Understanding the role of transcription in evidential consistency of police interview records in England and Wales

EMMA RICHARDSON [a], MAGNUS HAMANN [a],
JAMES TOMPKINSON [b], KATE HAWORTH [b]
AND FELICITY DEAMER [b]

[a]Loughborough University, England
[b]Aston University, England

ABSTRACT

Evidential records of investigative interviews serve an important institutional purpose within the legal system in England and Wales. Academic scholars have long recognized that little institutional attention is paid to the transformation process that occurs when written records of the spoken are produced, nor to the potential impact this has on later interpretation by users of the records during the investigation of crimes and later in court. We analyse twenty-nine digitally recorded investigative interviews and their corresponding official written 'Record of Taped/Videoed Interview' (ROTI/ROVI) transcripts, taking an ethnomethodological, conversation analytic (CA) approach to examine the social actions that are transformed in this activity by comparing the audio record of police interview evidence to the written transcripts. The intended outcome of this work, within the wider project of which this forms a part, is to better understand this process within the legal system, and to incite improvements. (Investigative interview, transcription, entextualisation, conversation analysis)

INTRODUCTION

The legal system, which is designed to uphold the law and ensure the delivery of justice, is a process conducted through language. The activity of investigating and prosecuting criminal activity is interactional, at each stage of the process. Given this, the language of the legal system has, for many years (see Bhatia 1987), attracted scholarly interest from linguists as well as from researchers working in other related areas such as psychology. In particular, much attention has been given to designing and evaluating models of interviewing and interview techniques (see e.g. Griffiths & Milne 2018; Shepherd & Griffiths 2021 for the input from psychology), including how interview questions are designed and issued (e.g. Heydon 2005; Grant, Taylor, Oxburgh, & Pankhurst 2015; Haworth 2017).

135

As a type of institutional discourse (Drew & Heritage 1992), these routinized, formal interactions in legal settings often abide by the turn-taking rules of the context, tied to the hierarchy of the setting. For example, the talk within investigative interviews is heavily structured with parties rarely diverting from a consistent interactional pattern (see e.g. Heydon 2005). This talk plays out with visible asymmetry (e.g. the professional parties talk more, including interrupting) (Momeni 2011). However, as we show in this article, when police transcribers produce official written records of the spoken interaction, much of this interactional detail is not included, meaning it is lost to any future observers, and the significance of such factors over the account produced by the interviewee is obscured.

The institutional function of the interview record

In England and Wales, there is a legal requirement for investigative police interviews with suspects to be audio-recorded (Police and Criminal Evidence Act 1984, Code E) with only a few exceptions (e.g. some terrorism cases). Sometimes, interviews are also video recorded. In addition to the audio/video record, a written version is also produced either 'in full' (an attempt to capture the talk 'verbatim') or, more commonly, a shorter version with parts summarised is produced. Written versions are referred to in the legal system as a 'Record of Taped/Videoed Interview' or ROTI/ROVI.

The mandatory audio recording of investigative interviews was introduced in response to records of investigative interviews being at best inaccurate (Dixon & Travis 2007) and at worst fabricated (see e.g. Coulthard 1996, 2002). Prior to the introduction of audio recording, the interviewer was expected to write down what happened post interview from notes and memory (see e.g. Coulthard 1996, 2002), which inevitably resulted in a record which contained substantially less detail and accuracy than the original interaction.

While the audio/video record remains to be the most accurate representation of what was said, producing a written record of the audio/video recording renders the interaction easily accessible to future audiences within the legal process (Walker 1990; Gibbons 2003). We even see, from the audio recordings of suspect interviews, parties orienting to these potential future audiences in their original spoken interaction (e.g. "For the benefit of the tape"; Stokoe 2009), who might be the investigating police officers, the Crown Prosecution Service (CPS), and members of the courtroom (Haworth 2013).

One of the central functions of the record in legal settings is for it to be tested and compared with accounts obtained at other points from the interviewee (given to a different audience, or in a different context, for example) and from others involved in the alleged crime. Our previous work (see Richardson, Haworth, & Deamer 2022) indicated that written texts are more 'referenceable', by which we mean easier for users to locate, mark, and reference part of the material from the whole than the audio/visual record, making them preferable to users over audio/visual materials.

Additionally, written records of the spoken suspect investigative interview are routinely presented in court in place of the audio or video recordings. The transcript is not only viewed and referred to as part of the prosecution evidence, but it is also used like a script, performed by members of the prosecution team, usually the prosecution lawyer and a police representative role-playing the interviewer(s) and the interviewee (see Haworth 2018). This reliance on the transcript rather than the audio/video recording is therefore deeply embedded in current legal practice. Although this is something we might seek to challenge and reform, this is an extremely ambitious target. Therefore, the focus of our current work is on ensuring the written record of the suspect interview is a more reliable representation of the original spoken interaction.

The process of entextualisation

Whether the record is an audio, video, or written text there are numerous, often unacknowledged, considerations when producing records of spoken interaction. To produce a transcription of a spoken interaction involves 'entextualization', 'the process of rendering discourse extractable, of making a stretch of linguistic production into a unit—a text—that can be lifted out of its interactional setting' (Bauman & Briggs 1990:73). This is later decontextualized, and then recontextualised—a process that Bauman & Briggs (1990) argue is an act of control and indexes social power. In creating records, 'producers' make decisions of focus, whether that is the positioning of the video cameras in an interview suite or police transcribers choosing what to include or exclude in the written transcript.

Haworth (2018) highlighted how interview evidence undergoes multiple significant transformations as it is captured and processed in the legal system. At each stage, the record is treated as an identical 'copy' of the original, without acknowledgement that the spoken data are substantially altered when converted into written format. Haworth (2018) makes a comparison between the practices and procedures associated with the handling of physical evidence (e.g. DNA, blood spatter, and fingerprint evidence) and the way in which spoken evidence is handled. Blackwell (1996:253) stated that 'the need for forensic linguists to develop an understanding of the transcription process is as pressing as ever'. Yet, there continues to be minimal recognition, certainly in legal contexts, of the non-equivalence of spoken and written text (Biber 1988; Halliday 1989).

The work detailed in the current article forms one part of a wider project, 'For the Record', comprising three distinct but interrelated strands. First, psycholinguistic experiments were designed to test our hypothesis that different formats (spoken/written) and transcription choices have an effect on interpretation (Deamer, Richardson, Basu, & Haworth 2022; Tompkinson, Haworth, Deamer, & Richardson 2023). Second, (reported here) we conducted a qualitative linguistic analysis of the two types of record, the audio/visual and the written. Third, focus groups with police transcribers and investigative interviewers were conducted to ensure

that the findings are firmly grounded in the practical realities of the professional context. The project as a whole is designed as a holistic examination of the process of producing investigative interview records, focusing in particular on the question of evidential consistency.

The quantitative findings from the first psycholinguistic experiment (Deamer et al. 2022) suggested that participants (representing potential jury members) presented with a transcript of a police interview were significantly more likely to (i) perceive the interviewee as *anxious* and *unrelaxed*, (ii) interpret the interviewee's behaviour as being *agitated*, *aggressive*, *defensive*, and *nervous*, (iii) determine that the interviewee was *un-calm* and *uncooperative*, and (iv) deem the interviewee's version of events to be *untrue,* than those participants presented with the original audio recording. Moreover, from the qualitative analysis, participants identified (a) *consistency*, (b) *phrase and lexical choice*, (c) *emotion (crying/upset),* (d) *hesitation and/or pauses* to be significant factors influencing their perception and interpretation of the interviewee and their story (Deamer et al. 2022). This highlights the potentially serious consequences of a legal system which relies on written records of the interview which do not include such features, rather than playing the original audio.

Building on Haworth (2018), and complementing the above findings, it is our intention in this article to highlight some of the ways in which producing these written records from the audio and video recordings, in fact, alters the content from what actually occurred in the original spoken interaction. Central to this is the process of transcription.

Transcription and existing notation systems

At the most basic level, transcription is the process of taking spoken language and producing from that, a written text. Transcripts can be produced by human transcribers or through automatic transcription software. All transcripts are reductionist, selective, and necessarily only capture details relevant to the institutional purpose they serve, some of which we have discussed elsewhere (Richardson et al. 2022; see also Fraser 2022). Transcripts are used across many settings for a wide range of purposes. Contrast, for example, a transcript that a journalist makes using specialist shorthand notation against a transcript made by a speech and language therapist to assist a client who has a swallowing disorder. Both transcripts serve to capture spoken language in written form, but the functions of these transcripts would be very different. The relationship between form and function is essential to ensuring that a transcript is fit for purpose. The function of a particular transcript will inform the specific aspects of speech that need to be captured in writing. Without full integration between both form and function, there will always be the potential for a transcript to be insufficient for its intended purpose. In the police interview context this can be especially difficult to achieve, due to the fact that these transcripts are used in various ways, as we have detailed elsewhere (Haworth 2013, 2018; Richardson et al. 2022).

Transcripts can be produced from 'live' interaction, although are more commonly produced from recorded speech, as in the case of ROTI/ROVI transcripts. As Fraser (2003) highlights, transcription is also inherently linked to human speech perception. Transcription relies on a listener perceiving and understanding the words that are spoken, and then having the appropriate system or tools to encode words and meaning within the transcript. Fraser's work on this topic (e.g. 2003, 2020, 2021, 2022) has elucidated some of the speech perception issues that can influence the quality of a transcript, including implicit listener-based biases, background noise, and/or poor-quality recordings, the continuous nature of speech, and the introduction of errors in hearing on the part of the transcriber.

Transcription practices differ in what details they capture and which they leave out, based on what the police transcriber considers important or relevant for the context, considering the purpose the record will serve and the intended user. As such, a good transcription system is able to capture those details that the institution is interested in and leave out those details that are deemed irrelevant. When it comes to transcription systems, their level of granularity varies significantly, from the broadest verbatim transcriptions to complex and detailed multimodal transcriptions that capture both the verbatim utterances, temporal, linguistic and paralinguistic resources, as well as embodied practices such as gaze direction, recognisable gestures, and limb/body movements (see Mondada 2018).

The simplest type of transcripts (orthographic transcripts) usually aims to represent only the words that are spoken, with minimal or no information about aspects of speech such as stress placement, intonation, emotion, or other extra-linguistic markers such as clicks and sniffs. Orthographic transcripts are most appropriate when the central function is capturing WHAT was said and capturing information about HOW words were spoken or the potential meaning of aspects of spoken interaction are of limited or no importance. One area of the legal system in which orthographic transcripts are frequently deployed is in so-termed 'forensic transcripts' (Ariani, Sajedi, & Sajedi 2014; Fraser 2018, 2020, 2021). These are transcripts of often poor-quality recordings which are produced with a primary function of assisting juries, lawyers, and judges to understand indistinct speech. Given that the purpose of an orthographic transcript is to represent the words spoken, and that orthographic transcripts are most commonly produced for general audiences with no specific linguistic or specialist knowledge, they are designed to contain no specialist notation and have a high level of general comprehensibility. The trade-off for this is that they lack standardised methods of representing aspects of speech which cannot be captured by words alone. For example, even within more 'formal' transcription systems, O'Connell & Kowal (1994) highlight variation in how aspects of speech such as emphasis, pitch, and loudness have all been marked using italics, capitalisation, and underlining.

Jefferson's transcription system (2004) is one that is widely used in social sciences, as it allows the transcriber to 'catch details' of actual occurrences of interaction which a researcher's imagination could not produce as hypotheticalized-

typical versions and assert as real (see Sacks 1992). This system aims to capture a number of details which are heard in talk but are not represented in an orthographic transcript. The system uses symbols common to any word processor to capture the timing and placement of talk, the onset of overlaps, pauses in turns, or gaps between turns. It also captures the sound qualities such as when a speaker stretches a sound or speeds up a word, emphasis, quietness/loudness, marked pitch changes, in and out breaths, laughter/crying, cut-off words or sounds, and other intonational features of the talk (Drew 2005).

A further development of the Jeffersonian transcription system is Mondada's (2018) system for transcribing embodied actions as they are coproduced with spoken utterances. Mondada (2018) argues that transcribing cannot be separated from analysis and proposes transcription conventions which aim to implement the principles of multimodal analysis in a systematic, coherent, robust, and explicit way. The system enables the transcriber to focus in on an unlimited range of embodied conducts and anchor them in within the temporality of action as it unfolds.

Another example of a specialist transcription system is the International Phonetic Alphabet (IPA), designed to capture the sounds of the languages and give the user a one-to-one sound-to-symbol notation system to provide users with a method to finely capture HOW words are produced by speakers. The handbook of the IPA (International Phonetic Association 1999) identifies that the goal of the system is to ensure that sounds can be transcribed and represented consistently, and it provides an unambiguous set of symbols for the phonetic transcription of both segmental and suprasegmental aspects of speech. This has a clear advantage over the imposition of these kinds of features in an unstructured way in orthographic transcripts, but the disadvantage that it requires users, including readers, to have specific training in IPA transcription.

As we have shown, there are so many possible ways to transcribe interaction; choices are made based on the most appropriate system for the intended purpose. For our own analysis presented here, we use the Jefferson (2004) transcription system, employed by conversation analysts, since—in line with the point made above—we feel this best suits our purpose here, in terms of the level of detail and the type of interactional feature we seek to analyse. We use CA since it is crucial to our analysis to consider the subtle nuances of the talk, as these can be consequential for the way recipients hear and respond to talk. In turn, this is consequential for the way talk is organized (Sidnell 2010). A Jeffersonian transcript renders those interactional features analysable and referenceable; it also enables us to present clear and direct comparisons with the official transcripts.

Further, in CA, transcripts are not treated as 'data' as such; data are the original audio or video recordings and the transcripts a representation of it (Hutchby & Wooffitt 2008). This is because in CA transcripts are acknowledged to be selective in the details represented and thus are never treated as a replacement for the data (Hepburn & Bolden 2012). We do not claim that a Jeffersonian transcript is more 'neutral' or 'objective' than others listed above; as Fraser (2022) rightly

points out, it is not. However, Jefferson transcription tries to capture as much of that interactional nuance as is available on the recording, for the purpose here of evidencing transformation of *social actions*, while also acknowledging that any transcriber (ourselves included) is unlikely to capture all relevant information when transcribing. This strongly aligns with our position that a transcript should never be taken as a straightforward replacement for the original, as we see routinely occurring in the legal process.

DATA AND METHODS

Data

We were granted access by a police service located in the England and Wales jurisdiction to digitally recorded investigative interviews and their corresponding transcripts. The dataset comprises twenty-five audio-recorded suspect interviews and four achieving best evidence (ABE) video-recorded interviews with victims or witnesses, and their accompanying transcripts. Witness interviews and their records play a different role in the criminal justice system, especially ABE video interviews which can be played in court as the witness's evidence in chief instead of them giving that evidence in person (s.27 Youth Justice and Criminal Evidence Act 1999). However, a transcript is still produced in the same way as with suspect interviews, so we included them in our analysis of police transcription practices. The digital recordings and transcripts were made as part of the routine work of the organization and not recorded for research purposes. The processing and use of the data is governed by a data sharing agreement between the police service and us as research partners. All personal information was redacted from the interviews on site at the police station, and the recordings anonymized. Where names, locations, and dates are shown in the data extracts presented here, they are pseudonyms. All data were stored on encrypted systems that are specialized for forensic and legally sensitive materials. The project received ethical approval from the university of the research team.

We acknowledge the limitation that this only includes data from one police force, and we are aware that each police force has its own practice in the production of interview records. We therefore make no claims of generalisability here; however, this approach enables us to gain deep insight into professional practice in this one location, as well as producing findings and outcomes which are maximally relevant to our partner force.

The method of analysis

For the purposes of this strand of the wider research program, we take an ethnomethodological approach (the study of how people produce social order through their social interactions (Garfinkel 1984) and draw on CA. CA is both a field of study and a research method for analyzing naturally occurring conversation, which has shown

that, and how, conversation is systematic (not 'messy'); that there is 'order at all points' (even in one single turn at talk), and that much of its core machinery is universal across languages (Stivers, Brown, & Levinson 2009).

Conversation analysis requires 'the study of naturally occurring activities as they ordinarily unfold in social settings, and, consequently, on the necessity of recordings of actual situated activities for a detailed analysis of their relevant endogenous order' (Mondada 2012:33). In essence, to understand the full range of practices including not just the words uttered but pauses, silences, spontaneous talk, pace, non-lexical vocalisations, and so on, it is necessary to examine actual recordings; the next step is then to evaluate the ways in which the social interaction observable in the recording is encapsulated (or not) in the official record. In so doing, we recognise and acknowledge that the recording is itself merely a partial version of the original; however, our purpose here is specifically to examine the process of converting these institutionally produced recordings into the institutionally sanctioned official transcripts.

An inductive approach to analysis was employed, informed by the findings from the preliminary experiments and focus groups. Our analysis proceeds to make available, systematically, the activities that comprise the complete interactions; the way those activities are designed, and how different designs lead to different trajectories and outcomes for speakers. In addition to describing the production of the talk by speakers, conversation analysis identifies the social actions that are performed by the speakers as they interact in real-time and, by documenting those using the Jefferson transcription system, can evidence the substantial change in the social actions that occurs in the transformative process of presenting these in the official written records of the spoken interaction.

ANALYSIS

Our analysis comprises four sections. First, we introduce the written record and describe the organization of this document. We then move to three sections analyzing the content of the record, focusing specifically on how silence, temporal features of the speech, and emotion are represented in the transcriptions.

The layout

The standard template on which records of taped interviews (ROTIs) are produced is the Manual of Guidance form number 15 (known as the MG15). An example of an MG15 is provided below. The form is laid out conforming to European culture of literacy norms which 'socializes its members to encode ideas not only from top to bottom, but from left to right of the writing surface' (Ochs 1979:49).

At the top of the form (see Figure 1), information about the suspect or witness is populated from information stated by the parties during the preliminary phase of the audio record. This includes verbalised details about the time, duration, and place of

142   *Language in Society* **54**:1 (2025)

Form MG15

RECORD OF INTERVIEW

Record Type: RODI

URN

Person interviewed: ███████████	Police Exhibit No: ███
Date of Birth:	Number of Pages: **4**
Place of interview: ██████ Police Station	Signature of transcriber producing exhibit
Date of interview: ████████	✎ ████████████

Time commenced: 1328 **hours**	Audio Tape Reference No: ████████
Time concluded: 1345 **hours**	Visual Image Reference No:
Duration of 17 **minutes** interview:	-
Interviewer(s): ████████████	-
Other persons present: ████████████	

Tape counter times	Person speaking	Text
		Introductions. Right to free and independent legal advice. Cautioned and explanation given.
3.20		Officer explains ████████ has been arrested on suspicion of an aggravated burglary which occurred at ██████████ ████████████
	Q	What can you tell me about that?
	R	I don't really know where that's come from at all.
		When asked for her movements last night ████████ states she and ████ were at a friend's house ████████ watching a movie when she received a phone call from the police asking her to attend ████ Police Station. ████ father was also at the house at the time. ████ lives on the ████████████ (number not

Page 1 of 4

FIGURE 1. Interview 4. Example of the MG15 (ROTI).

the interview, who conducted the interview, and any other present parties. There are three columns on the form: from left to right there is first a space to record tape counter times; next, in the middle, a column to record 'person speaking'; and on

the right, a column titled 'text' where the transcriber records what was said in the spoken interview. The text column is segmented into rows.

Our examination of the ROTIs revealed that although the form is standardised, there is variation in how the form is used by police transcribers to record the spoken interview. We note variation not only across the records, but within them. In the following sections, we describe the inconsistency present. Figures are taken from the official ROTIs and extracts are the research team's own representations of the talk using a Jeffersonian (2004) transcription system. Relevant transcription conventions are provided throughout.

The column labelled 'Tape counter times'.   The first of the three columns on the MG15 is titled 'tape counter times' (see Figure 1), despite the 'tape counter' now being obsolete technology as, in this force at least, they are now all digitally recorded. In this column, times are recorded in minutes and seconds which allows the user of the transcript to locate the text (transcribed in the third column) within the audio or video file. We observe no common practice for when transcribers link a textbox to the audio recording by inputting a time; transcribers do not write a time in the column for each textbox. The most common practice is to not write the time in the box (see Figure 1). When transcribers do note a time, we observe three different practices: to approximate where the text is on the audio record (Figure 2), to indicate the beginning of a new sequence (Figure 3) and to indicate where there is a point of interest in the record (Figure 4).

The first way in which we note this practice being used is to approximate where in the audio file this text corresponds. The police transcriber periodically records rounded time intervals (e.g. entering 8:00 at the eight-minute mark, and then 11:00 at the eleven-minute mark as shown in Figure 2). However, we do not find these to be included at consistent intervals.

The second way in which we see the tape counter times column being used by police transcribers is to enter a time at the beginning of a new sequence (as shown in Figure 3), for example, when the interviewer asks a specific sequence-initiating question about a particular part of the incident. Where the audio is heavily summarised this might result in frequent time entries in the tape counter times column. Additionally, we see times in this column when the police transcriber moves from one mode of recording to another, so from a summarised sequence, to the transcriber typing out what the speakers have said word-for-word.

The final way we observed this column being used was to indicate when something about what is being said in the interview is considered to be 'important' enough to require a reference (as shown in Figure 4).

The column labelled 'Person speaking'.   Any transcript needs to reflect which of the turns are attributed to which speaker. In this form, there is a column titled 'Person speaking' to show who the talk is attributed to. In our analysis, we find multiple ways in which the different speakers are represented. The first is to

Form MG15

		lockdown. Asked if he was aware that movement outside of his home was only for certain reasons, SMITH stated no comment.
08:00		There was a further allegation that on 23 July some number plates were stolen from another vehicle, a Renault Scenic, FW63 EUT, from 55, The Woods, Nearside in Midsizecity. The number plates were later found attached to the stolen vehicle mentioned previously. Asked if he had anything he wished to say regarding this, SMITH stated no comment.
		SMITH stated no comment when asked the following questions; Did you attend The Woods, Nearside on 17 July 2020. Did you enter the driveway of that property. Did you have to move anything to get onto the driveway. Have you got any reason to go on that driveway. Could you see any vehicles on the driveway of that address. Did you interfere with any vehicles in any way. Did you remove any number plates on any vehicles that might have been on the driveway. Did you speak to anybody at that address and specifically speak to the occupants of that address and tell them you were at the wrong address when they came out and asked what you were doing. Did you take any number plates away from that address and what was your intention with those number plates. Did you attach those number plates to a stolen vehicle; more specifically the vehicle which you have been asked about already. Will there be any of your DNA on the number plates and any reason why there would be. Why would you steal number plates.

Page 3 of 7

FIGURE 2. Interview 12. Tape counter times: approximation.

145

		Why did you change the number plates on the car already mentioned with the ones you had stolen.
11:00		The next incident was in relation to a burglary which occurred in Lowton near Highton on 19 July 2020 between 3.30 pm and 4.30 pm.
		SMITH stated no comment when asked the following questions. Do you know where Lowton is and have you ever been there. Have you ever been to Blue Street, Lowton and ever attended Treetop House, Blue Street, Lowton. Were you there on 19 July. Do you have any reason to attend that address. How did you get there. Did you drive a stolen vehicle, already spoken about, and where did you park it. Did you see anybody leave that property whilst on that street. Were you waiting for people to leave the address. Did you approach the property at all and did you go in the back garden for any reason. Did you force any windows on the property. What did you use to force the rear window on the property and why. Did you enter the property through the window mentioned.
		SMITH made no reply when asked the following questions: Why did you enter the property through the rear ground floor window. Did you have any permission to enter that property and would you say you were a trespasser in the property.

FIGURE 2. Continued.

Form MG15

		Stella.
9.52		JACKSON stated the children had gone to their step dad's so they had the house to themselves.
10.13		He stated there wasn't any cross words between them and it was like when they first got together. JACKSON stated they were both enjoying themselves.
11.25		JACKSON stated he noticed July's black eye when he woke up. It was her right eye. He stated he also noticed a mark on her hips. JACKSON stated when he asked her about the marks she told him they had been dancing the night before and they had fallen over.
12.18		JACKSON stated when he went downstairs he noticed some glass on the floor in the living room. He stated this is when he thought they must have been dancing the night before.
12.44		He stated he can remember them listening to music and having a slow dance.
13.03		JACKSON stated on the Saturday things were fine between him and July.
13.10		He stated he met someone knew and he thinks this is why July messaged him.
14.25		JACKSON stated he cannot remember them falling over.
14.47		He stated they both had consumed a fair amount of alcohol on the Friday night. He stated on a scale of 1 to 10 drunkenness he put himself at a 7 or 8.
15.17		PC HUNTER summarises JACKSON'S account.
9.14		PC HUNTER refers to July's statement
	Q	She's saying that about 6 weeks ago she's admitting to talking to

Page 4 of 8

FIGURE 3. Interview 25. Tape counter times: Marking a new sequence.

use a 'question-answer' or a 'question-response' practice to mark the 'speech act' of the textbox. Here, 'Q' is used to represent 'questions' and either 'R' or 'A' is used to represent the 'response' or 'answer'. This practice can be observed in Figure 4. The

7.20	Q	How have you go there?
	R	I drove.

FIGURE 4. Interview 4. Tape counter times: Point of interest.

other ways in which transcribers assign 'person speaking' to the adjacent text box is to use the interviewer's collar number, as shown in Figure 5, or to use the speaker's last name. This practice is often, but not always, used when there is more than one interviewer putting questions to the interviewee.

10:30	R	Yes.
	2412	Yeah. Okay. Can you tell me how these images got on your phone.
	R	I can't remember.
	2412	You can't remember okay. Did you download these images.
	R	Yeah.

FIGURE 5. Interview 8. Use of collar number.

The column labelled 'Text'.   In the 'Text' column, police transcribers use rows to represent what was said in the audio or video recording. They can do this by either recording word for word what was said or by producing summaries of the talk. It is not within the scope of this article to comment on how police transcribers make choices about what to type word for word and what to summarise. Here, we seek to present the way in which the text column is used and consider any inconsistent practice.

In Figure 6 below, we see most of the text boxes contain summarized talk, but, between two blank rows, a few turns have been typed out word for word. Note that '…' has been used to show a transition from summary to verbatim transcription.

We found that most often a new row is used for a new speaker, but this is not always the case. Sometimes, in conjunction with an entry in the tape counter times column, a new row with text is created to indicate a point of interest within the same speaker's turn (see Figure 7).

As shown in this image, this practice can create artificial segregation of turns-at-talk. Positioning the "yeah I've done it" separately in its own text box, with a time in the 'Tape counter times' column, visually highlights these words.

			When it was described how he had a go at Kieran who lived next door, DAVIDSON admitted he did that. He disputed the police took him away and said they went to him in the kitchen.
8:15		Q	…so that's what he says
		R	Yeah, yeah, that's fine
		Q	It's fair to say that you don't agree with that
8:19		R	Not at all, not one bit
			Jack BRAITHWAITE'S statement was then put to DAVIDSON, who questioned the gate and said he was nowhere near the gate.

FIGURE 6. Interview 14.

It also creates the risk that they are taken out of context and thereby misinterpreted, especially in terms of their evidential strength. The chosen layout has separated this part of the utterance from the very important conditional "if there's enough evidence to say I've done it I'll put my hands up and say…". It is only the lack of a new attributed speaker in the 'person speaking' column which enables us to infer that the highlighted talk follows on directly from

			Form MG15
			honestly know if, if someone said to me oh is this yours, you know what I mean, like a load of money or summate like that I don't know, I wouldn't say oh yeah it's mine I've lost it, so I don't honestly know. So if people are saying they see me do it then I put my hands up I've done it, but I don't remember actually doing it which is a bit frightening on my part. That's the truth and that is the truth. But I was annoyed that the motor was still there so, but I'm not saying, I'm not saying either way. All I'm saying is if there's enough evidence to say I've done it I'll put my hands up and say
6.10			yeah I've done it.

FIGURE 7. Interview 7.

the text box above. The layout choices made here by the transcriber thus give the misleading impression that an interviewee has made an admission which he did not.

Transcription resources

For the most part, police transcribers enter the words spoken into the boxes, but we also see a range of 'transcription resources' being used to indicate other interactional features such as periods of silence and temporal features of the talk, like overlapping speech and turns that are 'latched' to the prior. Our analysis revealed varied practice within and across the transcripts of how these are used which we unpack in the forthcoming sections, informed by the experimental work (Deamer et al. 2022; Tompkinson et al. 2023. We first consider the importance of silence in interaction and show the practices police transcribers have for representing silence. Second, we show how police transcribers attempt to capture temporal features of talk and provide commentary on what is gained and lost in doing so. Finally, in the last analytic section, we consider the representation of 'emotion' in the police interview.

Silence. In CA, silence is the absence of talk in relation to the sequential organization of turns (Hoey 2017). Table 1 below shows how, in CA, we use the Jeffersonian transcription system to represent silence and how we use it to demonstrate inconsistency in police transcriber practice.

TABLE 1. *Jefferson notation relevant for silence (Hepburn & Bolden 2013).*

(1.4)	The length of silence is recorded in parenthesis.
(.)	Silence below two tenths of a second is a 'micropause' and is recorded as a period in parenthesis.

A Jefferson (2004) transcript records silences to the tenth of a second, above two tenths. This enables the reader of any Jefferson transcript to understand how long a silence was, and the positioning of the silence enables the reader to know if the transcriber heard this as a pause within a turn or a gap between (Sacks, Schegloff, & Jefferson 1974). This is a feature often analysed as interactionally salient in CA.

We find police transcribers do attempt to represent episodes of silence heard in the audio files on the written record. Figure 8 shows an example.

In Figure 8 the square boxes indicate three places in which the police transcriber has represented SOMETHING taking place between parts of the suspect's ('R') talk using a series of periods. However, it is not clear from this transcript alone whether '…' is representing silence. To better understand this, we examined the

	Q	While they're on your external hard drive how often do you look at them?
	R	I don't, I don't ... this is probably the most stupid thing that people in this room are going to try and believe but I don't ... I look at them as have they completed and say yes, yes ... I might delete some straight away and then I put them in and then I don't look at them again. It's stupid really.
	Q	What do you think when you're looking at them?

FIGURE 8. Silence: Interview 20.

audio and then used the notations from Table 1 to produce a Jefferson transcript of the same talk (see extract (1)).

(1)   Silence: Interview 20 (INT: interviewer; SUS: suspect)
   1   INT: .hh >While they are on < your external hard drive?
   2   (0.2)
   3   INT: How often >do you look at them<.
   4   **(2.2)**
   5   SUS: I don't.
   6   **(.)**
   7   SUS: I £know- .hregh. .hh heh£ (0.2) £th(h)at is pr(h)ob(h)ably(h)
   8   .hhh hh £ (.) .hh- £the most stupid thing that people in this
   9   are going to try and belie:ve?=But£ **(0.5)** I don't.
 10   SUS: Now I- I- (.) I'll look at them: .hhhhh (as when) they're
 11   completed and say yes yes u::h **(0.2)** I might delete some
 12   straight awa:y and then I'll put them in and then: (0.4) I
 13   don't look at them again.
 14   (1.4)
 15   SUS: Stu[pid really?
 16   SUS: [What >do you think < when you are looking at them:,

From listening to the original audio, and producing the above transcript, we can observe many episodes of silence in the talk. Silences which were represented in the official record are shown in bold font, and those which were not represented, but which our analysis revealed, are highlighted in grey. Of those which the police transcribers did represent, two are slightly misplaced in the official record when we compare Figure 8 and lines 4 and 9 of extract (1). We might ask why are some and not others represented?

Silences are categorized by Sacks and colleagues (1974) in three ways. The first of these are 'pauses', a silence within a speaker's turn construction unit (TCU).

These are recorded at the point of silence. The next are 'gaps', which are silences between TCUs, at a transition-relevance point (TRP) and, in CA transcripts, are recorded on a new line before the next, or first, speaker continues. The last category of silence as categorized by Sacks and colleagues (1974) are 'lapses': at a TRP where the first speaker has not nominated a new speaker, they do not continue themselves and unlike a gap no new speaker takes the turn for such a duration that there is a 'lapse' in the flow of conversation.

Therefore, the Jefferson transcription enables us to distinguish between pauses, gaps, and lapses (visible through how they are represented in the Jefferson transcript), and therefore we can better understand the function the speakers are performing by pausing. The comparison of the two transcriptions enables us to see that not all 'pauses' (multiple between lines 7 and 12), 'gaps' (lines 2, 4, 6), or 'lapses' (line 14) present in the talk (audible on the audio record) are reproduced in the written record. We are not able to understand from this analysis why some might be included over others, or what the impact of including or excluding them might have on future audiences. What we can say is, without a standardized method for representing silence, and with limited transcription resources, it is not surprising to see inconsistency in representing episodes of silence.

We show in the next example where the police transcriber has chosen not to represent episodes of silence, as shown in Figure 9.

Q	Tell me what you're doing
A	Washing and yeah just washing my hair in one of them and it looks like in this one probably washing my private parts
Q	Can you describe where you are in those images

FIGURE 9. Interview 10.

If we then consider a Jefferson transcription of this same stretch of talk, we see there were some extended periods of silence.

(2)   Interview 10 (INT: interviewer; VIC: victim)
  1 INT: Tell me what you're doin'
  2 VIC: Uh:m washing and (2.2) yeah just washing my hair in
  3   one of them,
  4   (7)
  5   and it looks like I- in this one probably washing
  6   uh:m,
  7   (1.1)
  8   >sort of< private parts.
  9 INT: Ok.= Uh (0.8) can you describe where you are in those images.

In the Jefferson transcript, we have recorded a pause within the interviewee's turn construction unit (TCU) at line 2 and in the interviewer's turn at line 9. At the end of the victim's turn at line 3, there is a transition-relevance point (TRP), although the comma indicates that the speaker's turn does not end with falling intonation. Therefore, this gap is recorded on a separate line, as another speaker (here, the interviewer) could have said something in this space but chooses not to. This is also true of the gap at line 7. Compare this with Figure 9, where the speech appears continuous.

When we, as analysts, listen to and transcribe this stretch of talk we see there are multiple pauses and gaps of significant length which are not represented by the police transcriber. We might question, why? In this case, this interview has been video recorded. In the video record, we see (as the transcribers would) that during the seven second gap (line 4), and the shorter gap at line 7 the interviewee is leaning forwards and gazing toward images that are laid on the table. We cannot know from this study why the police transcriber did not choose to use '…' as in Figure 8/extract (1). However, in those previous examples we see there is some repair, some laughter, and some trouble producing the account. Here, the lengthy pauses directly relate to the interviewee moving her body into a position to view the images that she has been asked to describe.

Silence represented by '…' in some places/transcripts and not in others highlights how police transcribers are making judgements about when to represent pauses and gaps and when not to. We might assume that silence is recorded when it is considered to be interactionally important, or interesting, in the transcribers' opinion.

However, with minimal transcription resources available, we identify a practice of using '…' to do more than show pauses but also to represent a number of temporal features of the talk, which we unpack in the next section.

Other temporal and sequential features

CA, and the Jefferson transcription system, has revealed how precisely we as speakers co-ordinate our talk when interacting with others (Sacks et al. 1974). Some of the conventions to represent temporal features of the talk are represented in Table 2 below.

In this section, we consider how police transcribers attempt to represent temporal features of the talk using their limited transcription resources to represent the audible features. We first consider overlap, then cut-off talk, and finally latched talk.

Overlap. In our dataset we found that, as well as representing silence, police transcribers used a series of periods ('…') to represent overlapping speech when they transcribe sections of the interview verbatim. Figure 10 shows multiple uses of '…' in this extract of Interview 17.

TABLE 2. *The Jefferson notation relevant for temporal features.*

Over[lap] [Over]lap	The onset [and offset] of overlapping talk is encased by square brackets.
Latched= =talk	Marks no discernible silence between two turns. The absence of the 'beat of silence' (Jefferson 1984; Schegloff 2000)
Cut off-	The dash indicates the turn was cut off.
<	Arrows are used to indicate the <slowing down> or >speeding up< of talk. One < at the beginning of a turn indicates a 'quick' start, without a 'beat of silence'. Can be used in place of =.

		at her, she threw it at you. She says it was over moving the carpet that you'd discussed the day before and that you were in bed and had tooth ache and didn't want …
	R	Yeah …
	Q	…yeah …
	R	a very strong tooth …
	Q	…and didn't wanna get out. She said when she threw it back at you,

FIGURE 10. Interview 17.

To explore what these might be representing, we produced a Jefferson transcription of this same talk below.

(3)   Interview 17
    1   INT: She says it was over (.) moving the carpet. An',
    2   SUS: .Sni[hh]
    3   INT: [An::] that you'd discussed the day befo:re, and that
    4   you were in bed and had a (0.3) toothache?
    5   [And didn't wanna- yeah.]
    6   SUS: [Yeah very b- d-   ] strong tooth,
    7   INT: ((COUGH)) and didn't wanna get out.

If we compare the two extracts, in the Jefferson transcript we see the exact placement of the overlap (denoted by [ ]), which differs substantially from the officially produced transcript. Each of the text boxes shown in Figure 10 features the '…' convention. As this has also been used to represent silence, when we 'view' Figure 10, we could imagine a stretch of talk with periods of silence between the turns, as opposed to the direct overlap audible in the recording that we have identified.

			the dog. Okay. There were six videos of that description on your phone okay, so six videos of people having sex with dogs. Do you remember seeing those at all.
		R	No.

FIGURE 11. Interview 8.

CA has demonstrated that as speakers we project possible completions of TCUs and co-ordinate our turn-taking to minimise overlap, ensuring our talk begins at a transition relevance place (TRP) (Sacks et al. 1974). Therefore, when we do enter another's turn it is often to perform a social action. In this case, we see on line 4 the interviewer completes their turn, with questioning intonation, but then continues on line 5. However, the interviewee has begun to produce their own turn in overlap. Here, we see the interviewer quickly 'drop-out' of that overlap to minimize occasions of speaking over each other (Hepburn & Bolden 2012).

In Figure 11, and the corresponding transcript we have produced ourselves in extract (4), we show another kind of social action being performed with overlap. Please be advised of sensitive content in this example.

(4) Interview 8
```
     1   INT:   There were six videos, (0.7) of that (1) description on
     2          your (0.2) phone, okay. So s- six videos of people having
     3          sex with dogs. Okay. Do you remember seeing those at
     4          [all. ]
     5   SUS:   [No. ]
```

At line 5, the interviewee (SUS) projects the completion of the interviewer's turn and the early onset "no" (line 5) is produced (Jefferson 1984). The position of this performs a social action, to assert that the suspect does not remember seeing these images as early as possible (Vatanen 2018). From our experimental work, conducted as part of this wider project, we found participants made judgements about interviewees based on the temporal production of their talk. These judgements included how 'credible' and 'sincere' the interviewee was and how plausible their account sounded. When asked to justify their ratings, participants made comments in free text boxes about how quickly the interviewee 'latched' their answers to the officer's questions, or how they hesitated in their production (Deamer et al. 2022). These temporal features were commented on regardless of whether the participant was a reader of, or listener to, the extract. Their inclusion or omission from an official interview transcript is therefore potentially of consequence, especially if we consider how a jury might interpret the interaction.

8:57	Q	....(inaudible) he's saying you were....
	R	....Jack come running out from (inaudible) the house, out to the front of the gate, he was stood inside the gate, at the gate
	Q	Yeah
	R	With the...with the spade

FIGURE 12. Interview 14.

Cut off talk. Across the corpus, we observed the police transcriber practice of also using '…' to indicate a cut off, trailing off or in some way unfinished turn. An example of this is in Figure 12 below, indicated by the timestamp 8:57. Here we see the interviewer talk transcribed as "…. (inaudible) he's saying you were….". Then, in extract (5) we show a Jefferson transcription of the same talk.

(5)   Interview 14
    1   INT: Wh- when ( ) because he's saying you were:-
    2   SUS: <Jack's come running out from round the side of the house.
    3   Out to the front of the gate. He was stood inside the
    4   gate.<At the gate.
    5   (0.4)
    6   INT: Yeah-
    7   SUS: <With the, with the spade.

We have shown that elsewhere (Figure 6), '…' at the beginning of an entry with an empty box prior was used to indicate that the transcriber is moving from summarising to verbatim mode of transcription. We see this same practice being used here, but by producing the Jefferson transcript, we reveal this is being done to represent a different phenomenon. Listening to the audio we find the '…' at the end of the interviewer's ('Q') turn marked with the time stamp 8:57, and at the beginning of the interviewee's response ('R'), is being used in an attempt to represent that the interviewer cut off their own turn, and before they could proceed the interviewee began their turn. Interestingly, as shown in our own transcript, almost the same interactional phenomenon occurs in the immediately following turns (lines 6 and 7), but here the transcriber chooses not to use '…', or any other transcription convention, to capture this.

What we have shown in this section is that the police transcribers' practice of using '…' has many functions in the official transcript and is used inconsistently. This makes it challenging, if not impossible, for the reader to correctly interpret the meaning intended by the transcriber. It is through such mechanisms that meaning is thus lost or distorted through the transcription process. This is clearly of concern in a context where 'the reader' includes those tasked with determining

the outcome of criminal proceedings against participants on the basis of such evidential documents.

Emotion

We do not find emotion represented in the official transcripts in our dataset. However, in our previous experiments, we did find participants comment on audible 'emotion' when making judgements about interviewees. In this section, we choose to focus on laughter and crying in interaction, first by examining how it functions in interaction by representing it using Jefferson transcription conventions. CA considers the display of emotion as an interactional accomplishment (Hepburn 2004). Therefore, episodes of laughter or crying are analysable as part of the social action being performed, and decisions as to whether or not they are included in a transcript are of consequence.

Laughter.   First, there is a difference between what Goffman (1961) referred to as 'flooding out' laughter, which is uncontainable and disruptive to the talk, and laughter as placed (Jefferson 1985). There has been some limited discussion of the form and function of laughter in the police investigative interview (Stokoe & Edwards 2008; Stokoe 2009; Carter 2013). In Table 3, we show the Jefferson transcription conventions for laughter, which we use to highlight the placement of laughter in extract 6 below.

TABLE 3. *Jefferson transcription notation for laughter (Jefferson 1984, 2004).*

heh heh/hih hih/hah ha	Voiced laughter
£smile£	Pound-sterling signs on either side of talk can indicate suppressed, or pre-laughter or smiling made audible through the co-occurrence with talk (see Haakana 2010).
La(h)ugh(h)ter	Parenthesized 'h' indicates plosiveness. Can indicate laughter, crying and the breathy production of words.

To demonstrate how a speaker might use smiling/laughter as an interactional resource, we have produced a section of an interview where the interviewer asks a suspect of a car theft how they have come to be in possession of the keys to the car.

```
(6)   Interview 21
      1   INT:   How have you come to be in possession of those
      2          keys in the first plac[e,
      3   SUS:                   [s: s-
      4          S:::omeone £gave me the keys.£
      5          (0.2)
      6   INT:   Rig[ht,
```

```
 7  SUS:       [Asked me to dri:ve,
 8             (4.8)
 9  INT:       >Are you < willing to name that <person>?
10  SUS:       £ hn .hgef. is j- (come) on£.
11             (0.2)
12  SUS:       £That was a [stupid question£].
13  INT:                   [>I've got     ]
14             to a[sk<?
15  SUS:       [Yeah (go ahead).
```

In the interviewee's answer at lines 3–4, they produce the latter part of their turn with a 'smiley voice', indicating the question is being treated by the suspect as humorous. Talk enclosed with '£' represents an audible quality which can be heard as 'smiley voice' even when the facial expressions are not available to the interacting parties, such as when on the telephone (Jefferson 1984; Lavin & Maynard 2001). The design of this turn at line 4 is to attend to the implication that the person who gave the keys to the interviewee is either the owner, or the thief, of the car. At line 9, where the interviewer asks the suspect if they are "willing to name that person?", which would likely clarify the latter two positions, the suspect produces laughter particles ("hn .hgef.", line 10) before "come on" and continuing in smiley voice to say "£That was a [stupid question£]". The interviewer enters the suspect's turn at line 12 with "I've got to ask", accounting for the question as procedural. Stokoe (2009) found in police interviews laughter can mitigate the lack of affiliation associated with not responding adequately to a question. In this extract, by answering, the suspect would presumably incriminate another party. We can understand the interviewee's laughter to attend to this issue. We now examine how this same talk was recorded in the official written transcript.

		Q	How have you come to be in possession of those keys in the first place?
		R	Someone gave me the keys and asked me to drive.
			ROY stresses he is not willing to name this person.

FIGURE 13. Interview 21.

In Figure 13 above, the initial question (lines 1–2 of extract (6)) and response (lines 4 and 7 of extract (6)) have been typed out by the police transcribers 'verbatim', in that the words spoken are (essentially) preserved. Without the accompanying audio, readers would not have access to the 'smiley voice' audible in the delivery of R's response. We emphasize that we do not necessarily expect police audio transcribers to record this level of detail, we instead aim to draw attention to the fact that this omission leaves it to the reader to interpret how the answer

was delivered. We have shown in our above analysis of this short extract the inter-actional work being done by interviewer and interviewee here, which is lost in the official version.

In the next text box, we see the transcriber move from verbatim to producing a summary of the next stretch of talk. From listening to the audio, we know that the interviewer asked a follow-up question, "Are you willing to name that person?" (line 9) to which the interviewee had responded with laughter and talk in smiley voice. However, here the police transcriber has interpreted this as "ROY stresses he is not willing to name this person". This is the police transcriber's interpretation of the talk between lines 10–12 in extract (6). It could be that their interpretation of the speech is that it is stressed in some way, or that terms such as 'said' and 'stressed' are used by transcribers interchangeably. It is not possible to know from the current analysis alone. However, here we demonstrate the subjective interpretation of what is audible in the recording and suggest that this is possibly a somewhat misleading one.

Crying. Similar to laughter, in CA crying is seen as interactionally organised, and transcribing the crying helps analysts understand the form and function (see Table 4). One way to represent the crying is to do as Manzo, Heath, & Blonder (1998) did in their study of stroke patients, which is to label that crying has taken place. Whalen & Zimmerman (1998) attempted to move beyond this and represent the features of callers' distress when telephoning 911. Since then, Hepburn (2004) and later Hepburn & Potter (2012) have built on this and the work of Jefferson (1985) on laughter, subscribing to the need for transcription to do more than simply to note that crying had occurred. This is the approach we take here to represent the organisation of crying for our purposes. Hepburn (2004) and Hepburn & Potter (2012) outline the interactional features of crying and present the use of relevant transcription notations. Again, we are not advocating for police transcribers to use these methods; we employ them here for our own analytic purposes of highlighting exactly where crying is present in the audio.

TABLE 4. *Jefferson transcription notation relevant for crying (Hepburn 2004; Jefferson 2004; Hepburn & Potter 2012).*

.snih	Sniffs, transcribed as an inbreath.
°whisper°	Hearably quieter speech can indicate that a speaker is struggling to produce their turn.
↑ wh::y	Elevated pitch (e.g. ↑why) could indicate speaker is upset.
hhh .hhh	'Sobbing' disrupts the talk and like laughter is transcribed as in/out breaths with 'voiced vowels' (Huh/hih).
~ wobbly ~ #creaky#	What are commonly known in CA to be 'wobbly' and 'creaky' voice can indicate issues with production of speech when crying.
Brea(h)th	Parenthesized 'h' indicates plosiveness. Can indicate laughter, crying, and the breathy production of words.

In our experimental work, participants made judgements about interviewees based on crying when it was audible, and when it was represented in the written record (Deamer et al. 2022), indicating its potential significance in legal proceedings. Although we identified episodes of crying across our audio and video corpus, we did not find this to be recorded in the official police transcripts. Extract (7) is an example of an audibly distressed interviewee. We have used the transcription notation to show where crying is present within the talk.

(7) Interview 17
```
 1   SUS:   =~ >Sometimes I- < I will just (.) lost my (min:d) and nn n-
 2          don't do (0.2) w- n wu- bad things = like this morning (.)
 3          I wake up and this morning .hhh .snih. and she told me
 4          something (0.2) °en°- and #e- hh n:# d that happened~
 5          ~ ~becau = f:: °#e-n#° (1.0) .hnn-.hh- .ha- she: = O:h my god
 6          °#e-n-#° ~ ~,
 7          (0.9)
 8   INT:   Oka:y.=
 9   SUS:   =~ ~I don't want (.) >I don't < want her to hurt = because I
10          love he:r ~ ~.
11          (2.8)
12   SUS:   ~ ~°Uh: [huh- huh::° ~ ~
13   INT:           [Alright,
14          (1.0)
15   SUS:   ~ ~°uh:- Hu:h:::° ~ ~  ~ ~I come for- [I- ~ ~
16   INT:                                        [>So what happened<
17          today.
18   SUS:   ~ ~Ih:- ~ ~
```

Our transcription of this section of talk makes visible, sniffs, wobbly and creaky voice and sobs throughout. Between lines 3 and 4 the interviewee is having trouble producing their turns, which by line 5 and 6 has escalated into crying and an abandoned turn at talk. At line 12 and the beginning of line 15, the interviewee is audibly sobbing ("~ ~°Uh: [huh- huh::° ~ ~"), with a wobbly voice (~) and quietly (°°). But when we look at what was recorded in the official transcript in Figure 14, it is not represented that the interviewee was crying.

Our focus here is again on the use of the three periods. The police transcriber has attempted to represent SOMETHING after the interviewee says, "oh my God…" with these three periods, and again we see the same practice used here before the talk, "…I don't want". However, there is a time stamped "Okay" delivered in between by the interviewer ('Q'). Based on what we have seen elsewhere, '…' could represent that the interviewer's turn was delivered in overlap, or there was silence. Therefore, it is not clear from this record that the interviewee was too distressed to complete his sentence, or that there was sobbing during his talk which we have shown in extract (7).

160   *Language in Society* **54**:1 (2025)

	R	...she's called me bad, bad, bad, sometimes I will just lost my mind and do, do bad things. Like this morning, I wake up in this morning
		and she told me something and that happened because, she, oh my God ...
06:46	Q	Okay.
	R	...I don't want, I don't want to hurt her because I love her.
	Q	Alright.
	R	I come for ...
	Q	So what happened today?

FIGURE 14. Interview 17.

Understanding the interactional implications of distress and upset is particularly useful for studies of institutional settings. A party being so upset that they are unable to respond to turns can be highly consequential when 911 call-takers are gathering information, as Whalen & Zimmerman (1998) have shown. Understanding the responses to crying such as the use of tag questions by child protection officers can facilitate tellings of child abuse in calls to a child protection helpline (Hepburn & Potter 2010), and understanding how police interviewers deal with distressed, intellectually impaired child and adult victims of sexual offences (Antaki, Richardson, Stokoe, & Willott 2015) is conducive to more effective information gathering and better experiences for all. Our finding that crying and distress are not represented in our corpus of official police transcripts is therefore of potential concern, in that it removes this interactionally significant feature from the record relied upon by many future audiences in the legal system and beyond. This strengthens our position that it is far preferable for audiences to hear the recording, especially at the point when any extracts are being presented as evidence in court.

DISCUSSION

Through our analysis, we have shown inconsistencies, omissions, and distortions in official police interview transcripts. While the features that we identify here might not be consequential for justice in every case, the very possibility that they could influence judicial outcomes is surely sufficient cause for concern. Our intention in highlighting these issues here is to make them empirically available for producers and users, and to lay foundations for further academic and practitioner work, especially with regard to moving towards more standardisation of practice.

It is important to remember that police transcripts are not produced for linguistic analysis (Blackwell 1996); and to acknowledge that the purpose of the investigative

interview record is a complicated matter (see Haworth 2020). In a previous article (Richardson et al. 2022) we asked five questions of transcripts in legal settings: (i) Is the record produced an accurate representation of the spoken interaction?; (ii) Who has ownership of the record?; (iii) Who has agency, that is, whose 'voice' is represented in the recorded account?; (iv) How usable is the record?; and (v) How resource efficient it is to produce? In the current article we have explored the first of these points in more detail.

We have used CA and the Jefferson (2004) transcription system to reveal police transcribers' current practice. We previously outlined (Deamer et al. 2022) how the focus tends to be more on reproducing content, to ensure all or most of WHAT was said is transcribed, and less on HOW it is said, borne out in our analysis.

We have shown how police transcribers do attempt to represent certain features, such as silence, overlap, and cut-off turns, but they do so inconsistently, making use of available transcription conventions repeatedly, most notably, the series of periods (…).

In the absence of standardised guidance and training, police transcribers are left to use their personal judgement on when to include HOW, in addition to WHAT, was said. The difficulty here is for readers of the transcripts to know how to interpret the notations that are included. It is not possible to interpret the difference between indications of silence, overlap, moving from summarising to verbatim transcription, or crying if they are all denoted by '…'. Meaning therefore cannot be reliably communicated in this way from transcript creator to user.

Ultimately, the best way to solve the issues we highlight in this article would be for audiences to listen to the audio recording, rather than relying so heavily on a transcript. However, at least for now we have to accept the practical reality that transcripts do, and will continue to, have a role in legal proceedings. Therefore, we might consider recommending consistently including what scholars have found to be lacking in legal transcriptions: features such as disfluency in speech, overlapping talk, pauses, stress, and/or intonation (Bucholtz 2007; Fraser 2014; Komter 2019). However, we are reluctant to do so without knowledge of how this would benefit the users of the transcripts. Certainly, it is not as straightforward as recommending transcription conventions. As Deamer et al. (2022) has shown, adding a feature into a transcript, such as timed pauses, may in fact make this more salient to the user than its occurrence in the original recording, potentially distorting the interpretation as opposed to preserving it as closer to the original version. We are conducting experimental work to investigate this further, some of which is reported in Tompkinson et al. (2023).

The balance to be struck in collaboration with our police partners sits in the understanding that records do not need to be a 'perfect' representation of the interaction that took place. That is not reasonable nor is it required. Records need only to be adequate for the purpose(s) they serve (see Fraser 2022; Richardson et al. 2022), and it must also be feasible to keep pace with the demands and constraints of the justice process. Yet, at the heart of this is the fact that these records are ultimately used to secure the conviction of those charged with criminal offences, and as such, a

focus should be on ensuring that if they are to remain a central feature of the legal system, end users are made more aware of their limitations, and of the fact that any transcript is only one of innumerable possible interpretations of the interaction it purports to represent. As Fraser (2022:3) emphasises, they are viewing A transcript and not THE transcript.

CONCLUSION

In this article, we use conversation analysis and Jefferson transcription conventions to empirically demonstrate and make available to practitioners and academics alike the transformation that can occur in the process of producing written records of police investigative interviews. This novel contribution systematically highlights the issues present in consistency of transcripts, which are produced and used as evidence every day in the England and Wales criminal justice system, despite the availability of the audio/video recording.

We intend this work, as part of a wider program of research and in collaboration with our police partners, to contribute to guidance and training to improve consistency in the production of transcripts, and we also endeavour to increase awareness within the legal system of the benefits of using audio/video recordings in preference to their traditional reliance on transcripts to be something which they cannot be.

REFERENCES

Antaki, Charles; Emma Richardson; Elizabeth Stokoe; & Sara Willott (2015). Dealing with the distress of people with intellectual disabilities reporting sexual assault and rape. *Discourse Studies* 17(4):415–32.

Ariani, Mohsen Ghasemi; Fatemeh Sajedi; & Mahin Sajedi (2014). Forensic linguistics: A brief overview of the key elements. *Procedia - Social and Behavioral Sciences* 158:222–25.

Bauman, Richard, & Charles L. Briggs (1990). Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology* 19:59–88.

Bhatia, V. K. (1987). Language of the law. *Language Teaching* 20(4):227–34.

Biber, Douglas (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Blackwell, Susan (1996). Corrective measures: Some aspects of transcription in the British legal system. In Hannes Kniffka (ed.), *Recent developments in forensic linguistics*, 255–76. Frankfurt: Peter Lang.

Bucholtz, Mary (2007). Variation in transcription. *Discourse Studies* 9:784–808.

Carter, Elizabeth (2013). *Analysing police interviews: Laughter, confessions and the tape*. London: Continuum.

Coulthard, Malcolm (1996). The official version: Audience manipulation in police records of interviews with suspects. In Carmen Rosa Caldas-Coulthard & Malcolm Coulthard (eds.), *Texts and practices: Readings in critical discourse analysis*, 166–78. London: Routledge.

——— (2002). Whose voice is it? Invented and concealed dialogue in written records of verbal evidence produced by the police. In Janet Cotterill (ed.), *Language in the legal process*, 19–34. Basingstoke: Palgrave Macmillan.

Deamer, Felicity; Emma Richardson; Nabanita Basu; & Kate Haworth (2022). For the record: Exploring variability in interpretations of police investigative interviews. *Language and Law / Linguagem e Direito* 9(1):25–46.

Dixon, David, & Gail Travis (2007). *Interrogating images: Audio-visually recorded police questioning of suspects*. Sydney: The Sydney Institute of Criminology.

Drew, Paul (2005). Conversation analysis. In Kristine L. Fitch & Robert E. Sanders (eds.), *Handbook of language and social interaction*, 71–102. Mahwah, NJ: Lawrence Erlbaum.

———, & John Heritage (1992). Analyzing talk at work: An introduction. In Paul Drew & John Heritage (eds.), *Talk at work: Interaction in institutional settings*, 3–6. Cambridge: Cambridge University Press.

Fraser, Helen (2003). Issues in transcription: Factors affecting the reliability of transcripts as evidence in legal cases. *Forensic Linguistics* 10:203–26.

——— (2014). Transcription of indistinct forensic recordings: Problems and solutions from the perspective of phonetic science. *Language and Law / Linguagem E Direito* 1(2):5–21.

——— (2018). Forensic transcription: How confident false beliefs about language and speech threaten the right to a fair trial in Australia. *Australian Journal of Linguistics* 38(4):586–606.

——— (2020). Forensic transcription: The case for transcription as a dedicated branch of linguistic science. In Malcolm Coulthard, Alison May, & Rui Sousa-Silva (eds.), *The Routledge handbook of forensic linguistics*, 416–31. London: Routledge.

——— (2021). Forensic transcription: Legal and scientific perspectives. *Studi AISV* 8:1–19.

——— (2022). A framework for deciding how to create and evaluate transcripts for forensic and other purposes. *Frontiers in Communication* 7. Online: https://www.frontiersin.org/articles/10.3389/fcomm.2022.898410/full.

Garfinkel, Harold (1984). *Studies in ethnomethodology*. Cambridge: Polity Press.

Gibbons, John (2003). *Forensic linguistics: An introduction to language and the justice system*. Oxford: Wiley-Blackwell.

Goffman, Erving (1961). *Encounters*. New York: Bobbs Merrill.

Grant, Tim; Jennifer Taylor; Gavin E. Oxburgh; & Gary Pankhurst (2015). Exploring types and functions of questions in police interviews. In Gavin E. Oxburgh, Trond Myklebust, Tim Grant, & Rebecca Milne (eds.), *Communication in investigative and legal contexts: Integrated approaches from forensic psychology, linguistics and law enforcement*, 39–54. Oxford: Wiley-Blackwell.

Griffiths, Andy, & Rebecca Milne (eds.) (2018). *The psychology of criminal investigation: From theory to practice*. London: Routledge.

Haakana, Markku (2010). Laughter and smiling: Notes on co-occurrences. *Journal of Pragmatics* 42 (6):1499–1512.

Halliday, M. A. K. (1989). *Spoken and written language*. Oxford: Oxford University Press.

Haworth, Kate (2013). Audience design in the police interview: The interactional and judicial consequences of audience orientation. *Language in Society* 42(1):45–69.

——— (2017). The discursive construction of evidence in police interviews: Case study of a rape suspect. *Applied Linguistics* 38(2):194–214.

——— (2018). Tapes, transcripts and trials: The routine contamination of police interview evidence. *The International Journal of Evidence & Proof* 22(4):428–50.

——— (2020). Police interviews as evidence. In Malcolm Coulthard, Alison May, & Rui Sousa-Silva (eds.), *Routledge handbook of forensic linguistics*, 144–58. London: Routledge.

Hepburn, Alexa (2004). Crying: Notes on description, transcription, and interaction. *Research on Language and Social Interaction* 37(3):251–90.

———, & Galina B. Bolden (2012). The conversation analytic approach to transcription. In Jack Sidnell & Tanya Stivers (eds.), *The handbook of conversation analysis*, 57–76. Oxford: Wiley-Blackwell.

———, & Jonathan Potter (2010). Interrogating tears: Some uses of 'tag questions' in a child protection helpline. In Alice Freed & Susan Ehrlich (eds.), *Why do you ask? The function of questions in institutional discourse*, 69–86. Oxford: Oxford University Press.

———, & ——— (2012). Crying and crying responses. In Anssi Peräkylä & Marja-Leena Sorjonen (eds.), *Emotion in interaction*, 195–211. Oxford: Oxford University Press.

Heydon, Georgina (2005). *The language of police interviewing: A critical analysis*. Basingstoke: Palgrave Macmillan.

Hoey, Elliot (2017). *Lapse organization in interaction*. Nijmegen: Max Planck Institute PhD dissertation.

Hutchby, Ian, & Robin Wooffitt (2008). *Conversation analysis: Principles, practices, and applications*. 2nd edn. Cambridge: Polity Press.

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jefferson, Gail (1984). On the organization of laughter in talk about troubles. In J. Maxwell Atkinson & John Heritage (eds.), *Structures of social action: Studies in conversation analysis*, 346–69. Cambridge: Cambridge University Press.

——— (1985). An exercise in the transcription and analysis of laughter. In Teun A. Van Dijk (ed.), *Handbook of discourse analysis, vol. 3: Discourse and dialogue*, 5–34. London: Academic Press.

——— (2004). Glossary of transcript symbols with an introduction. In Gene Lerner (ed.), *Conversation analysis: Studies from the first generation*. Amsterdam: John Benjamins.

Komter, Martha (2019). *The suspect's statement: Talk and text in the criminal process*. Cambridge: Cambridge University Press.

Lavin, Danielle, & Douglas W. Maynard (2001). Standardization vs. rapport: Respondent laughter and interviewer reaction during telephone surveys. *American Sociological Review* 66:453–79.

Manzo, John; Robin L. Heath; & Lee Xenakis Blonder (1998). The interpersonal management of crying among survivors of stroke. *Sociological Spectrum* 18:161–84.

Momeni, Negar (2011). Police genre: Interruption and its classification as a sign of asymmetry in police interview/interrogation. *International Journal of Criminology and Sociological Theory* 4(1):615–25.

Mondada, Lorenza (2012). The conversation analytic approach to data collection. In Jack Sidnell & Tanya Stivers (eds.), *The handbook of conversation analysis*, 32–56. Oxford: Wiley-Blackwell.

——— (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction* 51(1):85–106.

Ochs, Elinor (1979). Transcription as theory. In Elinor Ochs & Bambi Schieffelin (eds.), *Developmental pragmatics*, 43–72. New York: Academic Press.

O'Connell, Daniel C., & Sabine Kowal (1994). Some current transcription systems for spoken discourse: A critical analysis. *Pragmatics* 4:81–107.

Richardson, Emma; Kate Haworth; & Felicity Deamer (2022). For the record: Questioning transcription processes in legal contexts. *Applied Linguistics* 43(4):677–97.

Sacks, Harvey (1992). *Lectures on conversation*. Volumes 1 & 2. Oxford: Basil Blackwell.

———; Emanuel Schegloff; & Gail Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4):696–735.

Shepherd, Eric, & Andy Griffiths (2021). *Investigative interviewing: The conversation management approach*. 3rd edn. Oxford: Oxford University Press.

Sidnell, Jack (2010). *Conversation analysis: An introduction*. Oxford: Wiley-Blackwell.

Stivers, Tanya; N. J. Enfield; Penelope Brown; & Stephen C. Levinson (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106 (26):10587–592.

Stokoe, Elizabeth (2009). 'For the benefit of the tape': Formulating embodied conduct in designedly unimodal recorded police–suspect interrogations. *Journal of Pragmatics* 41(10):1887–1904.

———, & Derek Edwards (2008). Did you have permission to smash your neighbour's door? Silly questions and their answers in police–suspect interrogations. *Discourse Studies* 10(1): 89–111.

Tompkinson, James; Kate Haworth; Felicity Deamer; & Emma Richardson (2023). Perceptual instability in police interview records. *International Journal of Speech, Language and the Law*, to appear.

Vatanen, Anna (2018). Responding in early overlap: Recognitional onsets in assertion sequences. *Research on Language and Social Interaction* 51(2):107–26.

Walker, Anne Graffam (1990). Language at work in the law: The customs, conventions, and appellate consequences of court reporting. In Judith N. Levi & Anne Graffam Walker (eds.), *Language in the judicial process*, 203–44. New York: Plenum Press.

Whalen, Jack, & Don H. Zimmerman (1998). Observations on the display and management of emotion in naturally occurring activities: the case of ‘hysteria’ in calls to 9-1-1. *Social Psychological Quarterly* 61:141–59.

Address for correspondence:
Emma Richardson
Loughborough University
School of Social Sciences and Humanities
Brockington Building
Loughborough, LE11 3TU, UK
e.richardson@lboro.ac.uk