The ‘Conditional Voice Recorder’: Data practices in the co-operative advancement and implementation of data-collection technology

Tim Hector / Franziska Niersberger-Gueye /
Franziska Petri / Christine Hrncal

WORKING PAPER SERIES | NO. 23 | SEPTEMBER 2022
Collaborative Research Center 1187 – ‘Media of Cooperation’
Sonderforschungsbereich 1187 – ‘Medien der Kooperation’

Working Paper Series
Collaborative Research Center 1187 – 'Media of Cooperation'

Print-ISSN 2567–2509
Online-ISSN 2567–2517
DOI https://doi.org/10.25819/ubsi/10168
dspace.ub.uni-siegen.de/handle/ubsi/2257
URN urn:nbn:de:hbz:467-22576

This Working Paper Series is edited by the Collaborative Research Center Media of Cooperation and serves as a platform to circulate work in progress or preprints in order to encourage the exchange of ideas. Please contact the authors if you have any questions or comments. Copyright remains with the authors.

The Working Papers are accessible online at:
https://www.mediacoop.uni-siegen.de/de/publikationen/working-papers-media-of-cooperation/

Print copies can be ordered by sending an email to:
workingpaperseries@sfb1187.uni-siegen.de

Publication is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 262513311 – SFB 1187.

Cover image: Sina van Oostrum

Universität Siegen
SFB 1187 Medien der Kooperation
Herrengarten 3
57072 Siegen, Germany
https://www.mediacoop.uni-siegen.de/
workingpaperseries@sfb1187.uni-siegen.de

Abstract Stationary voice-controlled systems are installed in an increasing number of households. The devices are operated primarily via voice-user interfaces, which evaluate the spoken commands cloud-based, and are aligned to the principles of interpersonal interaction. This raises questions about the integration of the devices into everyday practices carried out in the household: How is use of Smart Speakers negotiated situationally, embedded in interpersonal interactions, and (how) are aspects of data privacy, data processing and potential exploitation reflected by the users? The project "Un/desired Observation in Interaction: Intelligent Personal Assistants" in the CRC "Media of Cooperation" approaches these questions empirically.
However, such an investigation of smart speakers faces the methodological challenge that this requires voice data documenting not only the use of the smart speaker itself, but also the contexts of the use that go beyond mere "voice commands". Therefore, a "Conditional Voice Recorder" (CVR), a technology developed in Nottingham by Porcheron and colleagues (2018), was brought to bear to create audio recordings of usage contexts. These include not only the voice command itself, but also a few minutes before and after the smart speaker is addressed. However, the original device required further technical development to be compatible not only with Amazon's smart speaker model, but also products from other providers (Google and Apple).
The paper reflects on this advancement and the implementation of the CVR – i.e. our own research practices – as data practices. On the one hand, it makes visible which (otherwise opaque) data were collected and processed during the advancement, how the usage of the CVR itself is inscribed in the data recorded with it, and which data practices were carried out in the evaluation. On the other hand, it documents the advancement and application of the CVR to enable other studies with it (or similar technologies).

Keywords: Smart Speaker, Voice Assistant, Intelligent Personal Assistant, Voice-user interface, Human-Machine-Interaction, Data Practices, Research Practices, Hot-word detection, Data collection

1. Introduction

The Project *(Un)desired observation in interaction: "Intelligent Personal Assistants" (IPA)* of the Collaborative Research Center (CRC) 1187 "Media of Cooperation" at the University of Siegen aims to investigate empirically the spread of smart speakers in private homes using qualitative methods taking a perspective of media sociology and applied linguistics. In this paper, we reflect upon a technology which allowed us the data-driven linguistic analysis of smart speakers in private homes, namely the Conditional Voice Recorder (CVR, see Figure 1).

Figure 1: The setup for data collection: Conditional Voice Recorder (right) next to a Smart Speaker (left).

The use of stationary voice-controlled assistance systems such as Amazon's "Echo Dot" with "Alexa" or Apple's "HomePod" with "Siri" is a phenomenon that is not easily accessible to empirical investigation. While various studies from numerous disciplines have examined the use of such devices based on interviews with users (see e.g. Luger/Sellen 2016; Pins et al. 2020) or log files of such assistance systems (see e.g. Ammari et al. 2019; Bentley et al. 2018), experimental settings would not quite be suited for a conversation-analytically motivated examination of the integration of smart speakers into everyday domestic conversation situations (Bergmann 1985, 317–318; also see Hector in print for a discussion of praxeological approaches for the analysis of smart speakers) and log-file data provide insufficient contextual information about the conversation and the situation in question (Habscheid et al. 2021).

The CVR was essential for generating recordings of smart speakers in use in a domestic environment. The device developed and first used by Porcheron et al. (2018)[1] in the Mixed Reality Lab at the University of Nottingham allows audio signals to be saved depending on whether an activation word has been said. To this end, in standard mode the CVR uses a microphone to record sound in the environment where it is positioned. At the same time, the device deletes recorded audio signals after a period of 3 minutes so that its memory only ever stores 3 minutes of audio recordings. If the device recognises a pre-set activation word using hot word detection – originally Alexa – the recording continues for another 3 minutes before being saved which then results in 6-minute audio recordings. The recording continues for an additional 3 minutes if the activation word is recognised a second time within the 3 minutes after the activation word has been recognised for the first time.

The CVR hardware consists of a Raspberry Pi 3 combined with three LED lamps and a USB flash drive as well as a USB conference microphone. In its original configuration, the device's software consists of a Python script as well as a hot word detection library provided by Snowboy[2]. This version recognises "Alexa" as the activation word. While we were able to replicate the original version of the CVR with the valuable advice provided by Martin Porcheron and Stuart Reeves from Nottingham and extensive assistance offered by the FabLab of the University of Siegen[3], when using the device in the intended context, we faced the issue that it was designed to be activated by the single wake word "Alexa". The device did not support wake words of other smart speaker operators such as Google or Apple ("Hey/OK Google" or "Hey Siri").

In order to extend the functionality of the device to include the three additional wake words, it was not only necessary to involve an external service provider[4] for programming and configuration but also to collect and process additional data sets. This additional data collection primarily consisted of voice samples intended to be used for the development of speech recognition.

The present working paper aims at documenting the process of developing the CVR data collection technology while taking a reflexive perspective on our research practices with the intention of contributing to greater sustainability and opening of the research practice (Strauch/Hess 2019; Mosconi et al. 2019). We thus also consider the question raised by, among others, Ruppert et al. (2013, 25) of the "consequences of digital devices for social scientific ways of knowing" and accordingly address the following questions: What is the relationship between the phenomena of interest and the digital methods with which they are being examined as well as with the data generated and processed to this end? In what way are the inscriptions (Ruppert et al. 2013, 31 with reference to Latour 1990) of data practices reflected in the data themselves and how does this shape their analysis and, in turn, the genesis of knowledge? To this end, we rely on the concept of data practices which we will introduce as a theoretical framework (Section 3) after a brief description of the CVR in terms of technology and design (Section 2). We will then focus our attention on the additional collection of data used for the (originally planned) training of the speech recognition algorithm as well as on steps taken by the team of researchers to process and prepare that data (Section 4). Furthermore, we will turn to the use of the CVR during the main data collection and how the device is reflected in the recorded data (Section 5) as well as to the process of data processing, preparation and analysis. Thus, we reflect upon the practical research implications of our own data practices for the examination of smart speakers.

2. Conditional Voice Recorder: Instrument and research design

The CVR as replicated for the purpose of our project is based on the works of Martin Porcheron and Stuart Reeves (see above). It consists of several components:

1 GitHub documentation is available at https://github.com/mixedrealitylab/conditional-voice-recorder.

2 Since 2021, this library is no longer available (see below).

3 We are particularly grateful to Fabian Vitt.

4 In the course of the development, Kernel Concepts (Siegen) was commissioned to cooperate in the advancement of the CVR's software and configuration. We are particularly grateful to Ole Reinhardt and Simon Budig for their cooperation. For a description of their work, see: https://www.kernelconcepts.de/case-study-conditional-voice-recorder/.

– a Raspberry Pi 3 as processing unit which has to be connected to the power supply via a mini-USB port and built into a suitable Raspberry housing,
– an SD card on which the operating system and other required applications are installed,
– a USB flash drive connected to the device to save audio files in WAV format,
– three LEDs (green, yellow and red) indicating the device status (see below)
– as well as a conference microphone connected via a USB cable.

The processing unit and the microphone are placed close to the smart speakers in the homes selected for our study (see Figure 1, the smart speaker is located to the left of the CVR). The decisions which have led to this set-up were primarily of a pragmatic nature. Given its flexibility in terms of programming for our specific purpose, which also allowed us to adhere to data protection requirements, as well as its cost efficiency, the Raspberry Pi 3 and the installed Python application favoured by Porcheron and colleagues also suited the kind of data collection we had planned. Moreover, the major components (Raspberry Pi 3 in its housing and the conference microphone) are similar to a smart speaker in terms of dimensions. They can thus be placed unobtrusively in the vicinity of the smart speaker (also see Merkle/Hector in prep.). Placing the device not too far from the smart speaker is necessary for the recognition of the activation word to work properly. The multi-colour LEDs indicate whether any errors in the CVR configuration have been detected (e.g. missing USB drive for saving data, microphone not recognised). They also indicate whether the device is recording and whether it is saving the current recording. In this regard, the LEDs meet concerns expressed in the privacy discourse which played (and still plays) a major role both in examining smart speakers and designing the CVR. The CVR imitates the smart speaker, it "spies" on the "spy" (the smart speaker) and its use with its own means of speech recognition. (cf. Lau et al. 2018): It records data from the private homes of the study participants and is—in this regard—quite invasive. Several methods to address privacy issues affect the design of the CVR. These include the LEDs mentioned and the complete avoidance of an internet connection as well as the purely "physical" transmission of the audio data via USB flash drive[5]. In terms of function and design, we notably pondered the alternative of using a data collection method based on light sensing technologies as did Schönherr et al. (2020) in their study of "accidental triggers". One advantage of this method lies in its high reliability in terms of activation: Since this method involves starting the recording whenever the smart speaker's integrated lights signal that the device is in listening mode, a complete data collection can be assumed. By contrast, with the speech recognition-based version of the CVR as used for this study, it is possible that mentions of the activation word may not be recognised depending on background noise, volume, voice timbre and other acoustic factors. This also results from the fact that the CVR technology we replicated imitates speech recognition similar to that used in smart speakers by Google or Amazon, however, it proved considerably less performant than commercial products both in terms of hardware and software. Nonetheless, developing a device processing acoustic signals to collect data provided major advantages: First, the CVR could be more flexibly placed in the private homes. Given a planned data collection period of three to five weeks, this was quite necessary since it meant that the device had a secure location during this period at which its presence was not overly obtrusive and the smart speaker could continue to be used normally. By contrast, a light sensing technology requires that a sensor be placed immediately at the surface of the smart speaker. This also requires a more comprehensive and expert preparation of the device which was not needed for the CVR that only needed to be connected to the power supply. This meant that data collection—not least in light of social distancing imposed by the Covid-19 pandemic—could be performed remotely and we could post the CVR to the study participants who then set up the device themselves and posted it back once data collection was completed. This, too, occasionally led to difficulties in the maintenance of the devices when errors occurred and, in turn, to delays. However, personal encounters were thus avoided where possible.

This goes to show that CVR design and study design, while not deterministically related as such, do mutually influence one another and consequently affect the type of data generated and stored. In this regard, design decisions and the decision for this particular device and its corresponding revised versions may already be considered as a first component of data practices. Before we take a closer look at further data practices of developing the technology as well as of collecting and analysing the data, this concept requires theoretical contextualisation.

3. Data practices as theoretical framework

Data cannot be "conceived without the practices of accomplishment, settlement, evaluation, testing or valorisation that precede or follow them [...] They are conceptualised as data practices and contribute to the incremental accomplishment of data" (CRC Media of Cooperation 2019). As such, data practices constitute a

5 Conceptually, it would be worthwile to discuss the implications of data protection and privacy for design decisions further also against the background of a comparison between a smart speaker and a CVR, especially from the point of view of Agre's (1994) distinction between "surveillance" and "capture". Such a discussion is not within the scope of this paper but might inform further reflections.

response to the "practice turn" (Schatzki et al. 2001) and thus to a number of praxeological foundations as well as to the ethnomethodological research agenda of Harold Garfinkel (1967). At the same time, we understand data practices as a concept based on an understanding of practice as co-operation as proposed by Charles Goodwin in his numerous works and his monograph published in 2018 in particular: co-operation as "the mutual accomplishment of common goals, means, and processes" (Schüttpelz/Meyer 2017, 158). Means are either created incrementally or understood historically by those involved and thus facilitate the spontaneous emergence of goals—shared or not by those involved, but accomplished co-operatively in any case (Schüttpelz 2020). This view of co-operation cannot only be applied to visible interaction situations between two involved humans, but is very flexible both in terms of scaling and of objects (see Thielmann 2018 for instance). In this respect, it is also suitable for reflecting on data practices performed in the "background". We will return to this.

Looking at digitally connected media and infrastructures in their practical make-up and continuous creation through the lens of practices (as Gießmann 2018 claims for instance) has proved a very fruitful approach to (re-)define media as created by infrastructures and public audiences (CRC Media of Cooperation 2019). The flexibility of the concept of co-operation also allows to apply this understanding of co-operation and the idea of "reciprocal accomplishment" to data and to dissociate oneself from an absolute data concept accordingly (also see Gießmann/Burkhardt 2014; Dalton/Thatcher 2014; Burkhardt et al. 2022). Data are of a relational nature and must be seen in their respective context; the notion of "raw data", one which also has been mentioned in the context of data collection and processing in the present project more than once, is repudiated (Bowker/Star 2000; Bowker 2008) as is also demonstrated in the contributions compiled by Gitelmann (2013): Data do not simply exist, they are not a natural resource and just require interpretation, they are data of something for something as Gießmann/Burkhardt (2014, 3) emphasise.

Practice theory-oriented studies are confronted with the question pointedly raised by Hind et al. (2021, 2): "If the world is awash with practices, what then?". Likewise, a praxeological view of "data" must address the question of which added value a practice-based description is able to deliver. To this effect, we will also have to shed light on the repertoire of methods which recently has become more differentiated and that is available for such analyses of data. Several avenues have opened up. Thus, the avenue of data journeys (Bates et al. 2016) follows a rather socio-material approach and maps out "the movement of data through space and time". Researchers in such a methodological setting leave the outsider's perspective and, on their path through infrastructured environments, are "embedded alongside data" from production to variegated situated further uses. In particular, they refer to the configurations of materiality of the data in the digital space and build on such approaches that took into account the socio-spatial distribution of data (see for instance Sands et al. 2012; McNally et al. 2012; Beer/Burrows 2013).

Building on this and "critical data studies" (Dalton/Thatcher 2014; Dalton et al. 2016), Tkacz et al. (2021) conceptualised "data diaries", also aiming to understand data practices empirically. Using the ethnographic concept of the "diary" they align this more closely with single case descriptions than do Bates et al. (2016). The diaries endeavour to "follow the data" over a much extended period and "to understand what data do, how they move, are drawn upon or ignored, and generally co-constitute a given spatial 'situation'" (Tkacz et al. 2021, 2). This operates with a strong concept of situatedness which should be prioritised for presenting the diaries already (Tkacz et al. 2021, 4). Building on the concept of "situated knowledges" of Haraway (1988), Rettberg (2020) provides another methodological approach operating with the concept of "situated data" that takes seriously the fact that situatedness and point of view are inscribed into any representation of data (also see Lynch/Woolgar 1990; Coopmans et al. 2014). Rettberg (2020) thus carves out the power relationships between users and the platform operator of a fitness tracking app resulting from data situated in different ways.

Further applications of these methodological branches of practice-based description and analysis of data, which are not always clear-cut and partly overlap, are recent but variegated. In "Data Practices. Making Up a European People", Ruppert/Scheel (2021) propose a classification of data practices and use it to illustrate the relevance of data practices and a notion of situated data by means of statistical procedures for "making up" the European population. Likewise, Lämmerhirt (2021) examines data as practice in the context of data donations in the health care sector. Burkhardt et al. (2022) compile numerous additional areas of application and constitutive conditions for a praxeology of data.

To date, research practice and thus a reflexive use of the concept has received less attention as an application of a praxeological consideration of data. Using the example of advancing the described data collection technology with which data is processed that remain invisible at the surface, this paper carves out in detail, and thus documents, the various steps taken to inform future projects of this kind about the development and inspire planning procedures. At the same time, it emphasises to what extent audio data had to be processed in practice and divided into different states, situated and re-situated to bring about a data

collection technology able to compare to stationary voice-controlled assistance systems provided by the large players Amazon, Apple and Google, at least in terms of hot word recognition. Despite this ambition, the dimensioning of the processes in our research project cannot be described as even a fraction of data practices performed by the large players on a daily basis, both in terms of the sheer mass of data collected and, much more so, the de- and re-situation of data. The development process in our project thus adumbrates the basis on which data practice operations are performed which lead to the provision of products such as "Alexa", "Siri" or "Google Home". If, to this end, one intends to uncover the practices of accomplishment, testing or valorisation that precede or follow (cf. CRC 1187 Media of Cooperation 2019) and to document them in the interest of transparency of practical research—to present them as an example of the situationally bound states and transformation6 processes of data—, we believe, the best way is to follow the data as did others before us. In doing so, we ask how the "being present and available" of the data at different points in time during the process is negotiated between those involved and forms it, which intermediate steps had to be taken for transformation, and what kind of representations can be assumed by audio data in particular during that process. In this regard, the procedure is substantially inspired by previous approaches to data practical analysis such as "Data Journeys" and "Data Diaries" (see above). At the same time, it attempts to adapt it to the research reflective context such that it will simultaneously result in a conclusive documentation of advancing the CVR.

4. Data practices in hot-word-detection-development

Let us first turn to those data and data practices that remain opaque to the users of the data collection technology – namely the researchers – in the actual use of the CVR, that is those audio data which were necessary to further develop the CVR but which were subsequently not resorted to. These data were required to train the speech recognition models with the help of Tensorflow[7]. The company entrusted with developing the device needed 56 audio files of predefined activation words spoken by 100 people each, that is a total of 5,600 audio recordings. First, each of the 100 people had to record four relevant hot words ten times in a row, each time varying the recording slightly in terms of voice modulation or volume. Second, eight different 'false' trigger words had to be recorded twice each (e.g. "Torjubel" instead of "OK Google" or "extra" instead of "Alexa"). This approach intended to enable the trained AI model to distinguish hot words from similar-sounding words. The selection of people was supposed to be as diverse as possible in terms of characteristics such as age, voice pitch or native language, in order to prevent the trained AI from solely reacting to the voice characteristics of certain people. The company entrusted with the development needed every hot word and every negative example of each individual as a separate file adhering to certain technical standards in terms of quality and format. Ideally, each recording should be exactly one second long but no longer than 1.5 seconds. Our partner company only needed the recordings (without any additional information such as name, age or gender).

Before we were able to collect the data, a number of data protection issues had to be addressed. First, we had to clarify whether mere voice data constitute personal data when no additional information is collected. Consulting with the data protection office of the University of Siegen8 , we determined that—regardless of the type of voice recording—no anonymisation in the stricter sense of the word could be ensured as essential elements of a voice will be identifiable. This is why participants had to receive comprehensive information as to the purpose and subsequent processing of the recordings and this information had to be documented. In addition, it was necessary to sign a data processing contract with the contracting partner in accordance with Art. 28 of the General Data Protection Regulation (GDPR)[9] and to maintain what is called records of processing activities. When phrasing the declaration of consents for the participants we needed to make sure the document contained both an option to consent and to revoke and that this latter option could be matched, also in a blinded fashion, to the potentially revocating person. The declaration of consent had to be designed such that people were informed about the fact that their personal data was collected in

6 See Sterne (2012, 195) for the notion of the transformation of audio data; see Jänicke et al. (2017, 230–232) for the relevance of data transformations in the digital humanities; and see Latour's notion of the "circulating reference" (see in particular Latour 2000[1993], 65–72) for the general relevance of transformation processes in scientific work.

7 Tensorflow is a platform for the development of AI applications such as hot word recognition, see https://www.tensorflow.org/. The contractor company originally planned to use these data to develop the speech recognition for the activation words needed for our purposes. Since problems arose with the trained Tensorflow-model (detection was not precise enough and a very high number of false positives were recorded), the developers later on changed their strategy and used a pre-trained wake word engine from Porcupine. For more details, see https://www.kernelconcepts.de/case-study-conditional-voice-recorder/.

8 We are particularly grateful to Christina Schumann for her expert advice.

9 See https://www.bmj.de/DE/Themen/FokusThemen/DS-GVO/DSVGO_node.html (last accessed 30 September 2022).

accordance with Art. 6 (1) (a) of the GDPR. They were informed about the context of our research project as well as the purpose of collecting the voice data. The participants were also informed that the data would not be published, that the AI machine was trained offline and only a small number of employees had access to the data. Moreover, the fundamental rights of the data subjects were presented in accordance with the GDPR.

The voice samples were labelled with an identifier to enable finding the recording of a particular person and delete it should the need arise. To this effect, a number was assigned to each recording of the same person and the corresponding declaration of consent. We thus collected more data than was strictly necessary for the development of the device to meet the data protection requirements. The contracting company would not have needed the additional information to train the AI machine and received anonymised recordings. However, legal requirements mandate the collection of additional data (name, place, signature) by means of the declaration of consent. These had to be documented in the records of processing activities we prepared (see Art. 30, GDPR).

We were able to contact a sufficient number of participants and convince them to take part in the project via our private and professional networks. To keep effort for the participants to a minimum, we asked them to record the hot words and the "false" trigger words in a single voice recording and send this recording back to us. Initially, these recordings had quite different file names and were stored on the work computers of various team members. The team member in charge then collected, sorted and renamed the files to facilitate easy identification. At this point, the recordings had left the devices of the study participants and consequently witnessed a process of compilation and making them identifiable. The recordings were then adapted to the technical requirements of the external company and divided into separate voice samples with the help of audio editing software (see Figure 2).

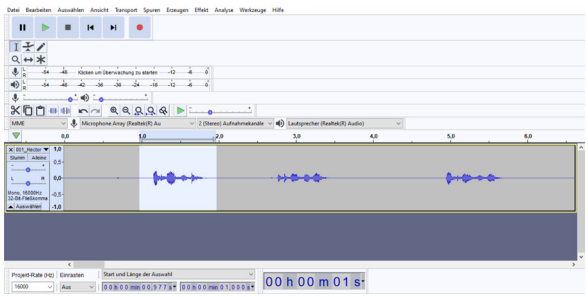

Figure 2: Editing of the voice samples

This clearly illustrates a process which had already been completed in the previous stages: the different visual representations of audio data; at the interface of the storage location on the one hand and in the oscillogramme of the audio editing software on the other. The editing process also changed the visual representation of each file and this procedure in turn transformed the existing data. After editing them, the individual voice samples were labelled with the respective identifier and hot word or the distinctive example to allow for clear identification without looking at the oscillogramme or listening to the audio file.

The data were then handed over to the company physically by means of an encrypted USB drive where they were then transformed, yet again, into other states and representations (the exact details of which largely remain opaque for the team of researchers). The developers used the data to train the AI machine – which in the end led to the use of another model (Porcupine) with pre-trained data (see above). What already becomes clear at this point is, that "data" mean quite different things in a process such as ours: While the data initially existed on the devices of the study participants, they were subsequently transmitted to the computers of the University of Siegen where they were identified and further processed by the team members. Likewise, the declarations of consent were transmitted and assigned. In this context, the practices of transformation and identification are of particular significance. The former refers to the transformation into various data formats and sizes as well as that into various forms of presentation which contribute significantly to the configuration of media and their production, distribution and reception (Volmar 2017, 9; also see Sterne 2012). The latter primarily refers to assigning the declarations of consent to the corresponding audio recordings as well as assigning the files to their corresponding content (see Gießmann 2020 for details on the media practice of identification as contrasted with registration and classification).

Up to this point, we spoke of data whose relevance only indirectly unfolds as part of the data collection technology. At no point in time (with the exception of this paper) have they been the subject of publications; they are not analysed in the scientific sense of the word and were collected with the intention to collect other data (with the help of appropriate AI training) – and, ironically, they are not even part of the finally used AI-models as the Porcupine-model provided a better data set. However, they tell a story of how data for a data collection could be collected, processed, identified and transformed and made their contribution to another data collection (our main study). The next chapter focuses on these data and the use of the CVR in the participating households.

5. Reflection of data practices in the collected data

For the people involved in the recordings and in whose homes CVR data was collected, data practices in using the hot word technology play a double role: first, in that they generate data themselves by using the smart speaker and second, in that these data thus generated are recorded by the CVR device as new or additional data. At times, the participants broach, and reflect on, such "observation" of their data practices when using the smart speaker made by the CVR recordings.
In the following example, Damaris (DL) addresses the recording of her routine use of the smart speaker by making sure with her partner Jan-Ole (JS) that it is okay to deactivate the recording device (l. 776/778).

Transcript (1): "Shall we switch that off now?"
776 DL: wollen wir das jetz (.) äh (mal) AUSmachen das (äh) diktiergerät?
do we want to switch this off now the uhm dictaphone
777 p: (4.9)
778 DL: also wir haben das jetzt auch LANge genug (.) oder?
so we have this long enough don't we
779 p: (0.8)
780 JS: ja.
yes
781 p: (4.3)
782 JS: (aber die) aLEXa dann auch.
but also the Alexa then
783 ((lacht))
laughing
784 DL:hm?
huh
785 p: (0.5)
786 JS: und die aLEXa auch aus?
and the Alexa off too
787 p: (0.4)
788 DL: ja.
yes

Interestingly, Jan-Ole suggests to switch off "but also the Alexa then" (l. 782) together with the recording device. It remains unclear in this example whether Jan-Ole intends to fully deactivate the smart speaker by disconnecting it from the power supply or to merely switch off the microphone or possibly just end the ongoing activity of the smart speaker.

In Transcript 2, Damaris informs Jan-Ole that she has activated the recording device and thus also calls to attention that anything being said will now be recorded (l. 482), which Jan-Ole, however, does not frame as cause for concern ("so", l. 486) as they "did not mention any names" (l. 490/493). While Damaris focusses her attention on the recording and thus the auditive observation of the conversation, Jan-Ole's remark frames the mentioning of names and thus potential identifiability of other people as possibly delicate.

Transcript (2): "I started the recording by the way"
482 DL: ich hab übrigens das dikTIERgerät laufen.
by the way I have the dictaphone running
483 p: (0.5)
484 JS: mhm,
mh
485 p: (1.2)
486 JS: und?
so
487 p: (1.2)
488 DL: wollt ich nur SAgen;
I just wanted to say
489 p: (2.0)
490 JS: HAM ja keine-
we did not
491 p: (0.9)
492 DL: [(ich wo-)]
I wa
493 JS: [NAmen genann]t.
mention any names
494 p: (0.3)
495 DL: ich wollt_s dir TROTZdem nur sagen;
I just wanted to tell you anyway
496 p: (0.3)
497 DL: also das war EINfach nur-
so that was just
498 JS: mhm,
mh
499 DL: als informaTION.
a piece of information
500 p: (4.5)

Subsequently, Damaris marks the observation of the conversation for research purposes and the related data practices as relevant (to herself) several times (l. 488/495/497 and 499), which Jan-Ole, however, does not approve (the relevance of the information is not confirmed interactively).
In the following example (Transcript 3) recorded in a household in which Sam (SR) and Andrea (AS) live together, Sam explicitly addresses the observation situation (l. 002) and the related limitations (l. 011–013). From the recording it is initially unclear whether Sam refers to the smart speaker or to the CVR newly placed in their home when he mentions the "permanent listener".

Transcript (3): "We now have a permanent listener"
002 SR: so andrea wir haben jetzt n ständigen ZUhörer.
so Andrea we have a permanent listener from now on
003 AS: WAS?
what
004 SR: wir haben jetzt ein ständigen ZUhörer;
we have a permanent listener from now on
005 AS: ham wir doch schon die ganze ZEI:T;
we have already had that all the time
006 SR: JA stimmt.

yes true
007 <<p> (nich die GANze zeit.)>
not all the time
008 heheheHE;
009 k: ((Hintergrundgeräusche, ca. 4.6 Sek.))
background noise, approx 4.6 sec
010 AS: woll_n wir jetzt FIcken,
do we want to fuck now
011 SR: !SCH!;
shush
012 RUhe.
silence.
013 So was darfst du jetzt nicht mehr SAgen.
you can't say things like that anymore
014 p: (1.7)
015 AS: das find ich aber KAcke.
that sucks
016 SR: WAS jetzt;
what now
017 (über nix ge-)
talk about nothing
018 nix mehr gesprochen werden was
irgendwie privater
naTUR ist-
talk about nothing that is somehow of a
private nature
019 wir sind jetzt nur noch REIN;
from now on we're only
020 EHM.
021 AS: oKAY?
022 SR: beruflich verBUNden;
connected professionally
023 p: (2.4)
024 SR: heheheHE;
025 p: (1.0)

Andrea appears to regard negatively the limitations related to the observation situation mentioned by Sam, judging from her subjectively framed remark (l. 015). Consequently, Sam restricts future topics of conversation even further (l. 018). It can already be suspected from Andrea's remark in line 005 that the permanent listener mentioned by Sam in line 002 does not refer to the smart speaker but to the CVR placed in their home for research purposes. Sam's remark in the further course of the recording as part of a test scenario corroborates this.

Transcript (4): "You now have a listener"
064 SR: (jetzt bin ich geSPANNT);
now I am curious
065: aLEXa?
066 p: (1.3)
067 SR: du hast jetzt einen ZUhörer.
you now have a listener
068 p: (2.4)
069 AL: das WEIß ich leider nicht.
I'm afraid I don't know.
070 p: (0.7)
071 SR: oh blöde KUH-
oh stupid cow
072 du weißt doch GAR nichts.
you don't know anything
073 p: (4.2)

It is not least this example that illustrates clearly that the recorded household members frame the research data (collection) practices, given their selective recordings, as much more problematic than the continuous "scanning" of the environment typical of the smart speakers.

6. Data practices in data analysis

During its first use, the CVR recorded 389 files in WAV format typically 6 minutes in length in the first phase. These recordings were initially sorted only by date and time and compiled in directories covering periods of two to four days. In order to manage this quantity of recordings, they first had to be listened to separately; again, the practice of identifying became relevant: the co-operative creation of an agreement between stored file and its content (also see Gießmann 2020, pp. 3–4). At the same time, data processing includes a transformation from a visual to an acoustic presentation.

Next, we inventoried the recordings in an Excel spreadsheet. We listed the file name generated by the CVR (consisting of the word recording followed by a ten-digit number), the new project-specific file name as well as time and date of the recording[10]. The inventory table as the most important tool for the research practice also contained information on whether or not the activation word ("Alexa" in this case) was mentioned at all, on the number of speakers at the time of calling the activation word and a transcript of the voice command as well as the remarks that were heard immediately before and after. For those table entries of recordings in which the CVR switched on by mistake, remarks were transcribed before and after the time when the voice command would typically have occurred to identify the false trigger. Information that would potentially have allowed to infer the identity of study participants was anonymised. We used a basis transcript to transcribe voice data following the GAT2 transcription conventions (Selting et al. 2011)[11], thus transcribing both the participants' remarks and

10 For data protection reasons, the CVR does not provide an internet connection and thus no system time. The time is a preset time which is not saved and is set to zero for every restart of the device. We were able to determine the actual times with the help of information provided by the study participants.
11 The transcription conventions are provided in the appendix.

background noises such as acoustic signals captured in the IPA. A further table column contained a description of the content. Thus, we noted the current conversation topic at the time of the recording, whether the TV was switched on or someone was on the telephone.

A large number of recordings were false positives which we had to delete then. Their share was 97%. However, we were able to determine that the amount of recorded data was plausible and the high number of false positives simply exceeded the actual use by way of comparing them to the log files, which document the use of the smart speaker, provided by the smart phone app accompanying the smart speaker (see Habscheid et al. 2021 for this). This resulted from a less accurate algorithm for hot word detection compared to that of commercial products. For data protection reasons these files were deleted from the devices of project employees, but they had to be listened to to identify them first. The entrusted company then, after a first round of collection, developed an additional feature to simplify handling the sheer number of false positives: The second in which the hot word detection algorithm recognised the (supposed) activation word was copied from within the recording and pasted at the start of the recording (separated from the rest of the recording by an acoustic signal). This made it possible for the team members to identify right at the beginning of the recording whether the file needed to be deleted or constituted a "hit". This feature demonstrates the importance of different representations particularly clearly: While the files saved on the USB flash drive look "identical" and, from the outside, cannot be distinguished, the latter is possible by "listening" to them. However, as the audio file type naturally progresses in time and only allows visual navigation inside the data set with the help of a visual representation such as an oscillogramme, the search for the activation word had to be simplified with the help of an additional feature. This feature reacts to necessary data practices of identification to be performed by way of a transformation within the data. Those recordings in which the smart speaker was activated and which were classified as suitable for analysis were subsequently presented in the described Excel sheet. They were then anonymised, pseudonymised and transcribed in full length using the GAT2 conventions with the help of EXMARaLDA.[12] This marked yet another "transformation". The spoken word is transcribed to make it more enduring and accessible to other forms of analysis. Transcription, it is worth noting, is a processual interpretation and already part of the analysis (Ochs 1979; Selting 2001) and cannot ignore the objectives and questions for which something is being transcribed. The act of transcription can be regarded as a data practice of "interpreting transformation" which facilitates the identification and analysis but at the same time necessarily pre-processes an interpretation of the material which accompanies the subsequent analyses and must be reflected upon constantly (also see Dittmar 2004; Duranti 2006; Davidson 2009). In order to be able to work with the data – in a first step in common data sessions for instance –, they have been transformed accordingly once again. On the basis of the inventory table, we selected audio/video clippings, which were in turn cut to form separate video or audio files. The corresponding transcripts, too, were shortened and compiled in accordance with the newly created clippings. The discursive deliberation of the data material as part of data sessions finally resulted in an interactive genesis of knowledge—again on the basis of data practices previously performed—which found its way into publications and presentations.

7. Conclusion

In reflecting upon data practices involved in the development and use of the CVR as a tool for data collection, three interlaced practices emerged: identifying, representing and transforming. Practices of representation are contingent on identification, that is, different forms of presentation depend on the purpose of identification. At the same time, forms of representation both enable and limit the possibilities of identification. Practices of transformation enter into this relationship to facilitate new forms of representation which in turn allow different identifications. This does not constitute a counter proposal to the triangulation of identifying, classifying and registering very convincingly proposed by Gießmann (2020) using the example of credit cards. Instead, our example demonstrates that different media and data practices are performed in different contexts and both are tightly interwoven.

These data practices (in our case identifying, representing and transforming) are inscribed in the data and inevitably shape the scholarly analysis to varying degrees at and below the surface. Practices of identifying inscribe which parts of the data material are actually accessible to an analysis. Representations, i.e. different forms of presentation (such as the Excel spreadsheet, the oscillogramme of an audio file, the transcript in various transcription editors) define how they can be analysed and can thus not be regarded as neutral entities in the process of analysis but as shaping that process. Finally, transformative processes may enable analysis in the first place but influence them at the same time, as was the case for the previously mentioned transcription as a transformation from speech to writing. Consequently, sense making is influenced and shaped to a large

12 See https://exmaralda.org/de/.

extent by the data practices which are performed in collecting, processing, analysing and sampling data.
The development of the CVR unveils yet another level: In developing the technology required for the data collection, a second process of data collection and data use is completed which remains opaque at first: For training the AI based speech recognition software, additional participants were involved who provided a voice sample for the analysis. These data are only analysed indirectly to gain knowledge about the use of smart speakers: They only serve the purpose of enabling data collection to generate such knowledge in a second step. However, at the same time, they significantly influence the database: it remains invisible, which data were not recorded at all, because of the AI model, trained with the audio-samples, simply not recognizing the activation word. Other reasons for failed collection include a defect in the conference microphone used or processing issues of the Raspberry Pi.
Furthermore, the collection situation is evidently inscribed in the data: users, sometimes explicitly, reflect upon the situation of double observation – by the smart speaker and the corporation behind it on the one hand and by the researchers on the other. It may be assumed, however, that this reflection takes place implicitly and various stages are and thus data collection is also influenced by the location of the CVR in the living quarters. This fact can be regarded as example of the observer's paradox as discussed more often for linguistic research (Labov 1972; Koerfer 1985).
In sum, the development of our CVR can be seen as an illustration of how the data practices that are accomplished during the advancement of appropriate research designs, processing and subsequent analysis of data do not only influence the kind of research that can be done but also what research actually is being done. Thus, we add another component to the practices identified, among others, in the context of laboratory studies which science adopted from the social everyday world and condensed ("verdichtet") for the purpose of generating knowledge (Knorr Cetina 1988, 99): the data practices that inscribe themselves into the generation of knowledge (also see Ruppert et al. 2013). Tentatively and for our case, we can pinpoint identifying, representing and transforming as the corresponding practices which are performed in co-operation. Further studies will have to show whether these three operations also apply to other contexts.
In addition to the reflection on the data practices as part of our research practice, this paper aimed to document and uncover the development, advancement and use of the CVR as data collection technology for linguistic investigations. Further studies using the CVR technology are possible. However, our discussion demonstrates that the use of such a technology will always require customisation for the specific context in which it is used. The resulting efforts of co-operation and the data practices performed will deserve scholars' reflective attention in the future.

8. References

Agre, Philip (1994): *Surveillance and capture: Two models of privacy.* In: *The Information Society.* 10(2), 101–127.

Ammari, Tawfiq/Kaye, Jofish/Tsai, Janice/Bentley, Frank (2019): Music, Search, and IoT. *How People (Really) Use Voice Assistants.* In: *ACM Transactions on Computer-Human Interaction* 26(3), 1–28.

Bates, Jo/Lin, Yu-Wei/Goodale, Paula (2016): *Data journeys: Capturing the socio-material constitution of data objects and flows.* In: *Big Data & Society* 3(2), 1–12.

Beer, David/Burrows, Roger (2013): *Popular culture, digital archives and the new social life of data.* In: *Theory, Culture & Society* 30(4), 47–71.

Bittner (1967): *«The Police on Skid-Row: A Study of Peace Keeping».* In: *American Sociological Review* 32 (5): 699–715.

Bentley, Frank/Luvogt, Chris/Silverman, Max/Wirasinghe, Rushani/White, Brooke/Lottridge, Danielle (2018): *Understanding the Long-Term Use of Smart Speaker Assistants.* New York: ACM, 1–24.

Bergmann, Jörg (1985): *Flüchtigkeit und methodische Fixierung sozialer* Wirklichkeit: *Aufzeichnungen als Daten der interpretativen Soziologie.* In: Bonß, Wolfgang/Hartmann, Heinz (eds.): Entzauberte Wissenschaft: *Zur Relativität und Geltung soziologischer Forschung. Sonderband 3 der Zeitschrift Soziale Welt.* Göttingen: *Schwarz*, 299–320.

Bowker, Geoffrey (2008): *Memory Practices in the Sciences*, Cambridge: MIT Press.

Bowker, Geoffrey/ Star, Susan (2000): *Sorting Things Out: Classification and Its Consequences.* Cambridge: MIT Press.

Burkhardt, Marcus/van Geenen, Daniela/Gerlitz, Carolin/Hind, Sam/Kaerlein, Timo/Lämmerhirt, Danny/Volmar, Axel (in print): *Interrogating Datafication. Towards a Praxeology of Data.* Bielefeld: *Transcript.*

Coopmans, Catelijne/Vertesi, Janet/Lynch, Michael/Woolgar, Steve (ed.) (2014): *Representation in Scientific Practice* Revisited. Cambridge: MIT Press.

CRC Media of Cooperation (2019): *Research Program 2020–2023.* Online: https://www.mediacoop.uni-siegen.de/de/forschungsprogramm/ (last accessed: 2022-02-03).

Dalton, Craig/Thatcher, Jim (2014): *What Does A Critical Data Studies Look Like, And Why Do We Care? Seven Points for a Critical Approach to ‚Big Data'.* In: *Society and Space.* 12.05.2014. Online: https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care (last accessed: 2022-02-03).

Dalton, Craig/Taylor, Linnet/Thatcher, Jim (2016): Critical data studies: *A dialog on data and space.* In: *Big Data & Society* 3(1), 1–9.

Davidson, Christina (2009): *Transcription. Imperatives for Qualitative Research.* In: *International Journal of Qualitative Methods* 8(2), 36–52.

Dittmar, Norbert (2004): *Transkription. Ein Leitfaden mit Aufgaben für Studenten*, Forscher und Laien. Wiesbaden: Springer.

Duranti, Alessandro (2006): *Transcripts, Like Shadows on a Wall.* In: *Mind, Culture, and Activity* 13(4), 301–310.

Gießmann, Sebastian (2018): *Elemente einer Praxistheorie der Medien*. In: *Zeitschrift für Medienwissenschaft* 19(10), 95–109.

Gießmann, Sebastian (2020): *Identifizieren. Geschichte und Theorie einer Medienpraktik*. Working Paper Series Media of Cooperation 17, 1–14. Online: https://www.mediacoop.uni-siegen.de/de/publikationen/working-papers-media-of-cooperation/ (last accessed: 2022-02-03).

Gießmann, Sebastian/Burkhardt, Marcus (2014): *Was ist Datenkritik? Zur Einführung*. In: *Mediale Kontrolle unter Beobachtung. Datenkritik* 3(1), 1–13.

Gitelmann, Lisa (2013): *„Raw Data" Is an Oxymoron*. Cambridge: MIT Press.

Goodwin, Charles (2018): *Co-operative Action*. Cambridge: Cambridge University Press.

Haraway, Donna (1988): Situated Knowledges: *The Science Question in Feminism and the Privilege of Partial Perspective*. In: *Feminist Studies* 14(3), 575–599.

Habscheid, Stephan/Hector, Tim/Hrncal, Christine/Waldecker, David (2021): *Intelligente Persönliche Assistenten* (IPA) mit Voice User Interfaces (VUI) als ‚Beteiligte' in häuslicher Alltagsinteraktion. Welchen Aufschluss geben die Protokolldaten der Assistenzsysteme? In: *Journal für Medienlinguistik* 4(1), 16–53.

Hector, Tim (in print): Smart Speaker in der Praxis: Methodologische Überlegungen zur linguistischen medienlinguistischen Erforschung stationärer Sprachassistenzsysteme. To be published in: *Sprache und Literatur* 51(2).

Hind, Sam et al. (2021): Introduction. In: Götz, Magdalens/Hind, Sam/Lämmerhirt, Danny/Neumann, Hannah/Och, Anastasia-Patricia/Randerath, Sebastian/Seitz, Tatjana: In the Spirit of Addition: Taking a ‚Practice+'-Approach to Studying Media. Working Paper Series Media of Cooperation 18, 30–34. Online: https://www.mediacoop.uni-siegen.de/de/publikationen/working-papers-media-of-cooperation/ (last accessed: 2022-02-03).

Jänicke, Stefan/Franzini, Greta/Cheema, Muhammad/Scheuermann, Gerik (2017): Visual Text Analysis in Digital Humanities. In: *Computer Graphics Forum* 36(6), 226–250.

Knorr-Cetina, Karin (1988): *Das naturwissenschaftliche Labor als Ort der „Verdichtung"* von Gesellschaft. In: *Zeitschrift für Soziologie* 17(2), 85–101.

Koerfer, Armin (1985): *Zum Beobachter-Paradoxon in der Sprachwissenschaft*. In: Kürschner, Wilfried/Vogt, Rüdiger/Siebert-Nemann, Sabine (eds.): *Sprachtheorie, Pragmatik, Interdisziplinäres*. Tübingen: *Niemeyer*, 187–200.

Lämmerhirt, Danny (2021): *Health Data Donations as Cooperative Practice*. In: Götz, Magdalens/Hind, Sam/Lämmerhirt, Danny/Neumann, Hannah/Och, Anastasia-Patricia/Randerath, Sebastian/Seitz, Tatjana: *In the Spirit of Addition: Taking a ‚Practice+'-Approach to Studying Media*. Working Paper Series Media of Cooperation 18, 30–34. Online: https://www.mediacoop.uni-siegen.de/de/publikationen/working-papers-media-of-cooperation/ (last accessed: 2022-02-03).

Labov, William (1972): *Some principles of linguistic methodology*. In: *Language in Society* 1, 97–120.

Latour, Bruno (1990): *Drawing things together*. In: Lynch, Michael/Woolgar, Steve (eds.): Representation in Scientific Practice. Cambriwdge: MIT Press, 19–68.

Latour, Bruno (2000[1993]): *„Zirkulierende Referenz. Bodenstichproben aus dem Urwald am Amazonas."* In: Latour, Bruno (ed.): *Die Hoffnung der Pandora*. Untersuchung zur Wirklichkeit der Wissenschaft. Frankfurt a.M.: Suhrkamp, 36–95.

Lau, Josephine/Zimmerman, Benjamin/Schaub, Florian (2018): *Alexa, Are You Listening?* In: Proceedings of the ACM on Human-Computer Interaction 2, 1–31.

Luger, Ewa/Sellen, Abigail (2016): "Like Having a Really Bad PA". In: Kaye, Jofish/Druin, Allison/Lampe, Cliff/Morris, Dan/Hourcade, Juan (eds.): Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM, 5286–5297.

Lynch, Michael/Woolgar, Steve (eds.) (1990): Representation in Scientific Practice. Cambridge: MIT Press, 19–68.

McNally, Ruth/Mackenzie, Adrian/Hui, Allison/Tomomitsu, Jennifer (2012): Understanding the 'Intensive' in 'Data Intensive Research': Data Flows in next generation sequencing and environmental networked sensors. In: International Journal of Digital Curation 7(1), 81–94.

Merkle, Benedikt/Hector, Tim (in prep.): *„Alexa, Stop!"* Zur Beredsamkeit symbolischer Ausdrücke mit IPAs. In: Borbach, Christoph/Haensch, Konstantin/Kaerlein, Timo/Miyazaki, Shintaro/Stock, Robert/Wirth, Sabine (eds.): Akustische Interfaces. Interdisziplinäre Perspektiven auf Schnittstellen von Technologien, Sounds und Menschen. Wiesbaden: Springer-Vieweg.

Mosconi, Gaia/Li, Qinyu/ Randall, Dave/Karasti, Helena/Tolmie, Peter/Barutzky, Jana/Korn, Matthias/Pipek, Volker (2019): Three Gaps in Opening Science. In: *Computer Supported Cooperative Work* 28 (3–4), 749–789.

Ochs, Elinor (1979): Transcription as Theory. In: Ochs, Elinor/Schieffelin, Bambi (eds.): *Developmental pragmatics*. New York: Academic Press.

Pins, Dominik/Boden, Alexander/Stevens, Gunnar/Essing, Britta (2020): *„Miss Understandable"* – A study on how users appropriate voice assistants and deal with misunderstandings. In: *Proceedings of Mensch und Computer* 2020 (MUC 20). Magdeburg: ACM.

Porcheron, Martin/Fischer, Joel/Reeves, Stuart/Sharples, Sarah (2018): *Voice Interfaces in Everyday Life*. In: Mandryk, Regan/Hancock, Mark/Perry, Mark/Cox, Anna (eds.): *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* - CHI '18. New York: ACM Press, 1–12.

Rettberg, Jill Walker (2020): *Situated data analysis: a new method for analysing encoded power relationships in social media platforms and apps*. In: *Humanities and Social Sciences Communications* 7(1), 1–13.

Ruppert, Evelyn/Law, John/Savage, Mike (2013): Reassembling Social Science Methods: *The Challenge of Digital Devices*. In: *Theory Culture Society* 30(4), 22–46.

Ruppert, Evelyn/Scheel Stephan (2021): *Data Practices. Making Up a European People*. Cambridge: MIT.

Sands, Ashley E./ Borgman, Christine. L./ Wynholds, Laura/Traweek, Sharon (2012): Follow the data: *How astronomers use and reuse data*. In: Proceedings of the American Society for Information Science and Technology 49(1), 1–3.

Schatzki, Theodore/Knorr-Cetina, Karin/Savigny, Eike von (eds.) (2001): *The practice turn in contemporary theory*. London: Routledge.

Schönherr, Lea/Golla, Maximilian/Eisenhofer, Thorsten/Wiele, Jan/Kolossa, Dorothea/Holz, Thorsten (2020): *„Unacceptable, where is my privacy?"* Exploring Accidental Triggers of Smart Speakers. ArXiv-Preprint 2008.0058v1, URL: https://arxiv.org/abs/2008.00508 (last accessed: 2022-06-24).

Schüttpelz, Erhard (2020): Chils Prädikation. Ein Kommentar zu Charles Goodwins Co-Operative Action. In: Kablitz, Andreas/Markschies, Christoph/Strohschneider, Peter (eds.) (2020): Prädikation und Bedeutung. Berlin et al.: de Gruyter.

Schüttpelz, Erhard/Meyer, Christian (2017): Ein Glossar zur Praxistheorie. „Siegener Version" (Frühjahr 2017). In: Navigationen – Zeitschrift für Medien- und Kulturwissenschaften 17(1), 155–163.

Selting, Margret (2001): Probleme der Transkription verbalen und paraverbalen/prosodischen Verhaltens. In: Brinker, Klaus/Antos, Gerd/Heinemann, Wolfgang/Sager, Sven (eds.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 2. Halbband. Berlin/New York: de Gruyter, 1059–1068.

Selting, Margret et al. (2011): A system for transcribing talk-in-interaction: GAT 2. Translated and adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten. In: Gesprächsforschung Online 12, 1-51.

Sterne, Jonathan (2012): MP3: The Meaning of a Format. Sign, Storage, Transmission. Durham: Duke University Press.

Strauch, Annette/Hess, Volker (2019): Von der Produktion bis zur Langzeitarchivierung qualitativer Forschungsdaten im SFB 1187. In: Bibliothek – Forschung & Praxis 43(1), 101–116. Online: https://edoc.hu-berlin.de/handle/18452/20452 (last accessed: 2022-02-03).

Thielmann, Tristan (2018): *Die bewegte Mediengeschichte des Fotofahrtenführers*: ein Co-Motion-Picture. In: Ghanbari, Nacim/Otti, Isabell/Schramm, Samantha/Thielmann, Tristan (eds.): *Kollaboration. Beiträge zu Medientheorie und Kulturgeschichte der Zusammenarbeit*. Paderborn: Fink, 147–180.

Thurow, Joseph (2021): The Voice Catchers: *How Marketers Listen in to Exploit Your Feelings, Your Privacy, and Your Wallet*. Yale: Yale University Press.

Tkacz, Nathaniel/Henriqe da Mata Martins, Mário/Porto de Albuquerque, João/Horita, Flávio/Dolif Neto, Giovanni (2021): Data diaries: *A situated approach to the study of data*. In: *Big Data & Society* 8(1), 1–16.

Volmar, Axel (2017): *Formats as Media of Cooperation*. In: *Media in Action* 1(2), 9–28.

Zuboff, Shoshana (2018): *Das Zeitalter des Überwachungskapitalismus*. Frankfurt/New York: Campus Verlag.

9. Transcription conventions (Selting 2011)

Sequential structure

[ ] overlap and simultaneous talk

[ ]

In- and outbreaths

°h / h° in- / outbreaths of appr. 0.2-0.5 sec. duration

°hh / hh° in- / outbreaths of appr. 0.5-0.8 sec. duration

°hhh / hhh° in- / outbreaths of appr. 0.8-1.0 sec. duration

Pauses

(.) micro pause, estimated, up to 0.2 sec. duration appr.
(-) short estimated pause of appr. 0.2-0.5 sec. duration
(--) intermediary estimated pause of appr. 0.5-0.8 sec. duration
(---) longer estimated pause of appr. 0.8-1.0 sec. duration
(0.5)/(2.0) measured pause of appr. 0.5 / 2.0 sec. duration (to tenth of a second)

Other segmental conventions

: lengthening, by about 0.2-0.5 sec.
:: lengthening, by about 0.5-0.8 sec.
::: lengthening, by about 0.8-1.0 sec.
? cut-off by glottal closure

Laughter and crying

haha hehe hihi syllabic laughter
((laughs)) description of laughter and crying
<<laughing> > laughter particles accompanying speech with indication of scope
<<:-)> so> smile voice

Continuers

hm, yes, no yeah monosyllabic tokens
hm_hm bi-syllabic tokens
?hm?hm with glottal closure, often negating

Other conventions

((coughs)) non-verbal vocal actions and events
<<coughing> > ...with indication of scope
( ) unintelligible passage
(xxx), (xxx xxx) one or two unintelligible syllables
(may i) assumed wording
(may i say/let us say) possible alternatives

((unintelligible, appr. 3sec))	unintelligible passage with indication of duration
((...))	omission in transcript
->	refers to a line of transcript relevant in the argument
<<surprised>	interpretive comment with indication of scope

Accentuation

SYLlable	focus accent
sYllable	secondary accent

Final pitch movements of intonation phrases

?	rising to high
,	rising to mid
–	level
;	falling to mid
.	falling to low