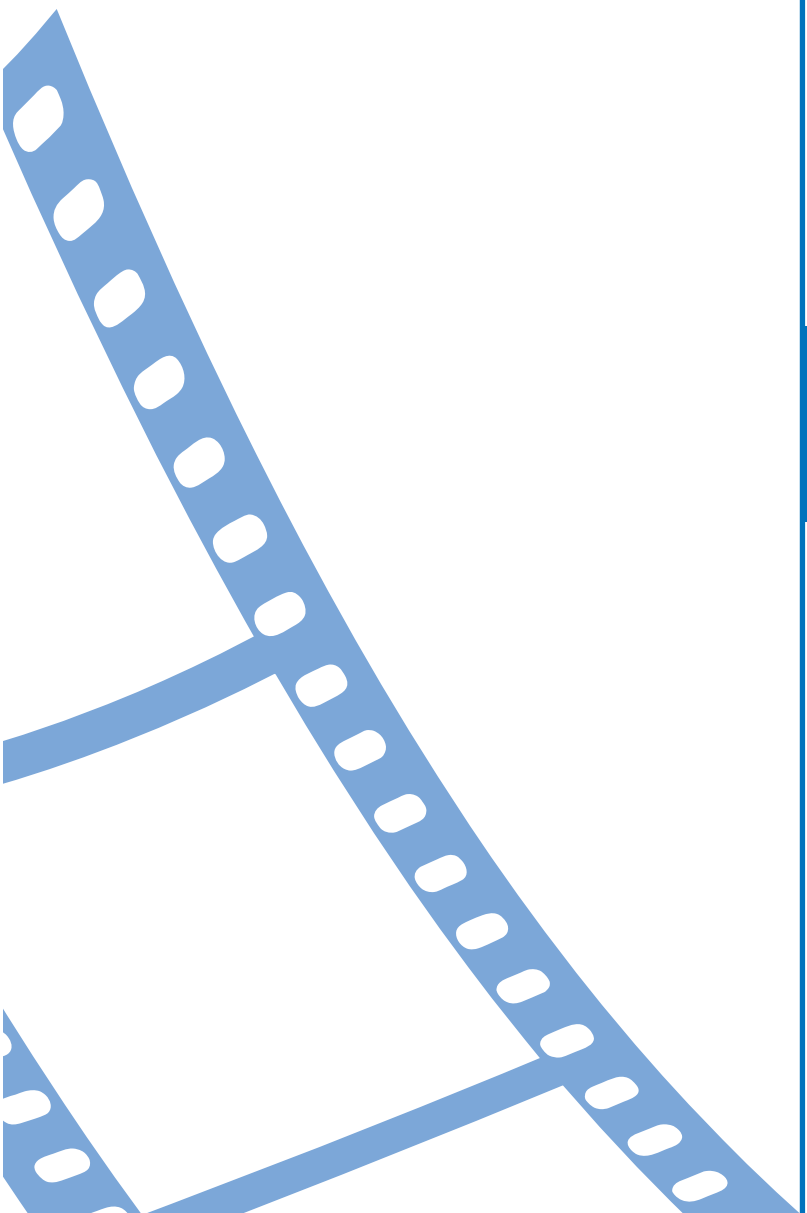
Understanding and Supporting Analysis of Audio and Video



Master's Thesis
submitted to the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Johannes Maas

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Jim Hollan

Registration date: 29.06.2020
Submission date: 09.09.2020

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Statutory Declaration in Lieu of an Oath

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)
Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel
I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
Para. 156 StGB (German Criminal Code): False Statutory Declarations
Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence
(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.
(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature

Contents

List of Figures

Abstract

Qualitative analysis of audio and video is a fundamental task in HCI research, yet researchers suffer from tedious and inefficient practices. Based on interviews with 10 researchers, mostly from the field of Human-Computer Interaction (HCI), we identify common workflows they employ when analyzing audio and video, examine the issues they face, and develop a software solution tackling a specific problem when analyzing videos.

Transcription is a common and valuable process for analyzing what people say. In our survey with 66 participants from HCI 67 % analyzed a transcript while 32 % reported a need to analyze the recordings directly.

From our interviews we identified a fundamental problem in audiovisual analysis. To answer a hypothesis, researchers need to investigate certain sections of the recordings. But finding these *investigable sections* in audiovisual recordings is difficult. In the worst case, the analyst needs to watch the entire recording to identify which sections are relevant for answering the question. But sometimes they can find the investigable sections via patterns that are easy to detect. An example for such a *detectable* is are finding voice assistant responses; if the analyst wants to review all the interactions with a voice assistant (the investigable sections), and they have a clean recording of the voice assistant's output, they can easily find all responses of the assistant by looking for bulges in the recording's waveform (the detectable).

Exploration goes beyond the research question. Analysts need to explore the recordings to find how the answers show up in the recordings. Therefore they cannot anticipate what they will be looking for, and in turn cannot predict what detectables might help them find investigable sections efficiently. Instead they need to retroactively find detectables.

To this end we implemented a prototype that extends ChronoViz with plugins to find detectables automatically. It allows an analyst to find all button clicks in a

screen recording by analyzing the average brightness of the button. Since the button turns darker when clicked, when the button's brightness is below a threshold a click is detectable.

We propose two further design recommendations for supporting analysis of audio and video: Synchronization and integration of all the different kinds of data available to the analyst can help them cross-compare. Offering filtering and combination of annotations can help the analyst navigate to investigable sections more easily.

Especially the notion of detectables has the potential to drive better support for working with audiovisual recordings directly in qualitative analysis software.

Acknowledgements

Krishna Subramanian, நன்றி. You are a helpful and patient supervisor. Working with you was collaborative and sometimes felt playful which made the experience rather pleasant. I hope you are as proud of our work as I am.

Prof. Dr. Jan Borchers, thank you for being the first examiner. You helped us avoid pitfalls and tune our reasoning with your feedback.

Prof. Dr. Jim Hollan, thank you for being the second examiner. You gave us motivating feedback on our research direction, which is especially valuable since you were involved in the development of ChronoViz which already offered much of what we needed to build our prototype.

Prof. Dr. Chat Wacharamanotham, thank you for dedicating a substantial amount of your time to give us feedback on the survey. We took advantage of every last brownie point, receiving critical feedback that undoubtedly and fundamentally improved the survey.

Alexander Eiselmayer, thank you for taking the time to help us with the survey. You provided clear and actionable feedback, a capability which I admire.

Dr. Adam Fouse, thank you for not only trusting me with ChronoViz's source code, but also helping me solve the more difficult issues towards implementing our extensions. It was exciting and rewarding to work on ChronoViz with you, which is amazingly well-designed.

Sebastian Hueber, thank you for saving me from headaches. Without you I would still be banging my head against a Mac.

We appreciate all our interview and study participants who selflessly dedicated their time to us. We hope that our work will eventually give you a better analysis experience in return.

Conventions

References to the main researchers of this work, namely the author, Johannes Maas, and his supervisor, Krishna Subramanian, will use the pronoun *we*. This aims to acknowledge the importance of the collaboration to the success of the research. We consistently use this style, even when expressing opinions of the main author.

We utilize singular they for increased anonymity.
Example: When this one user did something, *they* had a certain reaction.

Chapter 1

Introduction: Qualitative analysis of audio and video

To explore new fields or to better understand complex phenomena, researchers use qualitative analysis of audio and video recordings. The qualitative method is required because audiovisual recordings are difficult to quantify into numbers. (Section 1.1)

In qualitative analysis, the researchers code the data. They select text pieces and mark them with a category. Over the course of the analysis, the categories build up and evolve to help the analyst form a structured understanding of the data. (Section 1.2)

Before the analysis, oftentimes the recordings are transcribed into text, because working with the recordings is much more difficult and text-based analysis is well-established. Unfortunately, the process of transcription takes a lot of time and is considered tedious. (Section 1.3)

The recordings can also be coded directly, without first creating a transcript which is appealing because transcription is so tedious. With software tools, direct analysis has become feasible. (Section 1.4)

Unfortunately, direct analysis is not as well-supported as text-based analysis. Using direct analysis software is perceived as inefficient. We wanted to better understand these problems to find solutions that improve the support for direct analysis. (Section 1.5)

1.1 Methodology

Quantitative methods are based on numbers.

There are two approaches in scientific research: Quantitative and qualitative methods. Quantitative methods work with numbers and mathematical procedures. Therefore, to study a phenomenon from the real world it has been broken down into (quantified into) such numbers.

For subjects that need to be interpreted, researchers use qualitative analysis.

On the other hand, qualitative methods are employed when the researchers want to study something that is hard to put into numbers. This is usually the case when they are exploring something new, but it also happens when they study human behavior. We are doing many things unconsciously and they are therefore hard to put into rigid numbers. In that case, the researchers need to first interpret the data before they make sense out of it.

Qualitative analysis is an important part of many fields, such as Human-Computer Interaction (HCI), psychology, social sciences and many more. We will focus on the relevance for HCI, where qualitative analysis is employed to study how humans interact with computational systems. This can take different forms, such as interviews or prototype studies.

Qualitative analysis is considered subjective.

There is a tension between quantitative and qualitative approaches. Quantitative studies are considered more objective, because they rely on clear measurements and mathematical procedures. Qualitative methods on the other hand are intimately tied to the researchers' interpretation and thus considered more subjective.

To combat this subjectivity of qualitative analysis, approaches emerged that aim to ensure all findings are supported by the data. One of the prominent formulations

is *Grounded Theory* [Strauss and Corbin, 1994] that prescribes the principle of "grounding" the findings in the data through *constant comparison* [Boeije, 2002]. That is, after finding something new in the data, the researcher should go back, compare it with more of the data, and make sure that the data really supports that finding.

Grounded theory aims to combat subjectivity through constant comparison.

1.2 Coding

Qualitative analysis often revolves around text. Phrases in the text that are relevant for the analysis can be highlighted and assigned a *code* (sometimes called an annotation). A code is similar to a category; it is an description of what the analyst thinks the phrase means. The process of going through the text and marking pieces of text with a code summarizing their meaning is called *coding*.

During this process, the analyst will add more and more codes to more and more of the text. An important part of the process is to build relationships between the codes. A few codes might be similar and can be grouped together with a new code, or after finding more cases of another code, the analyst notices that it is better to split the code into more specific ones. This way, the analyst builds a hierarchy of codes that reflects their understanding of the text.

The constant comparison suggested by Grounded Theory happens for example when the researcher wants to study a specific code. They can then go through the entire text again looking for more cases that fit that code. Doing this, they might find new relationships or cases that evolve their understanding.

This process enables the researcher to both build a structured understanding of the data as well as ground the findings in the data by always having specific examples. There are software tools that help with coding text and improve the analysis by, for example, listing all the examples that belong to a specific code.

Analysts use coding to structure their understanding.

1.3 Transcription

Coding is well-supported on text, but video and especially audio recordings are commonly used to capture data for analysis. For example in an interview, instead of conducting the analysis based only on the notes written by the researcher, today it is cheap and easy to record the interview with a microphone.

Transcription is making text out of audio, enabling text-based analysis.

With recordings of the interview available, analysts can write out what the participants said word for word. Changing the form from an audio recording into text is called *transcribing*. Such a transcription is a textual representation of a recording, which can be analyzed using the methods and software tools that already exist for text-based analysis.

There are different types of transcript, depending on what level of detail is required in the analysis. Some fields such as linguistics are interested in preserving which words are stressed, how they are articulated, and how long the pauses are. They have developed very rich transcript notations like the Jefferson Notation [Jefferson, 2004].

Transcription is time-consuming and tedious.

Obviously, transcribing recordings in this detail is work that takes a long time and concentrated effort. But even simpler forms of transcript, for example verbatim transcripts that look like conventional text and only capture the words, take substantial time to complete. Transcription is a time-consuming and tedious process.

1.4 Direct analysis

Because the transcription process is so expensive and tedious, researchers try their best to find more efficient approaches. Especially since audiovisual recordings have become accessible and are now used routinely to record data, the time that researchers would need to spend on transcription has grown substantially.

Based on data from an analysis of all CHI '19 papers [Rein-

hard, 2020] we find that 40 % (283 out of 705) of CHI' 19 papers involve audio or video recordings in their analysis.[1] This means that a large proportion of analyses in HCI have to process recordings.

40 % of CHI papers use audio or video recordings.

The trend continues towards more recordings. Lifelogging is an emerging video recording method that captures entire days continuously, resulting in large amounts of video recordings to analyze [Gurrin et al., 2014]. We expect that audiovisual recordings will play an even bigger role in research in the future.

At the same time, dealing with such recordings has become easier in software. Programs were developed that allowed researchers to code recordings directly—by selecting sections of the recording instead of text. This removed the technical need for transcribing the recordings to access the established methods and tools of text-based analysis.

Direct analysis can be seen as an alternative to transcription.

Transcripts do have some benefits. During transcription the analyst familiarizes themselves intimately with the recordings, because they spend a lot of time with them. And a textual transcript can be easily skimmed, compared, and searched in; features which are not supported for audiovisual recordings. Transcripts can also serve to share understanding across a team of researchers, for example by describing what is happening in a computer game recording to those unfamiliar with the game [Woods and Dempster, 2011]. Transcripts remain relevant even with the option of direct analysis.

Transcripts are easier to navigate.

Nonetheless, there are disadvantages to transcripts beyond the effort of creating one. Transcription is a lossy process in which not all information is retained. While for many analyses keeping what is being said is perfectly sufficient, there are analyses that require the full richness of audiovisual recordings beyond what is easily described textually. Therefore there are cases in which transcripts are inappropriate.

Transcripts lose some of the recording's richness.

All in all, working directly with the recording is already an important part of research. It will likely become even

[1] See Appendix B.2.2 for details on how we determined this number.

more important as more audiovisual data is collected, for example in humanities and art Marsden et al. [2007]. Consequently, researchers should have access to tools that support them in their needs when analyzing audio and video.

1.5 Current issues

Available software tends to be geared towards text-based analysis. It does not support direct analysis well, because it lacks navigational aids that are particularly needed for audiovisual recordings. There is research software available that tackles specific problems. But overall there is little research into the current practices that analysts developed to deal with audiovisual analysis.

There is a plethora of software available for qualitative data analysis [Evers et al., 2010, Melgar Estrada and Koolen, 2018, Silver and Lewins, 2014, Saldaña, 2009]. The most widely used tools appear to be

- NVivo (`https://www.qsrinternational.com/nvivo/home`)
- ATLAS.ti (`https://atlasti.com`)
- MAXQDA (`https://www.maxqda.com`)
- Transana (`https://www.transana.com/`)
- ELAN (`https://archive.mpi.nl/tla/elan`)

Existing software is rather text-oriented.

These existing solutions tend to prioritize text-based analysis. Text-based analysis was established first and only later was direct analysis even possible. The established programs therefore integrated audiovisual analysis into their existing text-based toolset. As a result, working with audiovisual recordings feels tedious and inefficient.

There is little research into current practices.

While there is research around qualitative data analysis and direct analysis, as well as research software that tackles a specific problem, there is little research into the practices

that analysts currently employ when analyzing audiovisual recordings.

As a consequence, we wanted to investigate the needs of researchers analyzing audio and video recordings to identify problems and propose solutions.

Chapter 2

Related work: Easing qualitative analysis

2.1 Qualitative data analysis software

The use of qualitative data analysis software (QDA software) has undergone a discussion about its validity. Criticisms range from pointing out a friction between generalized software tools and the need to focus on peculiarities in the analysis [MacMillan, 2005] to a suspicion about the impact the software has on the methodology partly because of its adoption by novices [Fielding and Lee, 2002]. Proponents argue that the software aids the task, for example by allowing the analysts to "think visibly" [Konopásek, 2008]. The core of the argument that QDA software has a negative impact on the methodology is refuted [Woods et al., 2016, Bringer et al., 2006, Tummons, 2014, Smith and Hesse-Biber, 1996]. Instead, QDA software is considered to enable researchers to combine multiple methods more easily into "thick analysis" [Evers, 2015].

QDA software is helpful.

Some of the programs listed in Section 1.5 have been compared against each other [Melgar Estrada and Koolen, 2018]. There are introductions and design documents for

NVIVO [Bazeley and Jackson, 2013], ATLAS.ti [Hwang, 2008, Friese, 2019] and ELAN [Brugman and Russel, 2004].

KWALON is a Dutch journal that organized a comparative analysis of QDA software [Evers et al., 2010]. It provided a data set including audio and video files that different teams analyzed using different software. Three teams analyzed the audio or video files: Woods and Dempster [2011], Dempster and Woods [2011], and Mavrou et al. [2007].

There is little research into issues of direct analysis.

Overall, this research is more focused on discussing the general methodology or investigating the role of transcription. We will focus instead on the issues with direct analysis and the practices that have evolved.

2.2 Transcription

A fundamental decision in the analysis process is whether to transcribe the recordings into text. Even with support for direct coding, transcription can be beneficial, because transcripts do not necessarily need to represent what is spoken verbatim, but can be descriptions of what is going on [Evers, 2010]. With this flexibility, transcripts can make explicit observations of a recording to ease collaboration [Woods and Dempster, 2011], and are especially helpful when synchronized with the recording.

Transcription can be automated to a degree.

Since transcription is a tedious process when done manually, there is big interest in automating it. One approach is using algorithmic speech recognition [Matheson, 2007, Whittaker et al., 2002]. Even with errors present, there are mitigation techniques like focusing on key phrases which are reliably detected [Désilets et al., 2002], or making the playback for checking the transcript more efficient through time-compression techniques [Ranjan et al., 2006]. It is possible to automatically segment the recording by the speakers [Kimber et al., 1995, Heeman and Allen, 1995].

A less automatic approach is to crowd-source transcription [Vashistha et al., 2017], but there are concerns about confidentiality with the recordings that may disqualify such a

method.

All this work shows that the problem of transcription is being tackled from many angles. While it remains a somewhat tedious process, it is an important and robust analysis tool. As we will see in Section 4.7, it is the dominant analysis approach, but there is also a need for direct analysis. We will focus more on the peculiarities of working with audiovisual recordings than transcripts.

Transcription is well-supported. We focus on direct analysis.

2.3 Multimedia analysis support

Transana supports video analysis by allowing researchers to create and synchronize multiple transcripts to a video file [Dempster and Woods, 2011, Woods and Dempster, 2011].

There are different tools built for QDA analysis.

On the direct coding side, DIVA offers an innovative visualization of codes [Mackay and Beaudouin-Lafon, 1998], as well as a stream manipulation algebra for combining or isolating coded events. Savanta similarly offers a tree-based view of the codes and can restrict playback to a selection of codes, offering the ability to find overlaps between codes [Hauglid and Heggland, 2008]. VCode and VData are companion tools offering streamlined annotation and inter-coder agreement checks [Hagedorn et al., 2008].

There are specialized programs, like Oudjat for facial analysis [Dupre et al., 2015] and BORIS for annotating animal behavior [Friard and Gamba, 2016].

These programs are specialized to solve a very specific problem. But in HCI there are exploratory analyses where you cannot predict how the analysis will go. We are interested in finding general requirements in such analyses and formulating general design recommendations for audiovisual analysis software.

They did not study or solve the issues in exploratory analysis.

2.4 Media navigation

Navigating audio and video has received ample attention because it is necessary in a wide variety of contexts. In each situation, different characteristics can be exploited to tackle the problem of navigating such media.

There are many techniques for navigating audio and video.

For audio, time compression and variable playback rates may help finding desired content [Lauer and Hürst, 2007], as well as scrubbing, whereby a playhead is moved through the audio faster than real-time [Lee and Borchers, 2006]. The elastic audio slider is an interaction design allowing variable playback speed, and has been used to show that audio content is understandable at a playback rate of 1.8 and the overall topic classifiable at rates up to 3 [Hürst et al., 2006].

For video editors one problem is the need for different granularities of zoom. This is tackled by offering multi-level timelines [Casares et al., 2002] or a magnifying timeline [Mills et al., 1992]. In a similar vein, there is a framework describing the different requirements when trying to compare different videos [Tharatipyakul and Lee, 2018].

Videos may also be summarized in different ways [Smith and Kanade, 2005]. The contained movement can be visualized using a third graphical dimension [Nguyen et al., 2012]. It can even be automatically analyzed to extract relevant keyframes [Girgensohn et al., 2001].

Similar to the summarization, videos may be searched in for color patterns or shot compositions [Zhang et al., 1995]. The search for video tutorials can be aided by highlighting segments relevant to the keywords [Fraser et al., 2019].

Synchronizing transcripts or data is beneficial.

Synchronization of multiple media allows for interesting new navigation techniques. A synchronized transcript, as available in Transana, allows researchers to re-listen a portion of the transcript easily when the transcript lacks richness [Woods and Dempster, 2011, Dempster and Woods, 2011]. Synchronized audio, video, and data logs of everyday phone use were exploited to select relevant segments

for analysis, enabling an in-vivo analysis based on an otherwise overwhelming dataset [McMillan et al., 2015].

These techniques do not show up in QDA software.

While these are important technical tools that enable media navigation for specific cases, they are not usually integrated in qualitative data analysis software. We will investigate how such techniques can be applied to solve problems in audiovisual analysis.

Chapter 3

Interview study: Understanding workflows and issues

To get a general understanding of how researchers execute qualitative analysis involving audio or video, we conducted 8 interviews with such analysts, mostly from HCI. (Section 3.1)

We found that exploration goes beyond the research questions; our participants usually also needed to explore how the answers to their questions would manifest themselves in the recordings. For such exploration, they employed navigation techniques to help them familiarize with the recordings as quickly as possible while retaining just the necessary information. (Section 3.2)

Since working with audio and video recordings is difficult, our participants tended to first extract relevant information into a more convenient form. This mostly happened through transcription, but with partial transcription or extraction of clips there is a pronounced risk of premature filtering that can bias the findings with preconceived ideas. (Section 3.3)

Our participants needed to identify and find sections in the

recordings to further investigate. They used different techniques that helped identify these *investigable sections*. During data gathering and exploration, moments deemed interesting for future investigation were marked with backlinks such as timestamps. Additionally, analysts exploit their knowledge about the study structure to navigate the recordings. We introduce the term *detectables* to describe the approach to search for easily detectable symptoms that help efficiently identify the investigable sections. When writing up their findings, our participants tended to look for illustrative examples which we call *reportables*. (Section 3.4)

3.1 Method

We had 10 interviews in total.

The author first conducted three exploratory interviews (Section 3.1.1), followed by six focused interviews (Section 3.1.2). The findings in this section are derived from these interviews. After the survey (Chapter 4) we conducted follow-up interviews (Section 3.1.4) from which we will include supportive quotations.

We will refer to the interviewees from the exploratory studies as E1, E2, E3 (E for exploratory); those from the focused interviews are I4, I5, I6, I7, I8 (I for interview); the follow-up interview participants are F9 and F10 (F for follow-up).

The interviewees were mostly PhD students from HCI, with moderate analysis experience working in Germanic countries, summarized in Table 3.1 and the following list.

- Gender: 8 men, 2 women.
- Country: 5 Germany, 2 Switzerland, 1 Japan, 1 Sweden, 1 UK.
- Language: 6 English, 4 German.

All 10 interviews were conducted via teleconference[1] and the audio recorded. I4's recording had occasional lapses

[1] We used Jitsi (`https://meet.jit.si`) and Skype.

ID	Field	Status	Finished analyses[1]	Duration[2] (in minutes)
E1	HCI	PhD	1–2	71
E2	HCI	Prof	3–5	41
E3	HCI	PhD	1–2	60
I4	HCI	MSc	1–2	43
I5	HCI	PhD	3–5	28
I6	Psy	PhD	3–5	29
I7	HCI	MSc	1–2	30
I8	HCI	Prof	>20	31
F9	HCI	PhD	1–2	17
F10	HCI	PhD	6–10	64

[1]Self-reported number of analyses involving audio or video recordings. Reported in ranges for anonymity.
[2]The length of the recording that was transcribed. Does not include introduction and wrap-up.
HCI: Human-Computer Interaction;
Psy: Psychology.
PhD: Doctoral student;
Prof: Professor;
MSc: Master of Science student.

Table 3.1: Overview of the most important characteristics of our interviews.

due to an unknown technical problem, but most of it was still comprehensible. All interviews were fully transcribed verbatim, except for F10 as described in Section 3.1.4.

3.1.1 Exploratory interviews

We started with exploratory interviews.

The first three interviews were kept open and explored the interviewee's analysis approach, workflow and issues. The aim was to get a detailed understanding of how their analysis was conducted and was problems they faced. Consequently, they were substantially longer and less structured than the following interviews.

The pilot study, E1, was conducted with a participant who was familiar with the research goal of this thesis. But because we focused on understanding a specific analysis, we

do not expect a significant bias and included the interview in the analysis.

3.1.2 Focused interviews

We conducted further interviews.

Based on the notes of those first interviews, the research questions and scope of the thesis was defined: We wanted to focus on what analysts look for and how they look for it. The five following interviews were thus more focused, taking less time.

The five focused interviews had a protocol[2] and started with the presentation of our research question: what the analyst looked for and how they looked for it. Then we gathered basic demographic information and made our interviewees recall a recent analysis, asking for its research question, the data gathered, and a general walkthrough. This provided us with a general understanding of their analysis and situated[3] them back into it. We proceeded to ask what specific things they looked for in that analysis, and discussed in detail how they found each of those things.

3.1.3 Analysis

We analyzed in the style of an affinity diagram.

After conducting the first eight interviews, we categorized their contents in the style of affinity diagramming [Holtzblatt et al., 2005]; the author read through the transcripts and picked out statements that he deemed interesting for further analysis. German statements were translated to English.

Then the author and his supervisor categorized similar statements in a mind map into a hierarchy of themes, de-

[2] A summary of the protocol is available in Appendix A.3.

[3] Interviews are not objective and the researcher influences the data [Miller and Glassner, 2011, Qu and Dumay, 2011]. Similar to a "go-along interview" [Carpiano, 2009], we aimed to recreate as much of the analysis experience as possible. To this end we asked them during recruitment to have as much material from the analysis available as possible, to facilitate recall of details.

liberately resisting predefining categories and developing the themes on the grounds of the statements. After the first session, the author went through the transcripts again and gathered more statements that were integrated in the diagram in two subsequent sessions. In total we had three sessions spanning about 6 hours of affinity diagramming.

3.1.4 Follow-up interviews

After the survey, we had only two follow-up interviews.

After the analysis and the survey (Chapter 4) we were interested in finding more examples of detectables (Section 3.4.4). We recruited participants from the survey who indicated that they employed direct analysis[4] by displaying an invitation to participate in an interview to them.

In the end, we conducted two more interviews, F9 and F10.

Note that F10 did not match the anticipated typical video analysis, but offered comprehensive insight into a participatory study that used videos as prompts in an interview. Due to its length it was not fully transcribed, but first fully paraphrased and then selected portions transcribed verbatim.

All in all, we did not discover as many examples as desired. Since, however, these interviews offered insight into two more analyses, we will use their quotes to support and illustrate our findings in this chapter.

3.2 Exploration

Some of our participants deliberately explored novel areas, consciously starting with a rough research question to be developed and defined along the way. We noticed

[4] We invited participants who reported using no transcript in question 11, or requiring information not present in the transcript or primarily using the recordings in question 12. The survey questions are available in Appendix B.1.

that some exhibited a shift from an exploratory to a confirmatory approach which is apparent when they switched from watching the recordings chronologically to skipping between recordings looking for a specific aspect. (Section 3.2.1)

We give an overview of the techniques our participants reported to familiarize themselves with the recordings as efficiently as possible. (Section 3.2.2)

3.2.1 Research question

Exploratory analysis typically refers to an open research question.

What it typically means for an analysis to be *exploratory* can be condensed to its research question being flexible. The researchers will start familiarizing themselves with the data with only a vague hunch. Having some insight into the data, typically patterns will emerge and the researchers will focus on the ones they find most interesting, thus shifting and refining their research question until it is focused enough to generate testable hypotheses.

The counterpart of the spectrum is confirmatory.

The counterpart to exploratory is *confirmatory*. If question is known beforehand, for example in the form a hypothesis, the analysis serves to confirm—or deny—the validity of the hypothesis. For quantitative analyses, this is the normal mode of operation, as the measures are chosen specifically to generate data for this hypothesis. Qualitative analyses on the other hand tend to be substantially exploratory; either because they deliberately explore uncharted territory, or because they are studying difficult to quantitatively measure qualities.

An example of a rather exploratory interview from our study is I8. They studied how people use a certain teaching medium by videotaping study participants using the medium. After an initial analysis, they became interested in the interaction with a specific feature. They focused on all instances where this feature was used and studied the different purposes it was used for. Their analysis thus started exploratory and ended confirmatory of the different interaction purposes discovered.

Already starting out rather confirmatory were I4 and I7 who built artifacts that they validated in a study. Both of them had clear ideas of which of the artifact's features they wanted to test, and they designed their studies around those.

We found a noticable shift from exploratory to confirmatory.

It might seem that counting an analysis as exploratory rather than confirmatory is arbitrary. But we discovered a symptom of when an analysis switches from exploratory to confirmatory.

> So for the first third [...] I always heard the entire interview and then I checked my three individual thematic areas. [...] Theme one, theme, two theme three. Next interview, theme one, theme two, theme three. [...] In the second half [...] I did it theme-wise. [...] always only theme one, next person theme one, third person theme one. (I4)

During exploration the analyst looks at data individually, combing through it chronologically. For confirmation they look across the data for a specific aspect.

3.2.2 Opportunistic observation

Participants typically needed to familiarize themselves with the recordings.

In an exploratory analysis, the researcher does not have enough understanding of the data to formulate a concrete hypothesis to investigate. First, they must familiarize themselves with the recordings, for example to make surprising observations that spark concrete research questions. The goal is simply to observe the recordings until the analyst has come across something interesting.

Many times not every little detail is important for familiarization. Analysts therefore utilize different techniques to quicken opportunistic observation while retaining just enough detail.

Our participants reported playing the recording at a faster

They used techniques such as faster playback for opportunistic observation.

rate (E2, I5, I7), skipping sections they knew were irrelevant (E2), scrubbing through a video (I7), or even playing in the background:

> I then let it—run and you just heard on the side what the people are talking, right? And if some said [...] something's not working, then I have watched it again, if that right now is important for our data. (I5)

These techniques were also used for constant comparison.

Opportunistic observation can also be employed later in the analysis in the spirit of constant comparison. Interviewees reported a need for immersion in the data.

Participants wanted to cross-compare.

The most prominent reason is comparison across the entire data and across the duration of the analysis. Making these connections requires recalling aspects of some part when encountering another part of the data. I6 reported that they first fully transcribed their recordings to enable them to compare findings across all their data.

> [We] transcribed all focus groups fully and then [did] the thematic analysis [...] you have to—make connections between the things of the focus groups. (I6)

Participants needed to understand nuances.

Immersion is also required to form the necessary understanding of the material to conduct the analysis. Some analyses focus on behavior that needs to be watched carefully and repeated to understand its facets.

> The qualitative analysis, in the bottom-up manner relies on the fact that researchers [...] can, uh, understand the data, can read between lines. (E2)

I8 extracted video clips around moments of interest to analyze them in great detail.

> We would actually kind of sit through the videos—in a little bit more detail, and so we'd maybe spend—couple hours on each clip. (I8)

In summary, while opportunistic watching is most prominently employed for initial familiarization with the data, it is also used later for constant comparison across the data.

3.3 Extraction

Audio and video recordings are unsuited for analysis and researchers will try to extract the required information into a more convenient form. Most popularly this takes the form of transcribing the spoken words into text. But some of our participants transcribed only interesting parts of the recordings or extracted clips which has a pronounced risk of prematurely filtering out relevant information.

3.3.1 Transcription

Transcription practices varied wildly.

Three of our interviewees reported creating a full transcript (I6, I8, F10). Three reported partially transcribing the recordings (E2, E3, I5). Furthermore, three participants reported taking notes about what is happening (I4, I7, F9) and one mainly coded their videos (E1).

Transcripts were time-consuming.

The often-mentioned tedium of transcription manifests in the lengths of time our interviewees reported spending on transcription. E3 estimated spending 3 hours to extract interesting statements per hour of recording. I6 reported that a full transcription of 10 minutes of an interview required 50 minutes, and 90 minutes when inexperienced. These are substantial amounts of time spent to make accessible data in textual format.

Investing in a full transcription pays of by enabling sifting through the data more efficiently. Without a transcript,

you are bound to near-real-time playback speed, whereas a transcript can be skimmed and jumped around in easily.

Partial transcription can be verbatim or summative.

Not all partial transcription is the same. E2 and E3 extracted interesting statements for affinity diagramming into text by listening to the audio recordings, and I4 and I7 listened to the recordings and noted down interesting observations for their artifact validation. E2's and E3's analyses are more exploratory, building theory statement by statement, while I4 and I7 are more confirmatory, checking predefined aspects across their participants.

Transcripts may include non-verbal context.

The full and the partial transcripts were verbatim, i.e. they recorded all spoken words. The transcripts sometimes also contained additional information that was not strictly verbal, classifiable as "pragmatic transcription" [Evers, 2010]. E3 explained that they wanted to clarify which button a participant referred to, for example "I don't know what this one [back button] does.". Silence is another example of such contextual information:

> In our transcription, um, we might not have anything, because participant did not say anything while they are struggling [...] But we, if we watch the video, we will be able to note down that [...] the participant searched for—the function [...] and then—expressed some frustrations and stuff. (E2)

3.3.2 Clips

In addition to marking detectables, interviewees reported extracting portions of the recordings. In contrast to selection, extraction loses the context of the section but helps reduce the amount of data to an adequate amount.

Some participants extracted short clips for analysis.

I8 identified points of interest in the recording that they wanted to analyze in more depth. They extracted the sections around those moments into video clips that they watched and analyzed in much detail.

F9 also extracted video clips showing participants executing a specific action, in order to ask their rational in a follow-up interview.

It seems that this is a case where the researchers know that they only need to focus on specific parts of the recording. They preferred cutting out everything else to concentrate on these sections, which we will later call *investigable sections* (Section 3.4.1).

3.3.3 Risk of premature filtering

Participants are conscious about the risk of prefiltering.

Partial transcription and note taking risk prefiltering the data and discourages analysts from revisiting the untranscribed portions. Our interviewees were conscious about this risk of premature filtering.

> So to me the partial transcription is exteremely important. Because it already includes, so to say, a kind of pre-filter. (E3)

As E3 and their colleague split up the transcription work, they also met regularly to ensure they extracted similar statements. They went through each others extractions and recalled whether something similar happened in the section they had just transcribed and if necessary went back and extracted it.

Participants are aware of their biases.

I6, who transcribed fully, exploited the transcription process to familiarize themselves with the data before beginning the actual analysis. They deliberately resisted holding on to their initial ideas.

> This is, um, recommended by Clarke and Braun [See Braun and Clarke [2006]]. If you commit too quickly to things, then, um—you pursue what you but not what the data say. (I6)

This underlines the importance for analysts to move back and forth in all their data; after extracting statements or clips, they should not lose the ability to contextualize their findings in the other parts of the recordings.

3.4 Navigation

A key task in the analysis is to identify *investigable sections*. These are the parts of the recordings that are relevant for answering the research question or testing a hypothesis. They are not apparent and must be found before they can be investigated. Analysts have different techniques that help them navigate to those sections more efficiently. (Section 3.4.1)

During data gathering or exploration, analysts might already identify investigable moments. They can use *backlinks* to these moments so that during later analysis they can easily find them. (Section 3.4.2)

Analysts might also have knowledge about the study structure that helps them focus their search for investigable sections. (Section 3.4.3)

Another technique is to identify a symptom or pattern that makes investigable sections easier to detect, which we call *detectables*. (Section 3.4.4)

Lastly, at the end of the analysis when researchers report their findings, they often are looking for illustrative examples which we call *reportables*. (Section 3.4.5)

3.4.1 Investigable sections

Interviewees reported that their recordings contained sections not relevant to the analysis. These sections included setup, training, or digressions in an interview.

The recordings can not always feasibly be kept clean from

such sections. If the setup is at the beginning, the recording could be started later. But digressions in an interview can be due to accommodation towards the participant's comfort.

> Because you as the interviewer, um, you don't wanna sound too—formal and impersonal, so you tend to chit-chat a little bit, because that—always puts the interviewee at ease, and they open up more. Basic psychological tricks, right? [chuckles] (E1)

Participants needed to identify the investigable sections in the recordings.

Consequently, a key task in the analysis is to identify which sections of the recordings need to be further investigated. We introduce the concept of *investigable sections* to enable discussion of this problem. Basically, there are sections that the analyst wants to investigate, but they are surrounded by irrelevant sections and in audiovisual recordings it is difficult to differentiate them.

In the worst case, researchers need to watch the entire recording to identify all investigable sections, using a opportunistic observation as presented in Section 3.2.2. Sometimes however the researchers will be able to utilize techniques to identify the investigable sections more efficiently.

3.4.2 Backlinks

Timestamps link back to a moment that was deemed interesting.

A commonly used technique for navigating to investigable sections is backlinking. It typically consists of referring to a certain point in time of the recording. For example, E2, E3, I4, F9 reported adding the current timestamp of the recording while taking notes, for the ability to later revisit that moment in the analysis. Another example, as I6 reported, is to use QDA software to list all parts of a text that are marked with a code. Backlinking can take different forms, although timestamps are the most wide-spread.

Backlinking is especially useful to allow verification, such as adding timestamps to a transcript or to quotes used in

Timestamped (partial) transcripts are useful.

affinity diagramming. If the transcript or the quote require more context or appear to contain an error, they can easily be checked in the recording thanks to the timecode.

> In general I have not gone back to the audio file, except—um, when we for example noticed during the analysis that the text-snippet itself is not enough, or is a bit—ambiguous. (E3)

Some participants did not take notes.

However, I4 and I7 reported that they deliberately did not take notes during the study to make their subjects feel at ease.

> I tried my best not to make them nervous, so I didn't want to write stuff down while they were, while they were doing their studies. (I7)

I4 however noted down timestamps to mark moments of interest, thereby using at least basic backlinks.

> I basically intended to, that if important things come up, you don't write down what happens, but s-, so at least for example the timecode [...] [to] keep the conversation flowing. (I4)

Backlinking requires anticipation.

While backlinking is a straight-forward technique enabling targeted navigation, it requires the researcher to anticipate that they will want to revisit that point in the future. Therefore it is of limited use in an exploratory analysis where the analysts discover what they want to investigate during the analysis.

3.4.3 Implicit study structure

If no explicit links exist, analysts can make use of their knowledge of the study structure to navigate to points of

interest. I5's participants were exposed to a disturbance every three minutes. Naturally, these events were moments of interest. They were easy to find once the start of the study was known, since it was easy to jump ahead by 3-minute intervals in a recordings.

Participants used the study's structure to navigate.

Similarly, I5 reported on a different study for validating an artifact. If in the interview after the study the participant reported problems in a specific task, I5 could use knowledge about the length of the tasks to find the mentioned one. For example, if the first task took 5 minutes and the second task 10, then the third task would be roughly at the 15-minute mark from the start of the study.

3.4.4 Detectables

Qualitative research questions cannot be answered with the data directly, even if they are not exploratory. This type of data requires interpretation before it is of use to the researchers. Words and behavior are difficult to measure directly and usually need to pass through the analyst's head, before they can make sense of it.

With some familiarity with the recordings, however, the analyst tends to notice patterns that indicate points of interest and thus make investigable sections in the recording easier to find. These patterns are like symptoms; they indicate the presence of condition that is not directly observable. Our interviewees surfaced a few examples where there was a pattern reliable enough that they focused on looking for that pattern specifically. We call such a symptom or pattern a *detectable*, because it helps the analyst more easily detect the investigable sections.

Sometimes the investable sections can be identified by detecables.

Consider the following example that is inspired by a use case mentioned by F9. To understand how people use a voice assistant,[5] we study its use in the participants' homes. We make two recordings, one from a microphone in the room that records what the participants say and one from

[5]Examples of voice assistants are Apple's Siri, Google Assistant, and Amazon's Alexa.

the voice assistant's speaker that only contains what the voice assistant outputs.

Naturally, all interactions with the voice assistant are investigable sections, but the recordings are long and the assistant is only used occasionally; finding these sections is difficult. However, the speaker recording only contains the voice assistant's output. By viewing this recording's waveform,[6] it is easy to spot when the assistant speaks, because the waveform can show the difference between silence and sound graphically. In this case, the hills in the waveform are detectables for the sections where the assistant responds which we want to investigate.

Detectables can take many different forms.

We intend detectables to be a general concept and take many different forms. For example I7 scrubbed quickly through the video looking at a screen recording to find a specific task. This was possible, because they knew how the screen would look in that task, and because looking at the screen recording let them know whether to go back or forward to find that task. In this situation, one can describe the specific screen layout they were looking for as a visual detectable.

Detectables depend on the right perspective on the data.

Many things can be detectables, as long as there is "perspective" on the data that makes them stick out and thus easy to find. This is where we see potential for improvement; researchers should have the flexibility to extract and filter their data to enable new perspectives. For example, without a waveform, the voice assistant example would lack a detectable. There would have been no efficient way to find the assistant's responses other than listening to the entire recording. Visualizations and other transformations of the data can provide the analyst with a perspective that makes the investigable sections detectable.

Similarly, it is feasible for software to automatically mark the investigable sections given a suitable detectable. We will give an example of a detectable that can be found by a machine in Section 5.2.

[6] A waveform is a visualization of audio that is flat when there is silence but bulges depending on the loudness of the sound.

3.4.5 Reportables

Towards the end of the analysis, some participants looked for something different. When for example writing their paper, they searched for illustrative examples to support their argument, which we will call *reportables*.

At the end of the analysis, participants looked for reportables.

It is helpful to separate reportables from detectables, because the search for them differs substantially. Depending on the degree of exploratoriness of the analysis, the analyst might not yet know what detectables they will select. But reportables are illustrations of a specific argument, so the analyst will have a clear idea of what they are looking for. Additionally, they will usually have structure in place that helps them easily navigate to reportables, for example timestamped notes, codes, or simply remembrance.

Therefore, finding reportables is much less an issue than finding detectables.

Reportables are typically easy to find.

Examples for reportables are quotes (E1, E2, E3, I5), images or screenshots (E1, I5, I8), and video clips used to edit a summary video to go alongside a paper (I5).

Chapter 4

Validation survey: Common analysis approaches

The interviews suggest that there are two major approaches in analysis of audio and video in HCI: text-based analysis of a transcript and direct analysis. We wanted to quantify how many of these analyses relied on a transcript, and how many needed to work directly with the recordings.

We further wanted to quantify how often such analyses are exploratory. While it is difficult to measure, we were aiming to generate insight into how much flexibility analyses software needs to support.

Lastly, we wanted to understand what data in addition to the recordings was available. Our interviews and literature review suggested that notes and transcript are important sources of information, but that sometimes additional data is collected and used in the analyses.

In short, our guiding research questions for the survey on qualitative analyses involving audio or video were:

- How many analyses use a transcript versus analyze the recordings directly?

- How exploratory are analyses in HCI?
- What is the role of notes, and what additional data available?

4.1 Method

We recruited mainly CHI '19 authors.

The survey took the shape of a Google Form and was distributed to mailing lists and some of the participants of the interviews. Additionally, we sent personalized emails to 245 CHI '19 authors that reported audio or video analyses in their paper. The papers were selected based on an analysis of CHI '19 papers that produced a data set containing information on what type of recordings were used in the study [Reinhard, 2020].[1] From each of these papers' PDFs we extracted the email address of the first author using a script by Wacharamanotham et al. [2020]. Emails that bounced were not resent, unless the response gave indications of a new email address.

We excluded people who had given feedback on the survey, to prevent any bias arising from having dealt with the questions in more detail.

The survey was to be filled out with one specific, recent analysis in mind. One participant could have filled out the survey multiple times for multiple analyses.

No measures were implemented to detect duplicates; we cannot rule out malicious participants or answers from multiple authors of the same paper as a result of forwarding of the invitations.

The survey ran two weeks and had 66 participants.

We closed the survey after two weeks, totaling 66 participants. Some participants responded to the invitation email indicating that they saw themselves as unfit to participate because they transcribed the audio and thus did not work with audio.[2] We responded to such feedback by explicitly

[1] For details on how we filtered out the papers, see Appendix B.2.2.

[2] Note that this might indicate a bias where researchers who analyzed a transcript of the recordings did not participate, and thus transcription-

inviting them to participate in the study.

Note that we will not present the answers in the order their questions appeared in the survey. Instead we present them in a order that best communicates the findings.

4.2 Population

Though we had no system in place to detect whether the answer came from a CHI '19 author or from our general reach-outs, we strongly suspect that the majority of participants are CHI' 19 authors, because 17 authors responded to the invitation email confirming their participation.

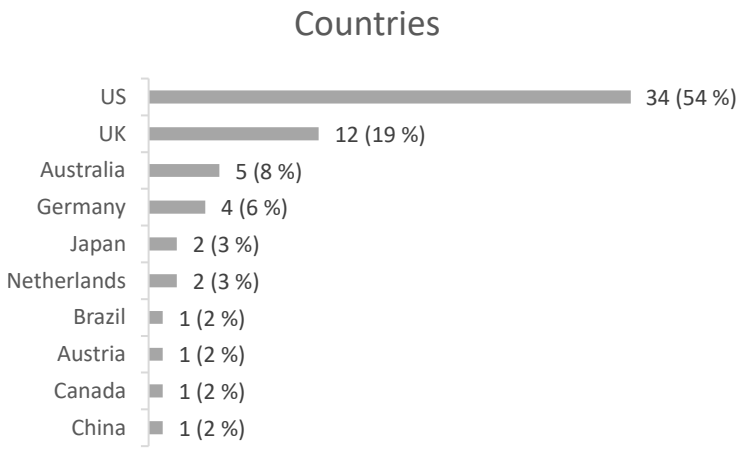


Figure 4.1: "Which country were you working from during the analysis?"
Free text input. Manually cleaned up answers. 63 out of 66 participants answered this question.

The options for the question "What was your status at the time of the analysis?" (Figure 4.2) were

- "Bacholor's student",
- "Master's student",

based analyses might be underrepresented.

- "Ph.D. student",
- "Professional academic researcher. For example post-doc, professor",
- "Professional industrial researcher",
- and a free text input.

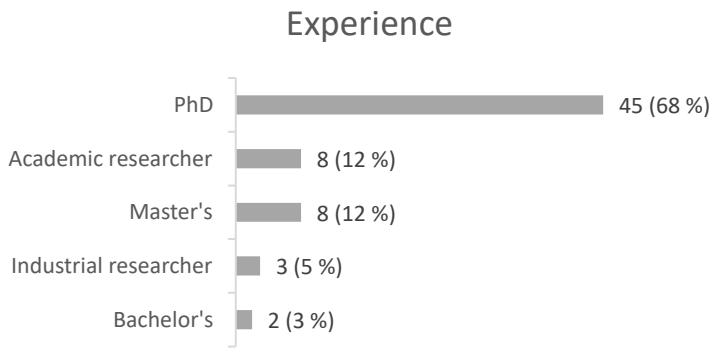


Figure 4.2: "What was your status at the time of the analysis?"
Single choice. All 66 participants answered this question.

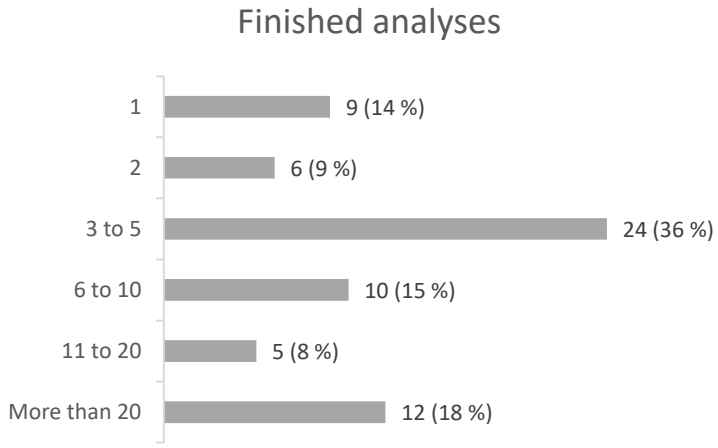


Figure 4.3: "How many analyses involving audio or video have you finished in your career as of today?"
Single choice. All 66 participants answered this question.

We reached people with different nationalities, though the majority worked from the US or the UK (Figure 4.1). $65\,\%$

were PhD students, with a quarter being academic researchers (e.g. professor) or Master's students (Figure 4.2). Only a quarter had finished fewer than 3 analyses, while most had finished 3 to 5, and 18 % accomplished over 20 (Figure 4.3).

The average participant was a PhD student from the US with 3–5 finished analyses.

4.3 Audio versus video

One of our question quantified whether analyses involved audio or video.

"Which type of recording was available for the analysis? (All types of video are included, for example, screen recordings and camera video.)"

Participants could choose one of the following answers.

- Only video. (No audio at all.)
- Only audio. (No video at all.)
- Video and audio. For example, video with sound, or video and separate audio recording.

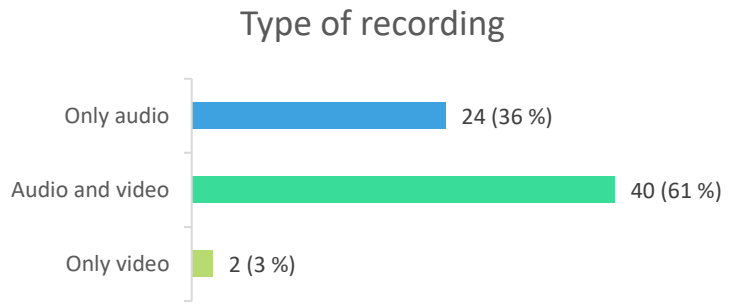


Figure 4.4: "Which type of recording was available for the analysis?"
Single choice. All 66 participants answered this question.

61 % involved both audio and video recordings, while 36 % used only audio (Figure 4.4). The remaining two answers

Most participants had both audio and video recordings, a third had only audio recordings.

(3 %) are left out from most of the analysis, since they are too few to offer significance.

There is a difference in focusing on audio and focusing on video.

This question suffers from ambiguity, because for those who answered "video and audio" we do not know whether the video played an important role in the analysis. When visualizing the answers to other questions we often see two peaks for those who answered "video and audio". We therefore suspect that there are two subpopulations:[3] Those who focused mainly on the audio and the video as safety or for clarification, and those who analyzed the video.

For an example from our interview study (Chapter 3), I4 and I7 recorded both audio and video for an artifact validation, but the video's role was to provide context for ambiguous statements such as "I don't understand what this button does." In such a case, the video does not play an important role and the focus is on what the participants said.

An example for focusing on the video is I8. They had video recordings which of course included audio. During their analysis, they extracted small moments from the recordings and watched these clips many times. Here the video clearly does not play an auxiliary role, but is central to the analysis.

4.4 Type of research question

We asked participants what kind of analysis they had done. We identified three categories based on our interviews and literature review:

- "*Theory building*. For example, modeling people's behavior when using a piece of technology, or understanding the meaning of objects in people's lives."

[3] Technically there is a third possible group. Since many devices that record video also record audio, it is possible that in some analyses only the video was used, and the audio was simply ignored. We only see two peaks in some graphs, not three, which supports our intuition that this case is rare.

- "*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."

Participants could choose multiple options, including giving a custom, free text answer.

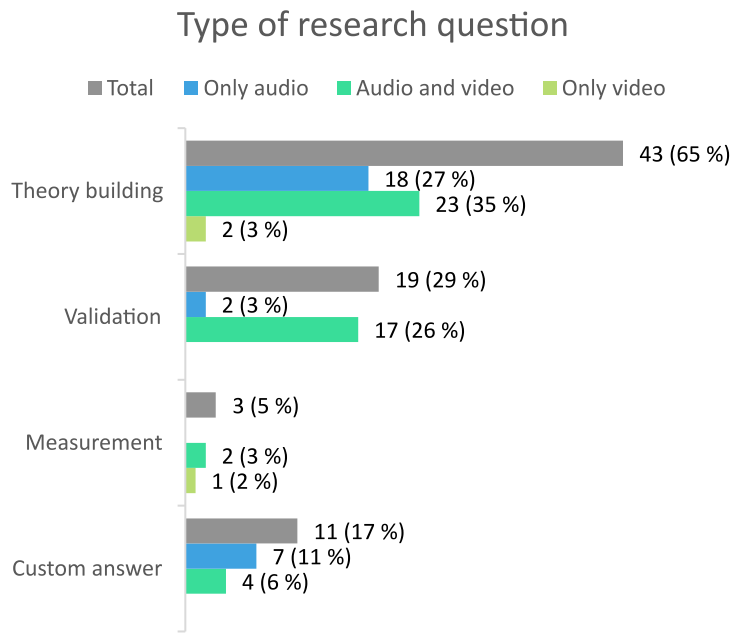


Figure 4.5: "What type of research question did you investigate in your most recent analysis?"
Multiple choices.

Most analyses are theory-building.

The majority of analyses (65 %) were theory building and only 29 % were considered validations (Figure 4.5). There were only 3 % of measurement studies. 9 participants checked more than one answer.

For 17 % of studies participants gave a custom category. The descriptions are short, which makes it difficult to correctly interpret them on our end. For that reason, even if

the answer sounded like it could fit the given categories, we chose to still only count it as a custom answer.

Of those 11 alternative categories, we interpret 4 to have an aspect of theory building, 3 to involve participatory design, 2 studying participant's perspectives, 1 proposing the term "empirical documentation," and 1 that we do not understand well enough to interpret.

Most validation studies use video and audio.

Grouping the answers by what type of recording they had available uncovers that the vast majority (89 %) of validation studies involve both audio and video recordings.

4.5 Exploratoriness

The question of how exploratory an analyses is cannot easily be posed in a survey. We opted to use qualitative statements describing aspects indicating exploratoriness and having participants rate these on a Likert scale [Likert, 1932].

For an overall impression of exploratoriness, we had two statements that are opposites of each other, in a sense:

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."

The key aspect of exploratory analyses that is relevant to this thesis is that exploratoriness prohibits planning. If the concrete measures and indications that help shape answers to the research questions are not known before or during data gathering, the data can be in an inadequate format, complicating the analysis and requiring manual extraction.

The previously listed statements gauge whether this is the

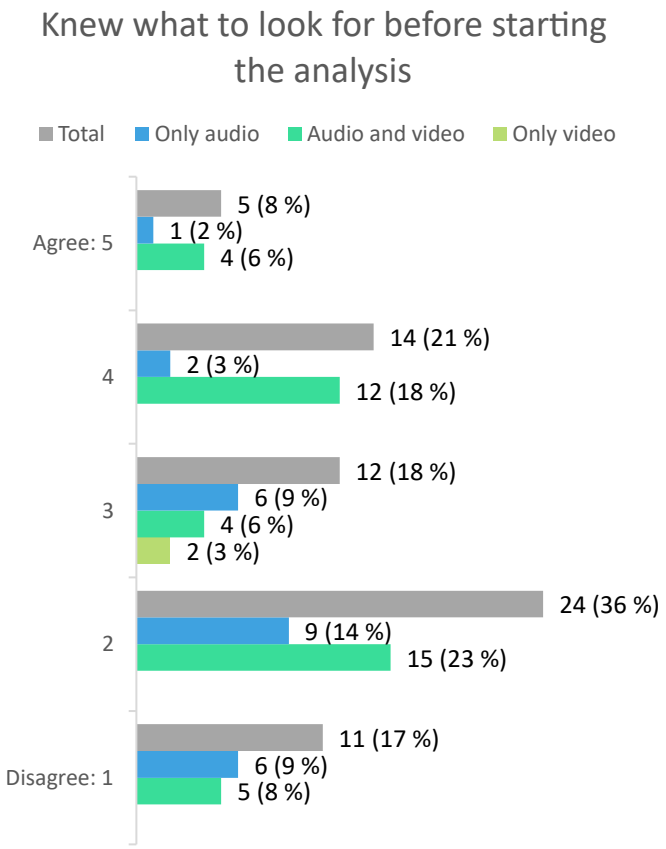


Figure 4.6: "Before starting the analysis, I knew clearly where in the recordings to look and what concretely to look for."
Likert scale. All 66 participants answered this question.

case and we expected their answers to be roughly opposite; if a participant did not know points of interest before the analysis, they should also have to watch the recordings before finding them.

Most analyses are exploratory.

Statement 2 indeed shows that the majority (63 %) indicated a need to observe the recordings before being able to formulate points of interest (Figure 4.7). In line with this, in statement 1 the majority (54 %) indicated not knowing points of interest before starting the analysis (Figure 4.6). However, in statement 1 there is a peak at rating 4 indicating that 21 % of participants *did know* what to look for. This peak is miss-

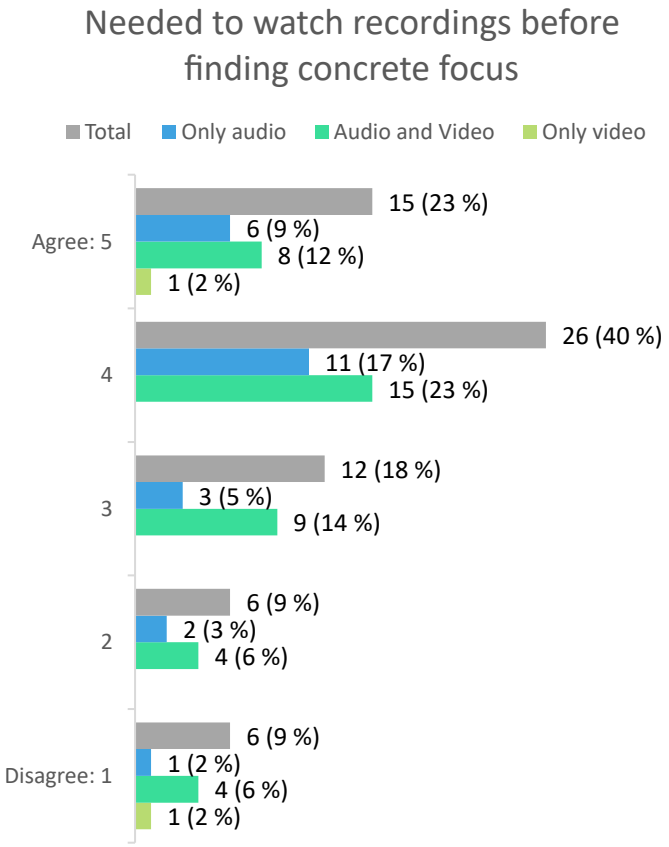


Figure 4.7: "Only after I watched (some or all of) the recordings did I find concrete events or aspects to focus the analysis on."
Liker scale. 65 out of 66 participants answered this question.

ing in statement 2; only 9 % gave the corresponding rating of 2.

Sometimes the research questions are predetermined, but the recordings still need to be watched.

This discrepancy between the answers can be explained by the first statement asking about the research question and the second involving the recordings: Analysts might have a predetermined research questions, but need to observe the recordings before being able to extract information to answer these research questions. This would lead to the single peak in statement 2. Note that this a rationalization which we cannot further substantiate.

4.5.1 Fixed coding scheme

Additionally we asked participants to judge whether they had a fixed coding scheme:

"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."

This question aimed to quantify how often the coding scheme needs to be evolved throughout the analysis, which requires supportive tools.

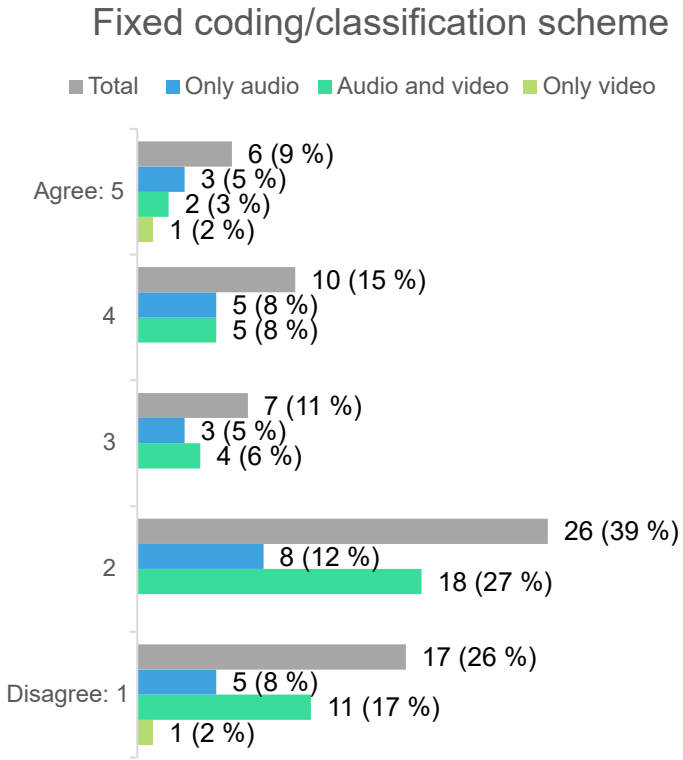


Figure 4.8: "I had a coding or classification scheme (from the start or after observing some of the recordings) that was mostly fixed and that I applied in the rest of the analysis." Liker scale. All 66 participants answered this question.

65 % indicated that their scheme was not fixed (Figure 4.8). Only a quarter had a somewhat fixed scheme. This finding

Most participants did not have a fixed coding scheme.

underlines the importance of supporting flexibility during coding.

4.5.2 Need for extraction or transcription

We also let our participants rate how much they needed to prepare the recordings for analysis:

"All information necessary for the analysis was first extracted from the recordings into a more convenient form for analysis."

This question was meant to quantify the reports of audio and video being tedious to analyze, and therefore transcribed or extracted into smaller clips.

Most participants need to prepare the recordings for analysis.

58 % of participants strongly agreed with the statement (Figure 4.9). Only 15 % indicated that they did not extract the information for the analysis. This underlines the difficulty of working directly with the recordings.

4.5.3 Custom answers

21 % elaborated with a comment.

Since these statements may not apply well to all analyses, we provided participants the opportunity to elaborate. 14 participants (21 %) wrote such a comment.

7 participants explained that an analysis may have both exploratory and confirmatory elements. 3 of them mentioned that they deliberately let their themes "emerge" from the data.

One answer explained that in their case the question about whether all information was extracted before the analysis is difficult to answer. They did use a transcript for an initial analysis, but the "heavy lifting" was accomplished by rewatching and coding videos manually.

Another comment indicated that not only the audio recordings were of interest. Other data and "data analysis activi-

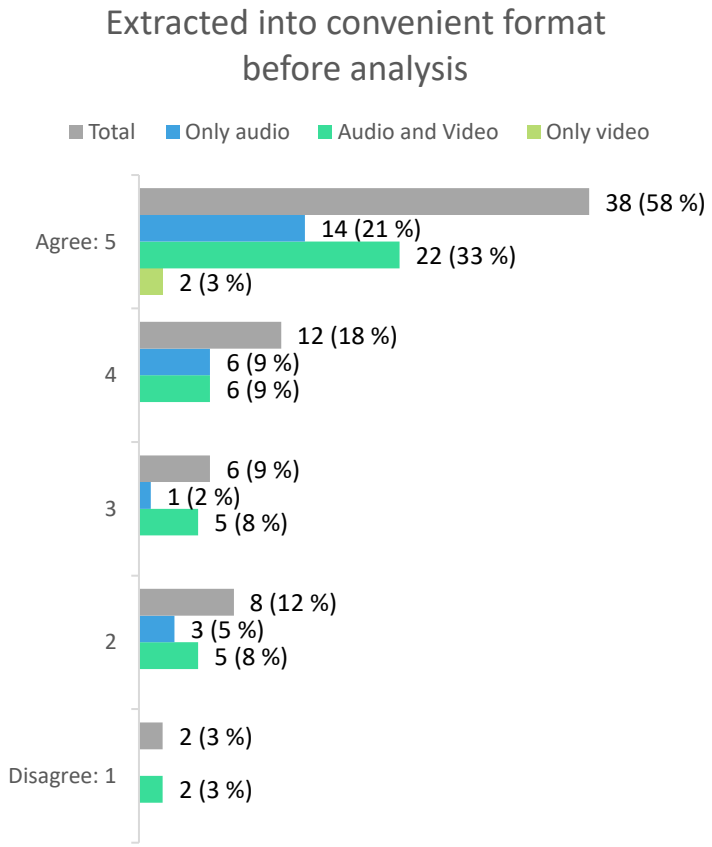


Figure 4.9: "All information necessary for the analysis was first extracted from the recordings into a more convenient form for analysis."
Likert scale. All 66 participants answered this question.

ties" such as debriefings or analysis workshops among researchers played an important part in the overall analysis. This resembles the process that I8 reported.

One commenter who had disagreed (rated as 2) with the statement about having a fixed coding scheme qualified their answer. They explained that they eventually settled on a fixed scheme, but "only after iteratively coding most of [their] data."

Another participant commented that because they had collected all the data, it was easy to "identify interesting

pieces" in their data. Their collaborators, though, struggled to "build an idea of what the data looked like" or understand what our commenter had in mind for the analysis.

After rating all four statements with 2, one participant explained that they used field notes and system logs to identify moments of interest and then "looked for those places in the video recordings."

A commenter suggested that the type of analysis—they give Grounded Theory as an example—might have been of interest for the survey.

One participant had difficulty answering the questions, because they had a participatory study in which the video recordings were produced by their subjects and then discussed in an interview. The analysis studied the interview, but not the recordings directly. This is the same situation as the one F10 reported.

4.6 Notes

We wanted to understand what kinds of backlinks (Section 3.4.2) participants used in their notes.

"While analyzing the recordings, how did you use notes to help navigate to moments of interest?"

Participants could choose one of the following statements about their notes, or give their own alternative.

- "No such notes were available for my recordings." (*No notes*)
- "The notes contained (approximate) timestamps." (*Timestamps*)
- "The notes were chronologically ordered (and did not contain timestamps)." (*Ordered*)
- "The notes did not have a special order or timestamps." (*Simple notes*)

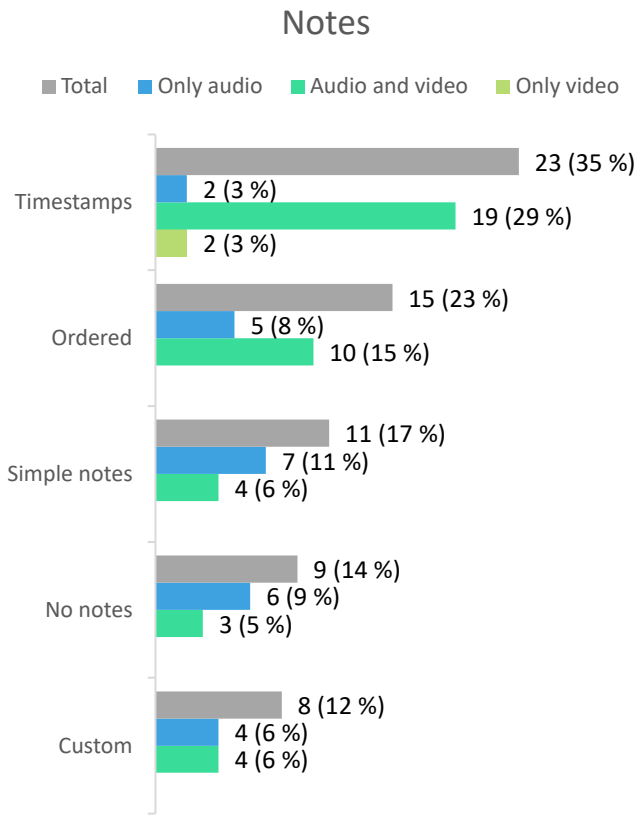


Figure 4.10: "While analyzing the recordings, how did you use notes to help navigate to moments of interest?"
Single choice. All 66 participants answered this question.

Two answers were cleaned up: One reported that the notes contained timestamps *and* were chronologically ordered, and it was counted as only "Timestamp" instead. The other reported that the notes were chronologically ordered and separated into tasks, thus "giving a rough sense of timing", and it was counted simply as "chronologically ordered".

Unfortunately the question might have been interpreted to include analysis notes when it was intended to ask about field notes. Nonetheless the results showcase different techniques being used.

The question is ambiguous: Field notes or analysis notes?

The participants who had only audio recordings tended to have no notes or unstructured notes, whereas those who worked with video tended to have timestamped notes (Fig-

Only audio was unstructured, audio and video was structured.

ure 4.10).

Of the eight custom answers, one mentioned that the notes contained screenshots taking during the session that marked important moments.

Six custom answers mentioned what they coded, transcribed, or other general statements about their analysis or notes, which did not fit the our intent with the question. One participant stated explicitly that they did not understand the question. These misunderstandings may be due to the ambiguity in "notes" in whether field notes or analysis notes are meant.

4.7 Transcription

The study contained two questions that aimed to illustrate the role of transcripts in qualitative analysis. We expected there to be analyses that rely solely on the transcript, whereas others would be concerned with details that cannot be adequately captured in a transcript.

First, we asked participants, how much of their recordings they transcribed.

"Which type of transcript did you use in the analysis?"

They could choose between the following options.

- "*Full transcript*. For example, all participants' responses."
- "*Partial transcript*. For example, only interesting statements."
- "*No transcript*."

Most participants fully transcribed their recordings.

Almost three quarters (72 %) fully transcribed their recordings. This is a clear indication that text-based analysis is the dominant approach in HCI.

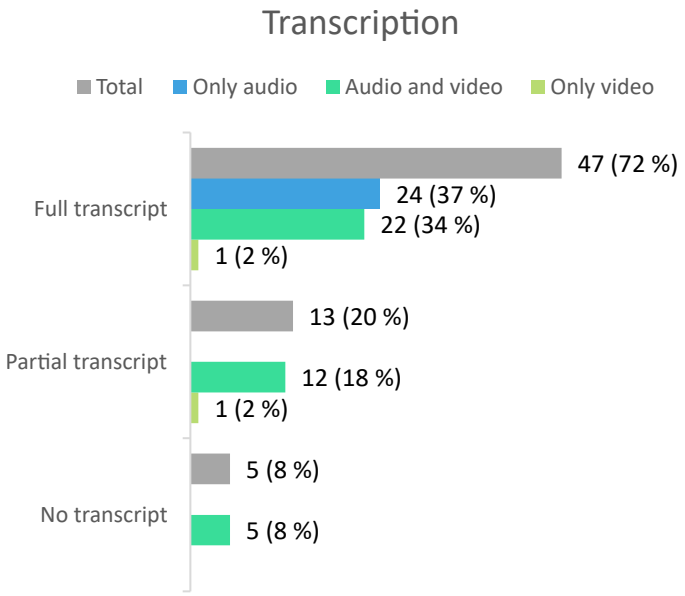


Figure 4.11: "Which type of transcript did you use in the analysis?"
Single choice. 65 out of 66 participants answered this question.

Interestingly, all (100 %) of audio-only analyses are fully transcribed (Figure 4.11). It is likely that the analysts knew early on that they would be focusing on the words, and they chose the according type of recording.

All audio-only analyses had full transcripts.

Note also that there is substantial part of the audio-and-video group (44 % of that group) that does not use full transcription. They are likely to benefit from improved support for navigation of audiovisual recordings and thus potentially part of our target group.

Almost half of audio-and-video does not have a full transcript.

Additionally, we asked how much they used the transcript in their analysis. Our goal was to differentiate whether the participants based their analysis more on the transcript or more on the recordings.

"After you had the transcript, how did you use the recordings for the analysis?"

Participants could choose from the following options or give a custom response.

- "I used only the transcript and did not revisit the recordings at all." (*Exclusively transcript*)
- "I used mainly the transcript and revisited the recordings only to clarify in case of ambiguity or errors in the transcript." (*Recordings for corrections*)
- "The analysis required information not present in the transcript, so I used both the transcript and the recordings." (*Transcript insufficient*)
- "I primarily used the recordings instead of the transcript." (*Primarily recordings*)

If they had a transcript, they used it.

Most of our participants who had a transcript relied primarily on it instead of the recordings (73 % of those with transcript responded "exclusively transcript" or "recordings for corrections"; Figure 4.12).

Transcipts are insufficient for 32 % of participants.

Note that all those for which the transcript was insufficient had audio and video recordings available. Together with the previous question we can calculate how many of our participants needed to employ direct analysis: 5 had no transcript, 15 had an insufficient transcript, and 1 analyzed primarily the recordings.[4] In total, 21 out of our 66 participants (32 %) could not rely solely on text-based transcription.

4.8 Additional data

With the survey we also wanted to get an impression of what other types of data are common in qualitative analysis.

"What additional data was available and did you use it in the audio or video analysis?"

We asked participants about different types of data. The examples were based on our interviews.

[4] The participant who answered "primarily recordings" did in fact report to have a full transcript available, but evidently did not use it.

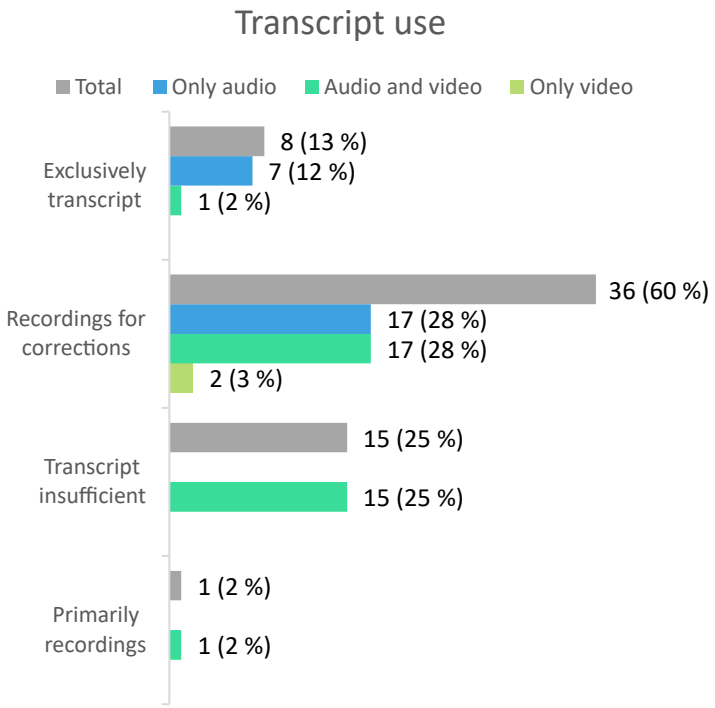


Figure 4.12: "After you had the transcript, how did you use the recordings for the analysis?"
Single choice. 60 out of 66 participants answered this question.

- "System logs. For example, user interactions with prototype, or event log of software system."
- "Time-series data. For example, participants' temperature during study, or noise levels during study."
- "Images."
- "Text produced in the study (not field notes or transcripts). For example, stories the participants wrote."

For each of these types we gave the following options.

- "Used in analysis"
- "Available, but unused"
- "Unavailable"

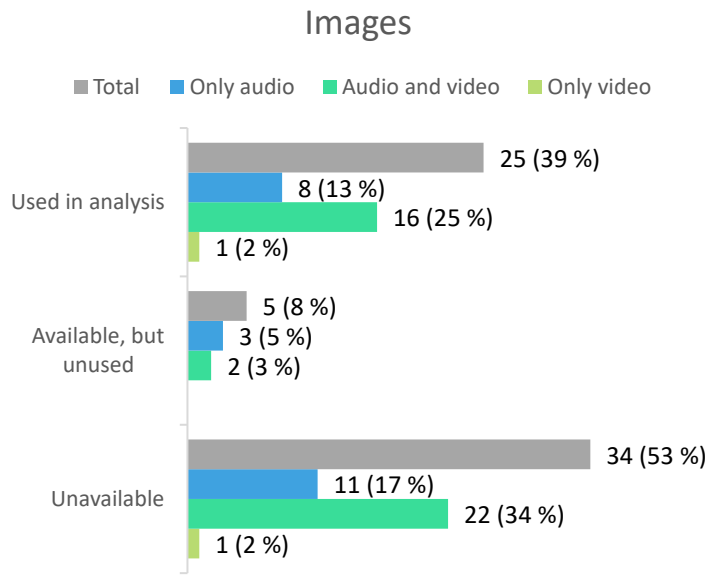


Figure 4.13: Were images available?
Single choice. 64 out of 66 participants answered this question.

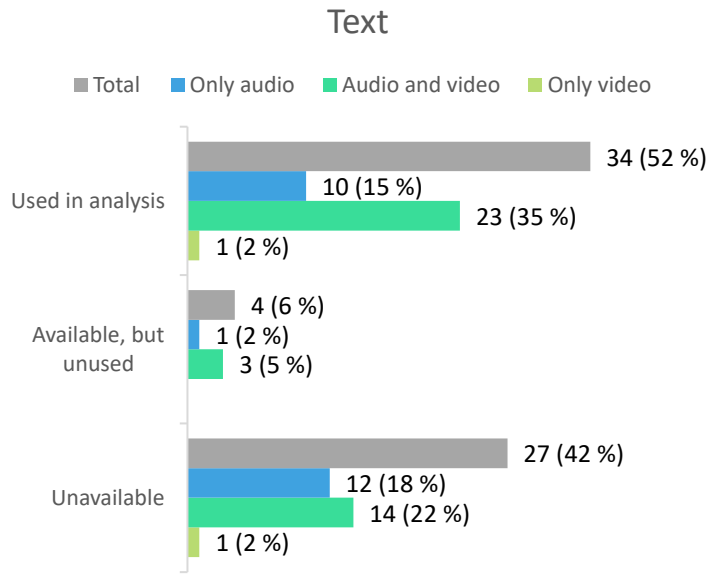


Figure 4.14: Was text produced during the study available?
Single choice. 65 out of 66 participants answered this question.

Images and text are often available.

Images (39 %; Figure 4.13) and text (52 %; Figure 4.14) are commonly used media in an analysis. Time-series data is

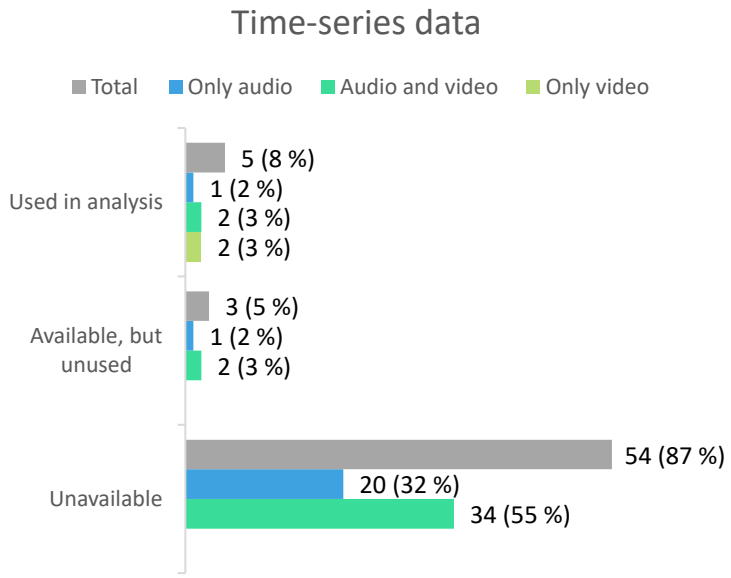


Figure 4.15: Was time-series data available?
Single choice. 62 out of 66 participants answered this question.

much less often used (8 %; Figure 4.15), but this is to be expected because qualitative analyses are used when gathering quantitative data is difficult. When additional data is available, it is rare to go unused (5–8 %).

System logs are quite often available for audio-and-video.

An interesting correlation is that system logs are predominantly audio-and-video analyses (74 % of analyses that used system logs are audio-and-video; Figure 4.16). One explanation is that these are validation studies that record the screen of the prototype. From Section 4.4 we know that 89 % of validation analyses use video and audio recordings. If it is a software prototype, it can be adapted to produce logs which are then available in the analysis. This would explain why system logs are mostly available for audio-and-video analyses. Of course this is hypothetical, we do not have insight into the actual analysis that was reported on.

27 % elaborated with a comment.

The survey also offered the possibility to elaborate on the data used, which 17 participants (27 %) did. 6 participants reported collecting artifacts produced in the study that were not available as options, ranging from a work-

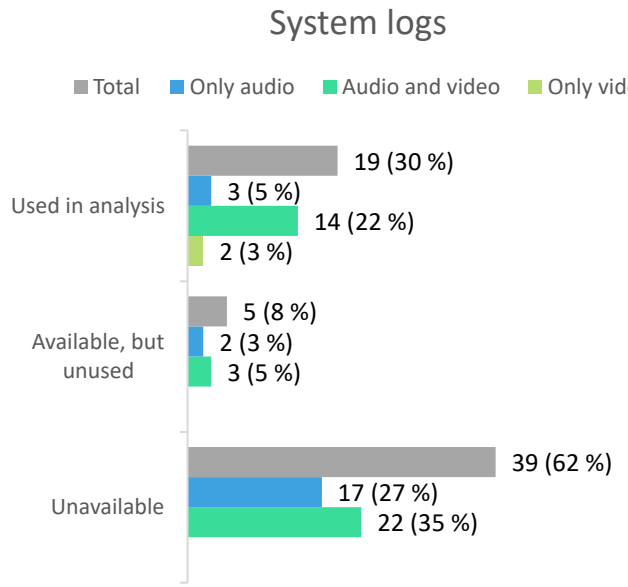


Figure 4.16: Were system logs available?
Single choice. 63 out of 66 participants answered this question.

shop's sketches, to ethnographical source code and documents, to pieces of art presumably created in the study.

One participant mentioned that they used biosignal data, which we assume is the time-series data they reported using.

Two participants reported that they triangulated their findings using notes or a survey. One reported on separately analyzing recordings from the study and a follow-up interview.

Three reported "interview" data, which we suppose may refer to notes about answers to a set of questions, or simply the recordings. One participants had questionnaire data in addition to interviews.

One participant suggested that it "depends on the research question being studied" and that we could have asked for the precise research question.

A participant mentioned that they checked "unavailable"

for images even though they did technically have them available, but "there was no need" to use them in the analysis. They reported looking for reportable images later. They also explained their understanding of "text produced in the study". Unfortunately, the short description was not clear enough to us, so it is hard to interpret; we suspect that they meant their notes about observations which they made "a sort of thematic analysis over." We decided to leave the responses in as they were.

Chapter 5

Design recommendations: Supporting video analysis

With this thesis we wanted to understand the difficulties of analyzing audio and video recordings in contrast to text-based analysis. While 67 % of our participants employed text-based analysis, we found that 32 % needed to work with their recordings directly (Section 4.7). Because text-based analysis is well-supported by established tools, we want to focus on design recommendations to facilitate the task of direct audio and video analysis.

From our studies we identified three important problems in video analysis that should be addressed by analysis software.

- Supporting synchronization and integration of all, possibly heterogeneous, data to enable cross-comparison. (Section 5.1)
- Support for finding detectables, enabling analysts to find the investigable sections more efficiently. (Section 5.2)

- Support for using codes and annotations for navigation, to enable the analyst to effectively filter out only investigable sections to concentrate on. (Section 5.3)

We focused on the automatic finding of detectables and built a prototype based on the multimedia analysis software ChronoViz [Fouse et al., 2011, Fouse, 2013]. It consists of analysis plugins that enable the extraction and annotation of button clicks from a screen recording. (Section 5.4)

5.1 Synchronization

Interview participants reported trouble syncing two recordings.

Some of our interviewees complained about technical synchronization issues. E3 had a video and an audio recording that were saved in separate files and that did not start at the same moment. E3 and F9 explicitly reported this problem and they both had to manually find the offset, calculate the corresponding position in the other recording, and navigate to it.

E3's transcription software did not allow them to import more than one recording, but the audio and video were separate files. Therefore, whenever they needed the video to clarify what a participant was talking about, they were forced to manually find the corresponding moment in the video.

> We did a, so to say, screen capture—of, um—so of what the participants were doing, and recorded audio. But, the problem was that we didn't start, or stop, them simultaneously. (E3)

F9 had similar trouble with looking up information across recordings. In their case, they had to deal with many different files and tried to organize them by creating an index of what file contained which trial from the study.

> Well, I didn't really "sync" sync them up, I just tried to identify the times [...] when the trials

> started in the video. Like trial one started at this time stamp, and then trial two is started at this time stamp. And then—I think GoPro also [...] saves it to multiple files, it's not just one big file. [...] So I then just tried to note down, ok, what file [corresponds to what trial.] (F9)

Many current analysis programs fail to support their users in even the simple cross-comparison need of comparing two recordings of the same event. But there are other flavors of synchronization that can facilitate the analysis.

Synchronized transcripts are helpful.

The literature indicates that synchronizing transcripts to their source recording is helpful. It makes it effortless to read and check the transcript during playback, and vice-versa to scrutinize the recording where the transcript is missing richness [Evers, 2010].

Other data can also be synchronized.

Not only recordings or their transcript can be synchronized. Other times-series data might be available and should be synchronized and integrated for analysis. This idea is at the core of the analysis software ChronoViz [Fouse et al., 2011, Fouse, 2013]. It can synchronize and visualize time-series data to video recordings, and enables direct coding of it all. It explored how to integrate different types of data like digitally recorded handwritten notes, GPS traces, and eye-tracking data, offering specialized visualizations and enabling cross-comparison and annotation.

Unfortunately, ChronoViz is limited to the Mac platform and its features have not yet found their way into more widely available solutions. As such, synchronization—while basically a solved problem—still needs to reach analysts' workplaces.

Synchronization enables cross-comparison.

All in all, synchronization of diverse data cuts down on manual overhead and enables a much broader analysis across all the different perspectives represented in the different sources.

5.2 Finding detectables

One of the motivations for this thesis was the difficulty my supervisor and I experienced in coding video. As computer scientists from HCI, we suspected that we could alleviate some of the pain of manually coding specific events through software.

We found that two of our participants, E1 and I8, had a similar use case. They both had a screen recording of an interface and they were interested in an interaction with the interface that was detectable in the screen recording. However, they needed to manually find all the investigable sections. Using an imaginary example that is based on these use cases, we want to show how software can help the analyst in this situation. We will show a prototype implementation that supports this example in Section 5.4.

Software can help find detectables.

Imagine that we want to study how people use Wikipedia.[1] We invite participants and ask them to research for example what the first video game ever created was. As they use Wikipedia to answer this question, we record their screen. Later, we analyze the recordings and start to notice that participants sometimes get into a dead end and backtrack to a previous page. It appears that there are different situations where this happens. One participant went back when they realized the page was not relevant for answering the question, another got distracted for a short while and realized that they needed to go back to answering the question.

We now want to investigate what the different reasons our participants have for backtracking. There is a detectable: When the participants go back, they do so by clicking the browser's back button. The button gets darker when it is clicked. With software we can detect the button's brightness and automatically detect any sections where it is dark. It is now easy to find all the button clicks and in turn look at the moments leading up to each click.

As our use case shows, it is possible to operationalize de-

[1] Wikipedia is an online, crowd-sourced encyclopedia. (`https://www.wikipedia.org/`)

tectables in a manner that allows the machine to extract them in analyst's place. This relieves the researcher from the chore of finding all investigable sections and allows them to focus on making sense of the findings.

Detecables can be operationalized in a machine-friendly way.

In addition, automatically extracting detectables can encourage analysts to try out more alternatives. If finding all instances of an detectables is cheap, the analyst might opt to see if they find something interesting, whereas traditionally the effort would have been too great.

Help finding detectables may encourage exploration.

The concept of automatically extracted information is not novel. A similar technique can be seen in MultimediaN, a concert video browser that automatically marks segments containing instrumental solos, applause, and visualizes excitement levels, all of which are points of interest when browsing concert videos [Naci and Hanjalic, 2007]. Relatedly, ChronoViz allows analysis of the imported data. They give the example of analyzing a flight simulation with altitude readings. ChronoViz can mark all the periods of time that the altitude is above a threshold via a plugin.

What we contribute is the insight that the detectables might only develop during the analysis. The altitude data was collected before the analysis, but when the analysis is exploratory, the detectables can usually not be planned (Section 3.4.4). In such a case, the analyst might still be able to recover usable information from what is available to them; e.g. from a screen recording or by combining different sources. We suggest that enabling the analyst to retroactively extract information offers them more options for finding and exploiting detectables. In turn, this has the potential to make the chore of finding investigable sections substantially more efficient.

Detectables cannot always be planned.

Note that analysts require transparency from the tools they use in the analysis [Marathe and Toyama, 2018]. Our participants made no explicit comments about this, but in the exploratory interviews we asked whether they would let others transcribe their recordings for them. All three were hesitant and argued that transcription requires expertise. Especially E2 and E3 wanted to keep the partial transcription in their hands, because it involves the prefiltering ex-

The tool needs to be transparent, no magic.

plained in Section 3.3.3. All in all, the process of extracting data and finding detectables needs to be transparent and stay in the control of the analyst. Otherwise they will lose trust in the system and the resulting findings.

Automatically finding detectibles makes analysis more efficient.

Enabling the analyst to extract and find detectables solves one of the most pressing problems in direct analysis: Navigating the recording to focus on the parts of interest.

5.3 Navigation of annotations

Implicit knowledge cannot be exploited by software.

In the interviews we found that analysts make use of implicit structures (Section 3.4.3). Without making such structures explicit the program is unable to aid them. For instance, it cannot automatically skip over irrelevant parts, such as study setup, in the recording, and it cannot juxtapose different recordings of the same task.

Annotations should enable better navigation.

Such sections could be marked with annotations. So why did we not encounter this behavior in our interviews? We suspect this is because annotations do not give any navigational benefits and are used exclusively to organize the meaning of the data.

Annotations currently only afford basic navigation. Analysts can list all annotated segments in MAXQDA or play an annotated segment in a loop in ChronoViz. But for example the need to skip irrelevant sections is not met by these features.

Software should allow filtering and combination of annotations.

To make such annotations useful for navigation, the user needs to be able to filter and combine them. If they could mark a section as setup and then hide it or skip it during playback, they could focus on the more relevant parts of the recording. And if they marked their study's second task across all recordings, they might want to juxtapose all instances of that task. They might want play all recordings in parallel to compare the progress participants made in that task, or they might want to compare the annotations of that task to discover higher-level patterns. The ability to filter out and combine marked segments in such ways would

give analysts a reason to use annotations for bringing the implicit structures in their head into the world.

There already are techniques that aim to make such structures explicit. Consider the "scratch-off" metaphor that highlights parts of the recording that were played more frequently [Fouse, 2013].[2] The playback behavior, even if it is not a conscious structure, can for example indicate what pieces of the recording are irrelevant to the analysis by automatically leaving them greyed out.

The "sratch-off" metaphor makes irrelevant sections explicit.

5.4 Prototype

We created a prototype that serves as a proof of concept mainly for the extraction of detectables. It extends ChronoViz [Fouse et al., 2011, Fouse, 2013], an analysis program that enables integration of many types of data in video analysis, which already implements our synchronization recommendation (Section 5.1) brilliantly.

We extended it by using the plugin mechanism it provides that allows automatic creation of annotations via Python. Our plugins enable the analyst to extract the average brightness of a region of the video and annotate when some data stays inside some bounds. This is an implementation of automatic detectable extraction (Section 5.2).

We extended ChronoViz with plugins for finding detectables.

In addition we prototyped a very simple navigational feature that allows playing back all segments of an annotation in sequence, skipping all sections that are not marked with such an annotation. This offers a navigational benefit based on annotations (Section 3.4.3), which might encourage users to make implicit structures explicit.

[2] The scratch-off metaphor is inspired by the "edit-wear" highlights [Hill et al., 1992] indicating which parts of a document were edited most.

5.4.1 Motivational use case

We developed the plugins with a specific use case in mind, as illustrated in Figure 5.1.

Our plugins enable finding button clicks in a screen recording.

In the Wikipedia example from Section 5.2, the analyst's goal is to find all the button clicks, to investigate what lead to those clicks. They notice that the button gets darker when it is clicked ①. The button's darkening is a detectable that they can find using the our plugins. They first extract the average brightness of the button by using the "Extract average brightness in selection" plugin. They draw a box around the back button to extract the average brightness from ② and hit the "Extract" button. The plugin runs and then the average brightness is visualized as a graph in the timeline ③. The graph makes the button clicks easy to find as there are visible drops in the average brightness whenever it is clicked.

The analyst stays in control.

However, the analyst notices a few false positives where the dark mouse cursor crosses the button and makes the average brightness drop. They decide to not rely on the graph and use annotations to mark the button clicks. They use another plugin, "Annotate inside bounds", and instruct it to annotate all sections where the average brightness is between 0.7 and 0.8. They then check every annotation and delete the false positives. In the end, they are left with annotations that precisely mark all button clicks ④, enabling them to easily investigate what lead up to each of them.

5.4.2 Implementation

ChronoViz offered a plugin architecture that already supported much of what we aimed to do. Code written in Python could inspect the available data as well as existing annotations and create new data or annotations that were then available to the analyst.

ChronoViz needed to be updated to run on the newest Mac OS.

First we needed to update ChronoViz to run on the current version of Mac OS, because Apple had removed APIs that ChronoViz relied upon. Adam Fouse, the creator of

Figure 5.1: Demonstration of using our prototype to detect back button clicks: ① In the screen recording, the back button gets darker when it is clicked. ② In a plugin we draw a green rectangle around the back button. ③ The plugin extracts the average brightness of that region. Clicks of the back button are visible as valleys in the graph, because the average brightness gets darker when the button is clicked. ④ We can use an additional plugin (not shown) to annotate all sections where the average brightness is between 0.7 and 0.8. The black annotations mark all button clicks.

ChronoViz, helped the migration towards the new media APIs. We decided to remove features that were not necessary for the prototype to reduce the migration effort needed. The migration introduced a few bugs and problems, but we managed to get ChronoViz into a usable state on the current Mac OS version, "Catalina".

We adapted the plugin system for our needs.

We then started to create plugins that would benefit the use case we imagined. The initial plugin architecture had intended that all parameters would be numbers and input before the analysis, by registering the inputs in Python while they were implemented in Objective-C. But a key plugin we wanted to develop needed to allow the user to select a region of the video. We thus changed the architecture

to allow the Python plugin to fully control its interface by exploiting PyObjC's[3] ability to call into Apple's UI toolkit, "Cocoa".

We implemented playback of only the annotations belonging to a category.

We also added a feature that allowed selection of an annotation category. During playback it would only play the annotations belonging to that category and skip all sections of the recording that were not annotated thusly.

5.4.3 Limitations

Unfortunately there are a few important limitations that hinder the productive use of our extensions.

Plugins freeze when they run.

The most important restriction is that we could not find a way to make the execution of the plugins asynchronous. This is especially relevant for the plugin that extracts the average brightness of a region, as it is quite slow. The interface will freeze up for the duration of the plugin's runtime, without any progress indication.

The prototype was not tested.

We did not test the prototype in a user study. As such it servers mostly as a showcase of our design recommendation to automatically find detectables. We discuss how to address this limitation in Section 6.2.2.

[3] PyObjC is a Python library that enables calling directly into Objective-C functions and that makes available many of Apple's APIs in Python. Documentation is available at `https://pyobjc.readthedocs.io/en/latest/`.

Chapter 6

Discussion: The future of video analysis

We have presented how our work motivates design recommendations for tools supporting qualitative video analysis in HCI.

Going further, our findings can be applied to other fields and for other activities than analysis. (Section 6.1)

We recommend future work to address limitations and remaining questions from our work. (Section 6.2)

Lastly, we conclude by summarizing the key findings and contributions we made. (Section 6.3)

6.1 Extended applications

Our recommendations can benefit other fields than HCI.

The design recommendations from Chapter 5 are not limited to video analysis in HCI. Other fields that rely on direct analysis might benefit especially from automatically finding detectables. Consider biologists who study an animal's feeding behavior in a zoo. They could track its location in

the enclosure and identify when it is near the feeding station, relieving them of watching most of the recording.

The benefits apply to partial transcription.

The benefits also apply to partial transcription, because the analysts there too need to identify the investigable sections to be transcribed. After the researchers have familiarized themselves with some of the recordings, they could exploit automatic search for detectables to identify more easily the investigable sections they want to transcribe for further analysis.

The techniques can be applied in participatory studies.

The fundamental ideas can also be applied to other contexts than analysis in which certain sections need to be identified in videos. Interviewee F10, for example, used a GPS track visualization, video clips, and accelerometer data as prompts in an interview. The video clips were recorded by the participants themselves, and as a consequence all the clips were largely investigable. All data was integrated in Google Earth[1] to show the videos at the places on the GPS track where they were recorded. These techniques are quite similar to what we recommend.

6.2 Future work

There are some considerations to mention both about the limits of our studies and our recommendations. We will suggest how these limitations may be addressed by future work.

6.2.1 Interviews and survey

Our studies focused on HCI because of our familiarity with the domain and access to experts in the field. We suspect that the findings and suggestions will apply to other fields that benefit from direct analysis, such as ethnography.

Generalizability is to be validated.

Whether our findings hold up could be determined with

[1] Google Earth is software that enables exploring the earth virtually.

similar interview probes and by presenting our survey to researchers in other fields than HCI.

6.2.2 Finding detectables

Flexibility of our plugin concept is to be validated.

We realize that the prototype is centered around a single use case. Whether this concept is flexible enough to be of practical use remains to be validated. This is in line with the criticism that software needs to be general enough to be applied over different analyses, whereas each analysis requires focus on novel and peculiar aspects [MacMillan, 2005]. It needs to be studied to what extent researchers can share plugins, and if it is efficient enough to program their own plugins where existing ones fall short.

Time savings are to be validated.

We also cannot make confident claims about how much time analysts spend looking for investigable sections. This is an important motivation for our most novel design recommendation, finding detectables (Section 5.2). We suspect that analysts spend a substantial amount of time manually identifying these sections, but future work would be needed to validate whether this makes up a big enough portion of the analysts work and whether systems like our prototype are effective enough to reduce this work significantly. A study that compares the performance between traditional QDA software and a system similar to our prototype can provide answers to these important questions.

6.3 Contributions

We found that transcription is an important approach to qualitative analysis of audiovisual recordings. Based on our survey, though, we estimate that a third of analyses involving audio or video recordings cannot rely solely on the established text-based analysis approaches. Instead, they need to work with the recordings directly.

`https://www.google.com/earth/`

Our interviews suggest that the analyses in HCI tend to be exploratory not only in their research question, but also in finding ways to answer these questions.

The central task is to analyze investigable sections, but they are difficult to identify in audiovisual recordings. This lead us to introduce the term *detectables*, which captures the tendency of analysts to find instances of abstract concepts through concrete, observable events in the recordings.

We created a prototype that enables researchers to find certain detectables automatically, drastically reducing the effort spent outside of the actual analysis. Furthermore, we recommend analysis software to support synchronization and enable navigation of annotations.

Our hope is that these recommendations will result in improved analysis software that can empower future analysts in their work.

Appendix A

Interview study

The following materials are presented both to offer more insight into the process for the interviews as well as to inform future work and provide templates.

A.1 Recruitment email

The following is the latest template for the email sent to interview candidates in May 2020.

> Dear participant,
>
> I am Johannes, a student at the Media Computing Group, RWTH Aachen. In my master's thesis I want to support researchers like you who have analyzed audio/video.
>
> I would like to interview you for about 30 minutes to understand your workflow in your most recent analysis involving audio or video recordings. It will be great if you can help me out! If you're available in the next week (until Sa, May 16), please choose a free time slot at *link to https://terminplaner4.dfn.de/* or contact me at johannes.maas1@rwth-aachen.de. I will

respond via email to confirm and discuss all further details.

Please consider forwarding this inquiry to people who have done analyses involving audio or video recordings.

Sincerely

Johannes Maas

A.2 Consent form

The following page shows the informed consent form that participants signed before the interview.

Informed Consent Form

Interview on finding relevant parts during analysis of audio and video

Principal Investigator Johannes Maas
Media Computing Group, RWTH Aachen University
johannes.maas1@rwth-aachen.de

Purpose of the study: We explore what analysts look for in audio and video recordings and how they find and extract relevant parts. We want to understand the process in order to improve upon it.

Risks/Discomfort: There are no expected risks when participating in the study. We aim for a 30 minute session, so you might become fatigued. Should you feel uncomfortable or want to terminate, tell us and we will abort the session immediately with no consequences for you.

Procedure: First, we will introduce ourselves and our research questions. Then we want you to explain how you performed an actual analysis. For this it is helpful, though not strictly necessary, that you open the analysis in the software used, to put you “back in the moment”. Subsequently we will look for situations in which you were looking for specific aspects and how you found and used the relevant sections.

Recording: We will record handwritten field notes. Additionally, with your consent, we would like to make an audio recording of the interview that is to be used exclusively for the analysis (e.g. for generating a transcript). We will ask for your permission at the beginning of the interview and before starting the recording.

Confidentiality: All information collected during the study period will be kept strictly confidential. Only the principal investigator, Johannes Maas, and his thesis supervisor, Krishna Subramanian, will have access to the recordings. You will be identified through a participant number. No publications or reports from this project will include identifying information on any participant.

_____ I have read and understood the information on this form.

_____ I have had the opportunity to ask questions and am satisfied with the answers.

By signing, you agree to participate in the study as detailed.

Participant’s Name	Participant’s Signature	Date
	Principal Investigator	Date

If you have any questions regarding this study, please contact Johannes Maas at johannes.maas1@rwth-aachen.de

A.3 Protocol

The first three interviews were exploratory and did not have a protocol.

All other interviews followed this template which was pre-filled on a sheet of paper. Each item would be ticked off when completed and there was enough space to add notes to the sections during the interview.

The follow-up interviews used this protocol as well, but with a condensed execution. Since the recruitment message set a time of 15 minutes, the interviewer aimed to spend less time on the walkthrough of the analysis and more on finding and discussing detectables in those interviews.

1. Introduce research question: What analysts look for and how they find it.
2. Confirm duration: 30 minutes.
3. "Questions before we start?"
4. Start the recording and tell interviewee.
5. Demography: Confirm professional position (e.g. professor) and field (e.g. HCI).
6. Experience: "How many analysis involving audio or video have you finished?"
7. Focus on analysis: "What was the theme of the analysis we will be discussing?"
8. Situate interviewee and get introduction to their analysis: "Please walk me through analysis process." (Data gathering, analysis technique, etc.)
9. Finding and discussing examples of aspects interviewee focused on (what we now call *detectables*).
10. Give rough summary and ask interviewee whether it is correct.
11. Stop recording and tell interviewee.

12. "Any feedback?" (On interview or any further remarks on the subject.)

13. Thank interviewee and wrap up interview.

Appendix B

Survey

B.1 Questionnaire

The following pages present a printout of the survey. Note that the section *Follow up interview* was not shown to participants who answered with either of the first two options of question 12, as indicated by the instruction to skip to question 13.

Understanding analysis of audio and video

My name is Johannes Maas and I am working on my Master's thesis at RWTH Aachen, Germany, to understand and support researchers like you who do qualitative analysis involving audio or video. With this survey you help me gather data on what analysis approaches are common in HCI.

Please think of your most recent qualitative analysis involving audio or video and answer all questions with regards to that one, specific analysis. (If you'd like to contribute information about multiple analyses, you can fill out the form multiple times; once for each analysis.)

This survey is subject to Google's terms of service and privacy policy, linked at the end of the page. In addition to that, the data is only visible to me, Johannes Maas, and my supervisor, Krishna Subramanian. We intend to publish the raw data alongside the thesis, with exceptions for certain questions explicitly stated.

1. What type of research question did you investigate in your most recent analysis?

Check all that apply.

- [ ] 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.
- [ ] Other: _____

How much do the following statements reflect your most recent analysis?

2. "Before starting the analysis, I knew clearly where in the recordings to look and what concretely to look for."

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

3. "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.

   Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

4. "Only after I watched (some or all of) the recordings did I find concrete events or aspects to focus the analysis on."

   Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

5. "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."

   Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

6. (Optional) If one of the previous statements was difficult to rate or you want to add details about your analysis, please explain.

The published dataset will not include your answer, only whether you responded or not.

Available data

7. Which type of recording was available for the analysis?

All types of video are included, for example, screen recordings and camera video.

Mark only one oval.

- ◯ Only video. (No audio at all.)
- ◯ Only audio. (No video at all.)
- ◯ Video and audio. For example, video with sound, or video and separate audio recording.

8. While analyzing the recordings, how did you use notes to help navigate to moments of interest?

If your technique does not fit the following options, please explain it in "Other".

Mark only one oval.

- ◯ No such notes were available for my recordings.
- ◯ The notes contained (approximate) timestamps.
- ◯ The notes were chronologically ordered (and did not contain timestamps).
- ◯ The notes did not have a special order or timestamps.
- ◯ Other: _____

9. What additional data was available and did you use it in the audio or video analysis?

If an answer does not fit here, please see the next question.

Mark only one oval per row.

	Used in analysis	Available, but unused	Unavailable
System logs. For example, user interactions with prototype, or event log of software system.	◯	◯	◯
Time-series data. For example, participants' temperature during study, or noise levels during study.	◯	◯	◯
Images.	◯	◯	◯
Text produced in the study (not field notes or transcripts). For example, stories the participants wrote.	◯	◯	◯

10. (Optional) Was other data available for analysis? Did you use it? Please explain.

The published dataset will not include your answer, only whether you responded or not.

Transcripts

11. Which type of transcript did you use in the analysis?

We define "transcription" to include any form of text describing the recording, including non-verbal information like descriptions of participants' actions.

Mark only one oval.

- ◯ Full transcript. For example, all participants' responses.
- ◯ Partial transcript. For example, only interesting statements.
- ◯ No transcript. (Skip the remainder of this page and press "Next" at the end.)
- ◯ Other: _____

12. After you had the transcript, how did you use the recordings for the analysis?

Mark only one oval.

- ◯ I used only the transcript and did not revisit the recordings at all.
  Skip to question 13
- ◯ I used mainly the transcript and revisited the recordings only to clarify in case of ambiguity or errors in the transcript. *Skip to question 13*
- ◯ The analysis required information not present in the transcript, so I used both the transcript and the recordings.
- ◯ I primarily used the recordings instead of the transcript.
- ◯ Other: _____

Follow up interview

You indicated that you did not use a transcript or analyzed information that was not present in the transcript.

If the analysis involved video, I would like to have a 15-minute interview with you to understand some of the concrete things you looked for in the recordings and how you found them. Based on such examples, I will try to develop ways of automatically finding commonly looked for things, so that you do not have to watch the entire video.

If you are interested, please pick a slot at https://terminplaner4.dfn.de/tH9fmDawZtKp7wJz or send me an email to johannes.maas1@rwth-aachen.de with the subject "Interview about A/V analysis". In both cases I will respond via email to give you more information, so that you can decide whether you want to participate in the interview.

Demographics

13. Which country were you working from during the analysis?

This question will be excluded from the published dataset.

_____

14. What was your status at the time of the analysis?

Mark only one oval.

- ◯ Bacholor's student.
- ◯ Master's student.
- ◯ Ph.D. student.
- ◯ Professional academic researcher. For example postdoc, professor.
- ◯ Professional industrial researcher.
- ◯ Other: _____

15. How many analyses involving audio or video have you finished in your career as of today?

Please include all analyses you've worked on, even if you were not the main analyst.

Mark only one oval.

- ◯ 1.
- ◯ 2.
- ◯ 3 to 5.
- ◯ 6 to 10.
- ◯ 11 to 20.
- ◯ More than 20.

B.2 Recruitment

We want to give an overview of how we recruited participants for the survey.

B.2.1 General recruitment

We sent the following message to a Slack[1] group comprising HCI researchers and our chair's mailing list.

> My name is Johannes and I am working on my Master's thesis at RWTH Aachen, Germany, aiming to understand how HCI researchers like you analyze audio or video recordings.
>
> If you are an HCI researcher (student, Ph.D. student, professor, etc.) and have analyzed audio or video recordings recently (in the last 12 months), I invite you to take a survey about this analysis. It takes 5–10 minutes to complete: *link to Google Form*
>
> Please consider forwarding the survey to colleagues who have analyzed audio or video recordings.
>
> Sincerely
>
> Johannes Maas

B.2.2 CHI '19 authors recruitment

Based on data on all 705 CHI '19 papers [Reinhard, 2020], we identified the following 245 papers that involved audio or video recordings. These numbers are the IDs present in the papers' file names and printed on each of their pages in the proceeding's downloads available

[1] Slack is a chat platform used in organizations. `https://slack.com/intl/de-de/`

at `https://dl.acm.org/doi/proceedings/10.1145/3290605#issue-downloads`.

Note that at the time of recruitment, the dataset was unfinished which is why the highest ID is 630, not 701 as it would be if the final dataset was used.

7, 8, 9, 12, 16, 21, 23, 24, 29, 30, 33, 34, 39, 42, 43, 52, 53, 62, 68, 70, 72, 73, 77, 82, 84, 85, 88, 89, 90, 91, 92, 93, 96, 98, 99, 100, 106, 109, 113, 116, 122, 126, 127, 129, 130, 133, 135, 136, 137, 139, 141, 142, 145, 149, 151, 153, 154, 160, 161, 164, 169, 171, 174, 178, 191, 195, 197, 198, 201, 202, 205, 206, 213, 215, 217, 218, 227, 228, 229, 231, 233, 234, 236, 237, 238, 243, 253, 260, 264, 265, 266, 267, 268, 270, 271, 272, 277, 283, 286, 287, 290, 295, 299, 300, 301, 304, 307, 308, 309, 312, 313, 314, 316, 317, 318, 320, 322, 323, 324, 328, 334, 335, 340, 341, 342, 343, 349, 350, 353, 356, 363, 366, 371, 372, 375, 376, 377, 378, 379, 380, 382, 383, 384, 385, 387, 388, 390, 393, 398, 399, 400, 401, 402, 405, 406, 410, 414, 415, 416, 421, 422, 423, 424, 425, 428, 432, 433, 434, 435, 436, 437, 439, 440, 441, 442, 447, 449, 455, 458, 459, 460, 461, 465, 466, 469, 470, 472, 475, 476, 481, 482, 484, 486, 492, 497, 499, 502, 504, 505, 510, 511, 514, 516, 522, 529, 532, 534, 536, 539, 544, 545, 547, 550, 558, 564, 565, 569, 572, 576, 577, 578, 582, 584, 586, 587, 588, 592, 593, 594, 595, 597, 599, 600, 601, 611, 613, 614, 615, 622, 623, 624, 626, 629, 630.

To obtain these results, we filtered the "Recording: audio / video / screen / other?" column to only show the following values.

"audio", "audio, notes", "audio, photos", "audio, screen", "audio, screen, usage data", "audio, usage data", "audio, video", "audio, video, photos", "both", "EEG data, audio", "photos, audio", "photos, video", "screen", "usage data, audio", "usage data, video", "video", "video, audio", "video, audio, photos", "video, audio, screen", "video, eye tracking", "video, photos", "video, screen"

Filtering the final dataset by these same values gives a total of 283 unique papers that involve audio or video recordings.

We sent these authors an email based on the following template.

> Dear *first author*,
>
> my name is Johannes and I am working on my Master's thesis at RWTH Aachen, Germany, aiming to understand how HCI researchers like you analyze audio or video recordings.
>
> Congratulations on your CHI '19 paper, *paper title*. Since it contains an analysis of audio and video, you would be a great candidate for my survey. (I extracted the first author's name and email from CHI '19 publications involving this type of analysis. In case some error slipped through, I would like to apologize.)
>
> You can contribute to my survey that asks questions about your most recent analysis (maybe the one from your CHI '19 paper). The survey takes 5–10 minutes to complete: *link to Google Form*
>
> If you feel that one of your co-authors can report on the analysis in more detail than you, it would be helpful if you would forward them this email. Please also consider forwarding the survey to your HCI colleagues who have analyzed audio or video recordings.
>
> Sincerely
>
> Johannes Maas

B.3 Anonymized dataset

The dataset can be downloaded from `https://hci.rwth-aachen.de/qualitative-av-analysis`.

Bibliography

Pat Bazeley and Kristi Jackson. *Qualitative Data Analysis with NVivo*. SAGE, Los Angeles London New Delhi, second edition, 2013. ISBN 978-1-4462-5655-8 978-1-4462-5656-5. OCLC: 822959812.

Hennie Boeije. A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity*, 36(4):391–409, November 2002. ISSN 1573-7845. doi: 10.1023/A:1020909529486. URL https://doi.org/10.1023/A:1020909529486.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006. ISSN 1478-0887. doi: 10.1191/1478088706qp063oa. URL https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa.

Joy D. Bringer, Lynne Halley Johnston, and Celia H. Brackenridge. Using Computer-Assisted Qualitative Data Analysis Software to Develop a Grounded Theory Project. *Field Methods*, 18(3):245–266, August 2006. ISSN 1525-822X, 1552-3969. doi: 10.1177/1525822X06287602. URL http://journals.sagepub.com/doi/10.1177/1525822X06287602.

Hennie Brugman and Albert Russel. Annotation Multimedia / Multi-modal resources with ELAN. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Lisbon, Portugal, May 2004. ISBN 2-9517408-1-6. URL http://www.lrec-conf.org/proceedings/lrec2004/.

Richard M. Carpiano. Come take a walk with me: The "Go-Along" interview as a novel method for studying the implications of place for health and well-being. *Health & Place*, 15(1):263–272, March 2009. ISSN 13538292. doi: 10.1016/j.healthplace.2008.05.003. URL https://linkinghub.elsevier.com/retrieve/pii/S1353829208000622.

Juan Casares, A. Chris Long, Brad A. Myers, Rishi Bhatnagar, Scott M. Stevens, Laura Dabbish, Dan Yocum, and Albert Corbett. Simplifying Video Editing Using Metadata. In *Proceedings of the conference on Designing interactive systems processes, practices, methods, and techniques*, page 157, London, England, 2002. ACM Press. ISBN 978-1-58113-515-2. doi: 10.1145/778712.778737. URL http://portal.acm.org/citation.cfm?doid=778712.778737.

Paul G. Dempster and David K. Woods. The Economic Crisis Though the Eyes of Transana. *Forum Qualitative Social Research*, 12(1), January 2011. ISSN 1438-5627. doi: 10.17169/FQS-12.1.1515. URL http://www.qualitative-research.net/index.php/fqs/article/view/1515.

Damien Dupre, Daniel Akpan, Elena Elias, Jean-Michel Adam, Brigitte Meillon, Nicolas Bonnefond, Michel Dubois, and Anna Tcherkassof. Oudjat: A configurable and usable annotation tool for the study of facial expressions of emotion. *International Journal of Human-Computer Studies*, 83:51–61, November 2015. ISSN 10715819. doi: 10.1016/j.ijhcs.2015.05.010. URL https://linkinghub.elsevier.com/retrieve/pii/S107158191500107X.

Alain Désilets, Berry de Bruijn, and Joel Martin. Extracting Keyphrases from Spoken Audio Documents. In Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, Anni R. Coden, Eric W. Brown, and Savitha Srinivasan, editors, *Information Retrieval Techniques for Speech Applications*, volume 2273, pages 36–50. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002. ISBN 978-3-540-43156-5 978-3-540-45637-7. doi: 10.1007/3-540-45637-6_4. URL http://link.springer.com/10.1007/3-540-45637-6_4.

Jeanine C. Evers. From the Past into the Future. How Technological Developments Change Our Ways of Data Collection, Transcription and Analysis. *Forum Qualitative Social Research*, 12(1), November 2010. doi: 10.17169/FQS-12.1.1636. URL http://www.qualitative-research.net/index.php/fqs/article/view/1636.

Jeanine C. Evers. Elaborating on Thick Analysis: About Thoroughness and Creativity in Qualitative Analysis. *Forum: Qualitative Social Research*, 17(1), November 2015. ISSN 1438-5627. doi: 10.17169/FQS-17.1.2369. URL http://www.qualitative-research.net/index.php/fqs/article/view/2369.

Jeanine C. Evers, Christina Silver, Katja Mruck, and Bart Peeters. Introduction to the KWALON Experiment: Discussions on Qualitative Data Analysis Software by Developers and Users. *Forum Qualitative Social Research*, 12(1), November 2010. ISSN 1438-5627. doi: 10.17169/FQS-12.1.1637. URL http://www.qualitative-research.net/index.php/fqs/article/view/1637.

Nigel G. Fielding and Raymond M. Lee. New Patterns in the Adoption and Use of Qualitative Software. *Field Methods*, 14(2):197–216, May 2002. ISSN 1525-822X, 1552-3969. doi: 10.1177/1525822X02014002005. URL http://journals.sagepub.com/doi/10.1177/1525822X02014002005.

Adam Fouse. *Navigation of Time-Coded Data*. PhD thesis, UC San Diego, California, US, 2013. URL https://escholarship.org/uc/item/6nx3x60t#author. ProQuest ID: Fouse_ucsd_0033D_13078 Merritt ID: ark:/20775/bb7578239n.

Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D. Hollan. ChronoViz: A System for Supporting Navigation of Time-coded Data. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, pages 299–304, Vancouver, BC, Canada, 2011. ACM Press. ISBN 978-1-4503-0268-5. doi: 10.1145/1979742.1979706. URL http://portal.acm.org/citation.cfm?doid=1979742.1979706.

C. Ailie Fraser, Tricia J. Ngoon, Mira Dontcheva, and Scott Klemmer. RePlay: Contextually Presenting Learning Videos Across Software Applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Glasgow, Scotland Uk, 2019. ACM Press. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300527. URL http://dl.acm.org/citation.cfm?doid=3290605.3300527.

Olivier Friard and Marco Gamba. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11):1325–1330, November 2016. ISSN 2041210X. doi: 10.1111/2041-210X.12584. URL http://doi.wiley.com/10.1111/2041-210X.12584.

Susanne Friese. *Qualitative Data Analysis with ATLAS.ti*. 2019. ISBN 978-1-5264-4623-7 978-1-5264-5892-6. OCLC: 1084318135.

A. Girgensohn, J. Boreczky, and L. Wilcox. Keyframe-Based User Interfaces for Digital Video. *Computer*, 34(9): 61–67, September 2001. ISSN 00189162. doi: 10.1109/2.947093. URL http://ieeexplore.ieee.org/document/947093/.

Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. LifeLogging: Personal Big Data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, 2014. ISSN 1554-0669, 1554-0677. doi: 10.1561/1500000033. URL http://www.nowpublishers.com/articles/foundations-and-trends-in-information-retrieval/INR-033.

Joey Hagedorn, Joshua Hailpern, and Karrie G. Karahalios. VCode and VData: Illustrating a new Framework for Supporting the Video Annotation Workflow. In *Proceedings of the working conference on Advanced visual interfaces*, page 317, Napoli, Italy, 2008. ACM Press. ISBN 978-1-60558-141-5. doi: 10.1145/1385569.1385622. URL http://portal.acm.org/citation.cfm?doid=1385569.1385622.

Jon Olav Hauglid and Jon Heggland. Savanta—search, analysis, visualisation and navigation of temporal annotations. *Multimedia Tools and Applications*,

40(2):183–210, November 2008. ISSN 1380-7501, 1573-7721. doi: 10.1007/s11042-008-0204-5. URL http://link.springer.com/10.1007/s11042-008-0204-5.

Peter A. Heeman and James F. Allen. Dialogue Transcription Tools. Technical report, University of Rochester, Rochester, NY, USA, 1995.

William C. Hill, James D. Hollan, Dave Wroblewski, and Tim McCandless. Edit Wear and Read Wear. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 3–9, Monterey, California, United States, 1992. ACM Press. ISBN 978-0-89791-513-7. doi: 10.1145/142750.142751. URL http://portal.acm.org/citation.cfm?doid=142750.142751.

Karen Holtzblatt, Jessamyn Burns Wendell, and Shelley Wood. *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Interactive Technologies. Elsevier/Morgan Kaufmann, San Francisco, 2005. ISBN 978-0-12-354051-5. URL https://doi.org/10.1016/B978-0-12-354051-5.X5000-9. OCLC: ocm56755725, DOI:.

Sungsoo Hwang. Utilizing Qualitative Data Analysis Software: A Review of Atlas.ti. *Social Science Computer Review*, 26(4):519–527, November 2008. ISSN 0894-4393, 1552-8286. doi: 10.1177/0894439307312485. URL http://journals.sagepub.com/doi/10.1177/0894439307312485.

Wolfgang Hürst, Tobias Lauer, Cédric Bürfent, and Georg Götz. Forward and Backward Speech Skimming with the Elastic Audio Slider. In Tom McEwan, Jan Gulliksen, and David Benyon, editors, *People and Computers XIX — The Bigger Picture*, pages 455–471. Springer London, London, 2006. ISBN 978-1-84628-192-1 978-1-84628-249-2. doi: 10.1007/1-84628-249-7_29. URL http://link.springer.com/10.1007/1-84628-249-7_29.

Gail Jefferson. Glossary of transcript symbols with an introduction. In Gene H. Lerner, editor, *Conversation Analysis: Studies from the first generation*, number 125 in Pragmatics & Beyond New Series, pages

13–31. John Benjamins Publishing Company, Amsterdam, 2004. ISBN 978-90-272-5367-5 978-1-58811-538-6 978-90-272-5368-2 978-1-58811-539-3 978-90-272-9528-6. doi: 10.1075/pbns.125.02jef. URL https://benjamins.com/catalog/pbns.125.02jef.

Donald G. Kimber, Lynn D. Wilcox, Francine R. Chen, and Thomas P. Moran. Speaker Segmentation for Browsing Recorded Audio. In *Conference companion on Human factors in computing systems*, pages 212–213, Denver, Colorado, United States, 1995. ACM Press. ISBN 978-0-89791-755-1. doi: 10.1145/223355.223528. URL http://portal.acm.org/citation.cfm?doid=223355.223528.

Zdeněk Konopásek. Making Thinking Visible with Atlas.ti: Computer Assisted Qualitative Analysis as Textual Practices. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol 9:No 2 (2008): Performative Social Science, May 2008. doi: 10.17169/FQS-9.2.420. URL http://www.qualitative-research.net/index.php/fqs/article/view/420.

Tobias Lauer and Wolfgang Hürst. Audio-based methods for navigating and browsing educational multimedia documents. In *Proceedings of the international workshop on Educational multimedia and multimedia education*, page 123, Augsburg, Bavaria, Germany, 2007. ACM Press. ISBN 978-1-59593-783-4. doi: 10.1145/1290144.1290166. URL http://portal.acm.org/citation.cfm?doid=1290144.1290166.

Eric Lee and Jan Borchers. DiMaß: A Technique for Audio Scrubbing and Skimming using Direct Manipulation. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing for Multimedia*, page 107, Santa Barbara, California, USA, 2006. ACM Press. ISBN 978-1-59593-501-4. doi: 10.1145/1178723.1178740. URL http://portal.acm.org/citation.cfm?doid=1178723.1178740.

Rensis Likert. *A Technique for the Measurement of Attitudes*, volume 22 of *Archives of Psychology*. New York, June 1932.

Wendy E. Mackay and Michel Beaudouin-Lafon. DIVA: Exploratory Data Analysis with Multimedia Streams.

In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 416–423, Los Angeles, California, United States, 1998. ACM Press. ISBN 978-0-201-30987-4. doi: 10.1145/274644.274701. URL http://portal.acm.org/citation.cfm?doid=274644.274701.

Katie MacMillan. More Than Just Coding? Evaluating CAQDAS in a Discourse Analysis of News Texts. *Forum Qualitative Social Research*, 6(3), September 2005. ISSN 1438-5627. doi: 10.17169/FQS-6.3.28. URL http://www.qualitative-research.net/index.php/fqs/article/view/28.

Megh Marathe and Kentaro Toyama. 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*, pages 1–12, Montreal QC, Canada, 2018. ACM Press. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3173922. URL http://dl.acm.org/citation.cfm?doid=3173574.3173922.

A. Marsden, A. Mackenzie, A. Lindsay, H. Nock, J. Coleman, and G. Kochanski. Tools for Searching, Annotation and Analysis of Speech, Music, Film and Video A Survey. *Literary and Linguistic Computing*, 22(4):469–488, September 2007. ISSN 0268-1145, 1477-4615. doi: 10.1093/llc/fqm021. URL https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/fqm021.

Jennifer Matheson. The Voice Transcription Technique: Use of Voice Recognition Software to Transcribe Digital Interview Data in Qualitative Research. *The Qualitative Report*, 12(4):547–560, December 2007. ISSN 1052-0147. URL https://nsuworks.nova.edu/tqr/vol12/iss4/1.

Katerina Mavrou, Graeme Douglas, and Ann Lewis. The use of Transana as a video analysis tool in researching computer-based collaborative learning in inclusive classrooms in Cyprus. *International Journal of Research & Method in Education*, 30(2):163–178, July 2007. ISSN 1743-727X, 1743-7288. doi: 10.1080/17437270701383305. URL http://www.tandfonline.com/doi/abs/10.1080/17437270701383305.

Donald McMillan, Moira McGregor, and Barry Brown. From in the wild to in vivo: Video Analysis of Mobile Device Use. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 494–503, Copenhagen, Denmark, 2015. ACM Press. ISBN 978-1-4503-3652-9. doi: 10.1145/2785830.2785883. URL http://dl.acm.org/citation.cfm?doid=2785830.2785883.

Liliana Melgar Estrada and Marijn Koolen. Audiovisual Media Annotation Using Qualitative Data Analysis Software: A Comparative Analysis. *The Qualitative Report*, 23 (13):40–60, March 2018. ISSN 1052-0147. URL https://nsuworks.nova.edu/tqr/vol23/iss13/4.

Jody Miller and Barry Glassner. The 'Inside' and the 'Outside': Finding Realities in Interviews. In *Qualitative Research*, pages 131–148. Sage, 3 edition, 2011. ISBN 978-1-4462-5957-3.

Michael Mills, Jonathan Cohen, and Yin Yin Wong. A Magnifier Tool for Video Data. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 93–98, Monterey, California, United States, 1992. ACM Press. ISBN 978-0-89791-513-7. doi: 10.1145/142750.142764. URL http://portal.acm.org/citation.cfm?doid=142750.142764.

Suphi Umut Naci and Alan Hanjalic. Intelligent Browsing of Concert Videos. In *Proceedings of the 15th international conference on Multimedia*, page 150, Augsburg, Germany, 2007. ACM Press. ISBN 978-1-59593-702-5. doi: 10.1145/1291233.1291264. URL http://portal.acm.org/citation.cfm?doid=1291233.1291264.

Cuong Nguyen, Yuzhen Niu, and Feng Liu. Video Summagator: An Interface for Video Summarization and Navigation. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, page 647, Austin, Texas, USA, 2012. ACM Press. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2207767. URL http://dl.acm.org/citation.cfm?doid=2207676.2207767.

Sandy Q. Qu and John Dumay. The qualitative research interview. *Qualitative Research in Accounting & Management*, 8(3):238–264, January 2011. ISSN 1176-6093. doi: 10.1108/11766091111162070. URL https://doi.org/10.1108/11766091111162070. Publisher: Emerald Group Publishing Limited.

Abhishek Ranjan, Ravin Balakrishnan, and Mark Chignell. Searching in Audio: The Utility of Transcripts, Dichotic Presentation, and Time-compression. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, page 721, Montréal, Québec, Canada, 2006. ACM Press. ISBN 978-1-59593-372-0. doi: 10.1145/1124772.1124879. URL http://portal.acm.org/citation.cfm?doid=1124772.1124879.

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

Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. Sage, Los Angeles, Calif, 2009. ISBN 978-1-84787-548-8 978-1-84787-549-5.

Christina Silver and Ann Lewins. *Using Software in Qualitative Research: A Step-by-Step Guide*. SAGE Publications Ltd, London, 2014. ISBN 978-1-4462-4973-4 978-1-4739-0690-7. doi: 10.4135/9781473906907. URL http://methods.sagepub.com/book/using-software-in-qualitative-research-2e.

Beverly A. Smith and Sharlene Hesse-Biber. Users' Experiences with Qualitative Data Analysis Software: Neither Frankenstein's Monster Nor Muse. *Social Science Computer Review*, 14(4):423–432, December 1996. ISSN 0894-4393, 1552-8286. doi: 10.1177/089443939601400404. URL http://journals.sagepub.com/doi/10.1177/089443939601400404.

Michael A. Smith and Takeo Kanade. *Multimodal Video Characterization and Summarization*. The Kluwer International Series in Video Computing. Kluwer Academic Publishers, New York, USA, 2005. ISBN 978-

1-4020-7426-4. doi: 10.1007/b99812. URL http://link.springer.com/10.1007/b99812.

Anselm Strauss and Juliet Corbin. Grounded Theory Methodology. *Handbook of Qualitative Research*, 17:273–285, 1994.

Atima Tharatipyakul and Hyowon Lee. Towards a Better Video Comparison: Comparison as a Way of Browsing the Video Contents. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, pages 349–353, Melbourne Australia, December 2018. ACM. ISBN 978-1-4503-6188-0. doi: 10.1145/3292147.3292183. URL https://dl.acm.org/doi/10.1145/3292147.3292183.

Jonathan Tummons. Using Software for Qualitative Data Analysis: Research Outside Paradigmatic Boundaries. In Martin Hand and Sam Hillyard, editors, *Studies in Qualitative Methodology*, volume 13, pages 155–177. Emerald Group Publishing Limited, November 2014. ISBN 978-1-78441-051-3 978-1-78441-050-6. doi: 10.1108/S1042-319220140000013010. URL http://www.emeraldinsight.com/doi/10.1108/S1042-319220140000013010.

Aditya Vashistha, Pooja Sethi, and Richard Anderson. Respeak: A Voice-based, Crowd-powered Speech Transcription System. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 1855–1866, Denver, Colorado, USA, 2017. ACM Press. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025640. URL http://dl.acm.org/citation.cfm?doid=3025453.3025640.

Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. Transparency of CHI Research Artifacts: Results of a Self-Reported Survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, Honolulu HI USA, April 2020. ACM. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376448. URL https://dl.acm.org/doi/10.1145/3313831.3376448.

Steve Whittaker, Julia Hirschberg, Brian Amento, Litza Stark, Michiel Bacchiani, Philip Isenhour, Larry Stead,

Gary Zamchick, and Aaron Rosenberg. SCANMail: A Voicemail Interface That Makes Speech Browsable, Readable and Searchable. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 275–282, New York, NY, USA, 2002. ACM. ISBN 978-1-58113-453-7. doi: 10.1145/503376.503426. URL http://doi.acm.org/10.1145/503376.503426. event-place: Minneapolis, Minnesota, USA.

David K. Woods and Paul G. Dempster. Tales From the Bleeding Edge: The Qualitative Analysis of Complex Video Data Using Transana. *Forum Qualitative Social Research*, 12(1), January 2011. ISSN 1438-5627. doi: 10.17169/FQS-12.1.1516. URL http://www.qualitative-research.net/index.php/fqs/article/view/1516.

Megan Woods, Trena Paulus, David P. Atkins, and Rob Macklin. 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):597–617, October 2016. ISSN 0894-4393, 1552-8286. doi: 10.1177/0894439315596311. URL http://journals.sagepub.com/doi/10.1177/0894439315596311.

H. J. Zhang, C. Y. Low, S. W. Smoliar, and J. H. Wu. Video Parsing, Retrieval and Browsing: An Integrated and Content-Based Solution. In *Proceedings of the third ACM international conference on Multimedia*, pages 15–24, San Francisco, California, United States, 1995. ACM Press. ISBN 978-0-89791-751-3. doi: 10.1145/217279.215068. URL http://portal.acm.org/citation.cfm?doid=217279.215068.

Index

Typeset September 8, 2020