



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/214921/>

Version: Published Version

---

**Article:**

Tompkinson, James, Haworth, Kate, Deamer, Felicity et al. (2023) Perceptual instability in police interview records: Examining the effect of pauses and modality on people's perceptions of an interviewee. *International Journal of Speech, Language and the Law*. pp. 22-51. ISSN: 1748-8885

<https://doi.org/10.1558/ijsll.24565>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



# Perceptual instability in police interview records: Examining the effect of pauses and modality on people's perceptions of an interviewee

James Tompkinson<sup>a</sup>, Kate Haworth<sup>a</sup>, Felicity Deamer<sup>a</sup>  
and Emma Richardson<sup>b</sup>

## Abstract

*This article examines whether the representation of linguistic features within transcripts and audio recordings of police interviews can influence people's perceptions of the interviewee. We specifically examine the influence of the representation of pauses through an experimental methodology. Participants were presented with a police interview either in audio format or in one of a series of transcript formats and asked to make a series of judgements about the interviewee. We manipulated both the presence and representation of pauses within the audio and transcript stimuli to assess how this would influence perceptions. Results showed differences between perceptions of the interviewee in the audio and transcript conditions, and that different representations of pauses within transcripts created perceptual instability between participants. The findings illustrate that the presence and representation of linguistic features in transcripts can affect perceptual judgements. We argue this should be explicitly considered by those using transcripts within the legal system.*

**KEYWORDS:** POLICE INTERVIEWING, HUMAN SPEECH PERCEPTION, TRANSCRIPTION, PAUSING BEHAVIOUR

## Affiliations

<sup>a</sup> Aston Institute for Forensic Linguistics, Aston University, UK <sup>b</sup> School of Social Sciences and Humanities, Loughborough University, UK  
emails: [j.tompkinson@aston.ac.uk](mailto:j.tompkinson@aston.ac.uk) [k.haworth@aston.ac.uk](mailto:k.haworth@aston.ac.uk) [f.deamer@aston.ac.uk](mailto:f.deamer@aston.ac.uk)  
[e.richardson@lboro.ac.uk](mailto:e.richardson@lboro.ac.uk)

Received: 21 November 2022 Accepted after revision: 5 May 2023

## Introduction

Research spanning almost a century has investigated how different aspects of language use can, often baselessly, influence social evaluations of a speaker (Allport and Cantril 1934; Coupland and Bishop 2007; Mileva et al. 2018). However, this is yet to be critically and thoroughly explored in the context of police interview records, where language is the primary source of the communication of evidence both in the interview itself and when interview evidence is presented in courtrooms. Taking an experimental approach, the focus of this article is to examine the role that judgements of an interviewee can play within the context of police interview records, and how this might impact on their use within the legal system. Specifically, we explore whether the presentation of a police interview in audio or written form affects perceptions of the interviewee, alongside the effects of varying how pauses are represented in interview transcripts.

## Background

Previous research has illustrated that people form impressions about others quickly (Willis and Todorov 2006), and that a range of factors can influence the impressions that they are prepared to make. These include, for example, a person's face (Oosterhof and Todorov 2008; Mileva et al. 2018) or voice (Zuckerman and Driver 1989; Belin et al. 2011). This research area has predominantly assessed impressions of unknown speakers, ensuring that existing familiarity is not a factor in the decision-making process. Focusing on vocal characteristics, it has been shown that a range of features including pitch (Ohala 1984; Tsantani et al. 2016; Mileva et al. 2018), voice quality (Laver 1994) and regional accent (Giles 1970; Coupland and Bishop 2007) can all influence perceptions of various personality and emotional traits. Such judgements have also been applied to legal concepts such as the perception of guilt in mock juror experiments (Dixon et al. 2002; Dixon and Mahoney 2004), as well as in the perceptions of legally relevant traits such as threat and intent-to-harm (Watt et al. 2013; Tompkinson 2018).

Previous research has shown that personality judgements can also be made from written texts. For example, McAndrew and De Jonge (2011) show that perception of anger in written emails was linked to increased use of third person pronouns, and that punctuation markers influenced impressions of a writer's emotional state. In an earlier study, Erickson et al. (1978) presented participants with audio and written versions of the same courtroom trial testimony and found differences between the two modalities in interaction with the sex of the speaker. The male speaker was judged to be less attractive than the female speaker in the audio modality, but their testimony was judged to be more believable than the female speaker's testimony when the modality was written. Furthermore, Murphy

(2007) found that perception of intelligence was affected by modality. In Murphy's (2007) study, video recordings were made of actors who were instructed to speak in a way that made them appear more intelligent. These actors were not provided with instructions to modify specific aspects of voice in order to sound intelligent, but were instead told to use their own judgement in order to appear 'smart, competent, and bright' (Murphy 2007: 328). These were produced alongside control stimuli where speakers received no instructions about how they should sound. Transcripts of the recordings in both sets were also produced, with participants assigned to either an audio or transcript condition and asked to rate how intelligent the speaker was. A difference was found between the two audio conditions but not between the two transcript conditions.

Relatively little attention has been paid to the role that person perception can play within the context of police interview records. This is the applied focus of this article, where our overall aim is to explore how judgements made about interviewees differ between an audio-recorded police interview and corresponding interview transcripts. Under the Police and Criminal Evidence (PACE) Act (1984), there is a legal requirement for all police interviews of suspects in England and Wales to be audio- or video-recorded. The motivation behind this change in legislation was to enhance transparency and to provide a move away from evidence drawn from memory or notes made by interviewing officers (Haworth 2018). In theory, this recording process would remove any doubt over what was said in the interview, and an exact copy would be preserved for use throughout the subsequent investigative and legal processes. Following the completion of an interview recording, a request can be made for a transcript to be produced of the interview. These transcripts, known as Record of Taped Interview (ROTI) or Record of Video Interview (ROVI) transcripts, are routinely produced by police-force employees for both investigative and evidential uses (Haworth 2018).

However, Haworth (2018) states that, despite these transcripts being treated as 'an unproblematic copy' of the interview recording, several processes can take place to essentially contaminate the original evidence. These problems include, but are not limited to, differences in the way that certain features are represented in the transcript, inaccurate or incomplete summaries of evidence, the lack of inclusion of features relating to emotion or pausing, and the subsequent presentation of this amended evidence in courtrooms when transcripts are then read aloud by legal representatives (Haworth 2018; Deamer et al. 2022). Richardson, Haworth and Deamer (2022) illustrate how ROTI transcripts compare to interview information in other modalities, highlighting that, while transcripts may be easier to use, they can lack accuracy and give additional power and agency to police employees (primarily the transcriber) compared with audio or video recordings.

Surprisingly, in England and Wales, there is currently no national guidance or set of standards relating to how ROTI transcripts should be produced, formatted or presented in courtrooms. Richardson et al. (2022) advocate for greater reflection and more research into transcript production, specifically to guard against some of the unqualified assumptions that can be made about transcripts. The most significant of these assumptions would be that ROTI transcripts are treated as being equivalent to the original spoken interaction within the legal system, despite this being a wholly inadequate assumption to make about this type of evidence (Haworth 2018).

In a small-scale experiment, Deamer et al. (2022) began to address the question of whether the modality of the representation of a police interview (audio or transcript) causes differences in the ways in which the interviewee is perceived. Deamer et al. (2022) found a range of significant differences depending on whether the interview was represented in audio or written form. Those who read the transcript perceived the interviewee as more anxious, less relaxed, more agitated, more nervous, more defensive, less calm, less cooperative and, perhaps most importantly, less likely to be telling the truth. Participants also identified a range of language and speech features including emotion and pausing, which they felt had influenced their judgements of the interviewee.

Pausing behaviour has been the subject of numerous linguistic investigations over several decades of research, both from a speech production and a speech perception perspective. For example, Duez (1982) highlighted the functional nature of pauses in different conversational styles, while Fox Tree (2002) examined how the introduction of a silent pause at turn intervals can contribute to perceptions of a speaker being both less comfortable with the topic of conversation, and also less honest. Fox Tree's (2002) work also highlighted the distinction between *filled* and *silent* pauses, and argued for them to be treated as separate phenomena in perceptual research. Later work by Lundholm Fors (2015) highlighted the differences in pause production between speakers, and evidenced the notion that perceptually pauses can influence memory and cognitive processing of utterances. Summarising the importance of pauses in conversation, Lundholm Fors (2015: 13) argues that 'silences can make or break the conversation'.

Given the previously highlighted importance of silences and pausing behaviour in conversation, it is relevant to consider how this could apply in forensic and legal settings, where speech information is routinely presented in transcript form. The representation of pauses is part of the Jeffersonian transcription system for conversation analysis, where silent pauses longer than 0.2 seconds are transcribed in (X.X sec) format, and shorter silent pauses are represented in (.) format (Hepburn and Bolden 2013). In police interview transcripts, Richardson et al. (submitted) highlight the use of '...' as a common representation and/or interpretation of

a pause, although the authors note that this is not always used or interpreted in a consistent manner. Kendall (2007) found a similar issue with the use of ‘--’ as a potentially confusing representation for a pause within courtroom transcripts, given that it was shown to be used to represent other aspects of speech in addition to pauses.

As well as variation in the representation of pauses, research has explored the relative merits of including silent pauses in legally relevant transcripts. Kendall (2007: 324) states the difficulty with balancing the requirement of a courtroom transcriber to go ‘beyond the verbatim’ to accurately convey meaning, while also remaining objective. Walker (1986) explicitly explores the relationship between so-termed ‘verbatim’ transcripts in the US court system and perceptions that are formed about the speakers from courtroom transcripts. In a survey where 45 judges were asked whether certain features should appear in transcripts, Walker found that a majority of those asked were in favour of representing non-verbal features in courtroom transcripts. This range of features included silent pauses, with 85% of judges in the survey in favour of the inclusion of pauses in transcripts. Kendall (2007) argues that the introduction of silences into a courtroom transcript can be essential for a reader to obtain a fuller understanding of the event, citing the example of a 32-second silence in a courtroom trial which corresponded to people turning pages to find a specific piece of documentation. Kendall (2007: 330) calls for pauses to be included in courtroom transcripts to maintain ‘temporal accuracy’, which he argues is ‘the most important component of a transcript’.

However, Walker (1986) highlights that, while participants were in favour of including pauses in transcripts, there was more disagreement over *how* pauses should be represented, with some judges arguing that pauses should only be included if they were timed, and others arguing that the inclusion of a pause should be determined by the speaker’s style and should only be included if they were considered to be longer than that speaker’s normal pauses. Walker (1986: 423) neatly summarises this tension, arguing that ‘while in the main they [judges] express their desire that a transcript accurately reflect as far as possible the exact meaning of the event, they do so in a way that suggests that they themselves have neither clearly identified the carriers of meaning nor adequately determined the delimitation of a reporter’s responsibility in transmitting them’. Although these words were written over 30 years ago, Walker’s discussion of courtroom transcripts could equally well apply to the current production of ROTI transcripts. Indeed, as Deamer et al. (2022) highlight, more work is needed to assist ROTI transcribers in producing transcripts which more accurately and reliably reflect the original spoken interaction. In their concluding remarks, Deamer et al. highlight the need for further research into how the representation of specific linguistic features in

transcripts can influence perceptions of interviewees. Exploring the relationship between person perception and linguistic representation within transcripts and audio recordings is the goal of the study presented in this article.

## Research aims

This article addresses whether perceptions of a suspect in a police interview vary according to the modality in which the interview is presented, and whether the presence and representation of pauses influences judgements made about the interviewee. We aim to assess whether the modality in which the interview is presented to participants would affect judgements that are made about the interviewee, and whether the findings in Deamer et al. (2022) could be replicated with a larger sample size and additional stimuli. Furthermore, given the relative importance that pauses have been shown to play in interactional discourse (Duez 1985; Wennerstrom and Siegel 2003; Kendall 2013; Deamer et al. 2022), another goal of the experiment is to assess whether the presence of pauses in the audio and the representation of pauses in transcripts would affect perceptions of the interviewee. The two key research questions are as follows:

1. Does the presentation of a police interview in audio or written format affect perceptions of an interviewee?
2. Does the presence/absence and mode of representation of pauses influence perceptions of an interviewee?

Taken together, these two research questions can help address the validity of changing the mode in which police interview evidence is presented to police officers, jurors, lawyers, judges and other stakeholders at various stages of the judicial process. This is especially important given the current treatment of police interview transcripts as ‘unproblematic copies’ of the corresponding audio recordings (Haworth 2018). We have taken an experimental approach as it is not possible to observe jury behaviour directly in the UK criminal context, and this is in line with a range of studies which have used experiments as a way of understanding human perception as applied to legal contexts (Carlson and Russo 2001; Dixon et al. 2002; Romero-Rivas et al. 2021).

## Method

The stimuli for the experiment comprised of a three-minute clip of an authentic police interview recording with a suspect who was being interviewed as a suspect in a murder enquiry. The police interview was conducted in the UK with a British English interviewee. The interviewee was a white British female speaker in her early 20s who spoke with a South West-area accent.<sup>1</sup> The recording was

available on YouTube (K.L.E.E. Photography 2015) and therefore in the public domain. It was the same recording used for the experiment in Deamer et al. (2022) to ensure comparable results. We extracted the audio from the original video as police interviews with suspects are more commonly audio-recorded than video-recorded. We selected the first three minutes to ensure participants had enough information upon which to base their answers, but so that the clip was short enough to hold participants' attention throughout the task. Additionally, the chosen section of the interview contained minimal amounts of identifying information from the case, and any personally identifying information such as names were removed from the experimental stimuli. The clip also contained a large amount of pausing and some emotional content such as sniffing and crying, which were found to be influential on people's perceptions in Deamer et al. (2022). Using stimuli which contained a high number of pauses was required to facilitate an appropriate test of the role that pauses play in people's perceptions of interviewee testimony. Our concern here is to investigate *silent* pauses as opposed to *filled* pauses such as 'uhm' and 'er'. We generated five versions of the interview clip for use in this experiment, as follows:

- **Audio 1** – The original, unedited version of the audio clip used in Deamer et al. (2022).
- **Audio 2** – An edited version of the audio recording in which all mid-sentence pauses longer than 200ms were removed, and pauses between sentences and speaker turns were reduced to 200ms in length. The 200ms threshold for what should be considered as a pause was taken from McDougall and Duckworth's (2017) definition under the Taxonomy of Fluency Features for Forensic Analysis (TOFFA) framework. The naturalness of the edited recording was discussed and agreed within the research team, but we did not conduct any other specific tests for naturalness with this version of the audio.
- **Transcript 1** – a transcript of the audio recording that contained no specific representation of pauses.
- **Transcript 2** – a transcript of the audio recording that showed pauses longer than 200ms in '(X.X sec)' format.
- **Transcript 3** – a transcript of the audio recording that showed pauses longer than 200ms in '...' format. Richardson et al. (2023, in press) found this representation to be a method of marking pauses in ROTI transcripts, where pauses are marked but not distinguished in length.

The transcripts were produced by the research team using the same principles as adhered to by Deamer et al. (2022). These were that the transcripts should be comprehensible by a lay audience and not contain too much technical information (of the kind which might be included in a transcript for Conversation Analy-

sis (see Jefferson 2004), but at the same time be accurate and detailed enough and to include information about some non-verbal aspects of speech. Each transcript line was numbered, with interviewer turns marked as 'IR' and interviewee turns marked as 'IE'. A transcription key was also included so participants were clear as to the meaning of the symbols used, although no pre-training was provided as part of the experiment. Representative examples of the three transcripts are shown in Figures 1–3

Audio editing was conducted using Audacity software. Audio files were exported in WAV format and embedded within SoundCloud for integration with the JISC Online Surveys experimental platform, which was used to collect and process the results. 250 participants (mean age 39; age range 19–77; 126 female) were recruited via Prolific (<https://www.prolific.co/>) to take part in a perception experiment, in which they were presented with one of the versions of the interview listed above and asked a series of questions about the interviewee. The experiment was conducted using the JISC Online Surveys platform (<https://www.onlinesurveys.ac.uk/>). Participants provided informed consent to take part in the research and received payment for their participation. All participants were United Kingdom (UK) nationals who also lived in the UK. This accurately reflected the requirements to sit on a UK jury panel. Participants were randomly and evenly split between the five experimental conditions (50 per condition). Participants were asked to either listen to the audio clip or read the transcript of the interview, depending on the condition, and assess how *relaxed*, *anxious*, *fearful*, *disgusted*, *surprised*, *happy*, *angry*, *sad*, *contemptuous*, *agitated*, *calm*, *panicked*, *friendly*, *cooperative*, *aggressive*, *defensive*, *assertive* and *nervous* the interviewee was. The use of these specific attributes follows the experimental approach

Line	Speaker	Content
1	IR:	How do I know that you weren't involved?
2	IE:	Again, I shouldn't have any DNA reason to be involved.
3		And again especially with my past. To think that I
4		could allow harm to come to somebody else like that
5		((sniff)) is highly unlikely. And again the fact that
6		as far as I knew he was y' know changing he wasn't as
7		violent anymore. And you know he was going to sort the
8		house out and talking about (blanked). You know. I'm
9		carrying his children, I didn't ((sniff)) to think
10		that I would know. Sorry.

Figure 1: Example from a transcript containing no pauses.

Line	Speaker	Content
1	IR:	How do I know that you (0.5 sec) weren't involved?
2	IE:	(2.3 sec). Again, I shouldn't have any (1.5 sec) DNA
3		reason to be involved. And again (1.7 sec) especially
4		(1.5 sec) with my past. (0.6 sec). To think that I
5		could (1 sec) allow (0.8 sec) harm to come to somebody
6		else like that ((sniff)) is highly unlikely. (0.9
7		sec). And again the fact that (0.5 sec) as far as I
8		knew (0.5 sec) he was y' know changing he wasn't as
9		violent anymore. (1 sec). And you know he was going to
10		sort the house out and talking about (blanked). You
11		know. I'm carrying his children, I didn't (4.5 sec)
12		((sniff))(1.1 sec) to think that (1.6 sec) I would
13		know (3.2 sec). Sorry.

Figure 2: Example from a transcript where pauses were represented in (X.X sec) format.

Line	Speaker	Content
1	IR:	How do I know that you ... weren't involved?
2	IE:	... Again, I shouldn't have any ... DNA reason to be
3		involved. And again ... especially ... with my past.
4		... To think that I could ... allow ... harm to come
5		to somebody else like that ((sniff)) is highly
6		unlikely. ... And again the fact that ... as far as
7		I knew ... he was y' know changing he wasn't as
8		violent anymore. ... And you know he was going to
9		sort the house out and talking about (blanked). You
10		know. I'm carrying his children, I didn't ...
11		((sniff)) ... to think that ... I would know ...
12		Sorry.

Figure 3: Example from a transcript where pauses were represented in ... format.

taken by Deamer et al. (2022), which was based on the emotion classification system developed by Ekman (1992). Also following the approach of Deamer et al. (2022), participants were also asked to state how *plausible* and *sincere* they thought the interviewee was. The ordering of questions was the same for all participants. These ratings were provided on a five-point Likert-type scale, where a

rating of 1 represented ‘not at all [TRAIT]’, and a rating of 5 represented ‘very much [TRAIT]’.

Additionally, participants were asked to say whether they thought the interviewee was telling the truth (‘yes’, ‘no’ and ‘don’t know’ were provided as options to answer this question). Participants were also asked to judge whether, from the information they were given, they felt the interviewee should be found *guilty* or *not guilty* of being complicit in the murder discussed in the interview.

Finally, for judgements of sincerity, plausibility, credibility and whether the interviewee was telling the truth, participants were instructed to state which aspects of the interviewee’s language use they felt influenced their judgements. These questions were designed to encourage participants to focus their attention on language use and provide more detailed commentary about their introspections on the way that language use shaped their judgements. These questions were free-text responses and participants were therefore able to give as much or as little information as they deemed relevant. The qualitative aspect to the experiment was designed to supplement the quantitative analysis and allow for individual responses to be given greater consideration alongside group-level trends. This was appropriate given the framing of the experiment in a jury context, where the opinions of individuals can be particularly important. Participants had unlimited access to the text or audio during the experiment and could read or listen as many times as they felt was necessary.

The use of the five different versions of the interview was designed to facilitate multiple comparisons to address the research questions outlined above. An assessment of perceptual differences between audio and transcript versions of the interview was conducted through a comparison of responses to the two audio versions of the data (Versions 1 and 2) against the three transcript versions of the data (Versions 3, 4 and 5). An assessment of whether the removal of pauses in the audio data would affect perceptions of the interviewee was facilitated through a comparison of the two contrasting audio versions (Versions 1 and 2). Finally, an assessment of whether both the presence and representation style of pauses in the transcript versions of the stimuli would affect perceptions of the interviewee was conducted by a comparison of responses to the three different transcript versions of the interview (Versions 3, 4 and 5).

## Results

### Perceptions of the interviewee in audio and transcript data

Our first research question was whether there would be differences in perceptions of the interviewee between the participants who heard the audio versions of the interview and those who read the transcripts. This aspect of the study was designed to assess whether the findings in Deamer et al (2022) would be rep-

licated with a larger sample size across an expanded study. An analysis of the distributions of participant responses in the audio and transcript conditions was undertaken, with the two audio conditions combined into a single ‘audio’ category and the three transcript conditions collapsed into a single ‘written’ category. This approach was taken to assess more directly whether the findings of Deamer et al. (2022) could be replicated with a larger sample. We later move to discussion and analysis of the differences between the two audio conditions and the three transcript conditions in order to ensure this is not overlooked and that our second research question is critically addressed. Statistical analysis of the numerically ranked traits (*relaxed, anxious, fearful, disgusted, surprised, happy, angry, sad, contemptuous, agitated, calm, panicked, friendly, cooperative, aggressive, defensive, assertive, nervous, plausible and sincere*) performed using Wilcoxon signed-rank tests. Statistical analysis of the categorically rated traits (*truth and guilt*) was conducted using chi-square tests. Data analysis was conducted using R software (R Core Team 2022).

Table 1 details the results of the statistical comparisons between participant ratings for the numerically ranked traits in the audio (Audio 1 and Audio 2 combined,  $n=100$ ) and transcript (Transcripts 1, 2 and 3 combined,  $n=150$ ) conditions.

Table 1: Statistical analysis of ratings traits between audio and written conditions

Trait	$W =$	$p$ -value
Credible	9815.5	<0.001***
Plausible	9519.5	<0.001***
Sincere	9283.5	<0.001***
Nervous	5389	<0.001***
Defensive	5220	<0.001***
Cooperative	10238	<0.001***
Panicked	5106	<0.001***
Calm	10048	<0.001***
Agitated	4744.5	<0.001***
Friendly	9160	0.002**
Sad	8964.5	0.007**
Surprised	6175.5	0.01*
Aggressive	6450	0.02*
Happy	8267.5	0.02*
Relaxed	8641.5	0.02*
Assertive	8652	0.03*
Anxious	6522.5	0.06
Contempt	6607.5	0.10
Disgusted	6757	0.17
Fearful	6845	0.21
Angry	6946	0.30

The results in Table 1 show significant differences between judgements of the interviewee in the audio and transcript versions of the interview in a majority of traits. There were significant differences between the two groups in how *credible*, *plausible*, *sincere*, *nervous*, *defensive*, *cooperative*, *panicked*, *calm*, *agitated*, *friendly*, *sad*, *surprised*, *aggressive*, *happy*, *relaxed* and *assertive* the interviewee was perceived to be. These findings support those of Deamer et al (2022), who also found differences across perceptions of a range of traits depending on whether participants were presented with a written or audio version of the interview. Figure 4 shows percentage distributions for each number on the rating scale (1, 2, 3, 4, 5) in order to illustrate the differences between responses to the audio and transcript stimuli for each trait yielding a significant difference. These charts are anchored around '3' on the scale as this was the middle value on the ratings scale, and each band shows the percentage of ratings provided for each value on the scale ('1' is represented in the darkest band to the left and '5' represented in the darkest to the right).

The plots in Figure 4 illustrate the differences between participant judgements in the audio and written conditions. Generally for this interview, the perception of the interviewee was more favourable from participants who had heard the audio compared to participants who read the transcript. The plots in Figure 4 show that, in the transcript condition, the interviewee was perceived to be comparably less *credible*, less *plausible*, less *sincere*, more *nervous*, more *defensive*, less *cooperative*, more *panicked*, less *calm*, more *agitated*, less *friendly*, less *sad*, more *surprised*, more *aggressive*, less *happy*, less *relaxed* and less *assertive*.

However, the plots in Figure 4 also highlight a potential issue in the data, where significant differences for certain traits were driven by differences between scores which represented one half of the ratings scale, such as in the judgements of *aggressiveness* and *happiness*. Therefore, while there were significant differences in the rating scores, the overall impression given by participants was that the interviewee was not *aggressive* and not *happy* in both the audio and transcript conditions. These traits contrast with judgements of traits like *defensiveness*, where the difference in ratings traverses the positive (scores of 4 or 5) and negative (scores of 1 or 2) parts of the scale. To further explore this, additional statistical analysis was undertaken to assess which traits showed differences between the positive and negative parts of the ratings scale. For each of the traits showing significantly different responses to audio and written representations in Table 1, ratings of 1 and 2 were collapsed together into a 'negative half' category, ratings of 3 were labelled as 'neutral' and ratings of 4 and 5 were collapsed into a 'positive half' category. Chi-square tests were conducted on the differences between the numbers of responses in the negative, neutral and positive sets in the audio and transcript modalities. This was designed to facilitate an assessment of whether

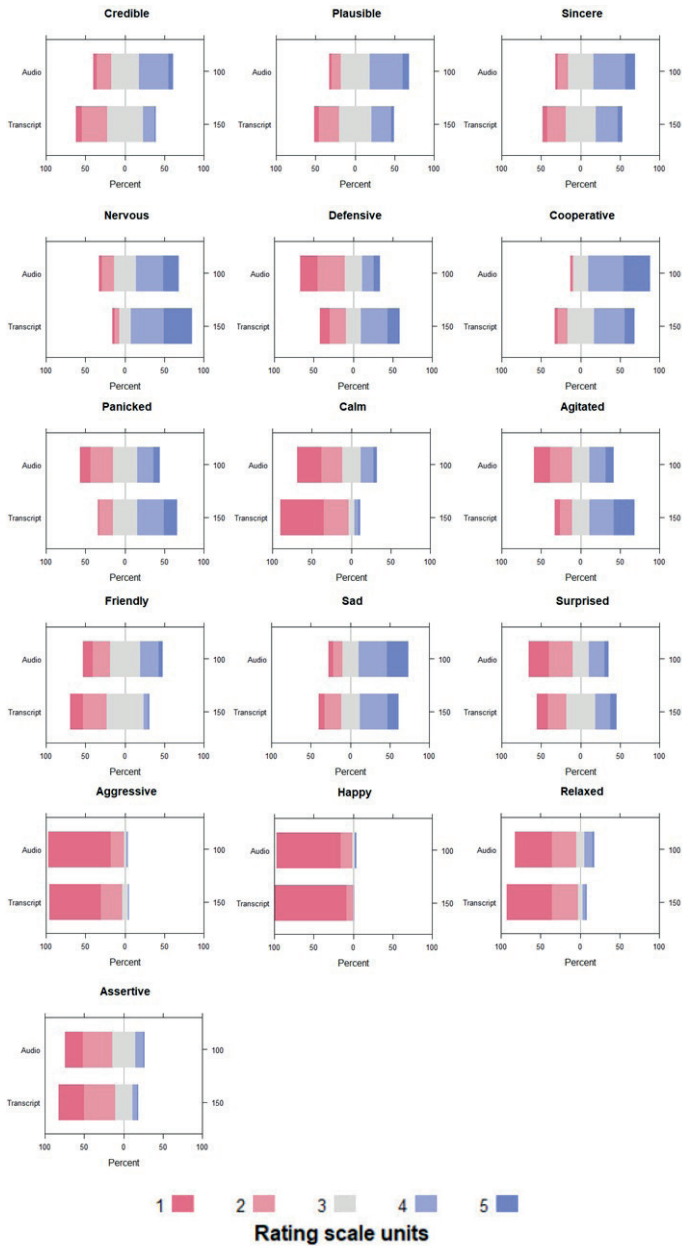


Figure 4: Percentages of numerical ratings assigned by participant for significantly different traits (1= not at all [TRAIT], 5 = very much [TRAIT]).

Table 2: Statistical analysis of ratings traits between audio and written conditions (positive, neutral and negative categories)

Trait	$\chi^2$	df	p-value
Calm	26.57	2	<0.001***
Credible	23.05	2	<0.001***
Agitated	22.61	2	<0.001***
Friendly	21.22	2	<0.001***
Cooperative	20.87	2	<0.001***
Defensive	19.21	2	<0.001***
Nervous	15.39	2	<0.001***
Plausible	14.96	2	<0.001***
Panicked	11.76	2	0.002**
Sincere	10.37	2	0.006**
Surprised	9.59	2	0.008**
Relaxed	7.76	2	0.02*
Sad	5.30	2	0.07
Assertive	4.14	2	0.13
Happy	3.01	2	0.22
Aggressive	2.27	2	0.32

the significant differences displayed in Table 2 also traversed the positive/negative boundary on the rating scale. The results of this analysis are displayed in Table 2.

The analysis in Table 2 shows that all traits aside from *sad*, *assertive*, *happy* and *aggressive* showed significant differences between participant ratings in the audio and transcript conditions when the differences between ratings of 1 and 2, and 4 and 5 were collapsed into single categories. This further shows that there were clear differences in participants' perceptions of these traits between the audio and transcript experimental conditions.

In addition to these numerically ranked traits, participants were also asked to state whether they thought the interviewee was telling the truth or not. Figure 5 shows the percentage number of responses to the question 'do you think the

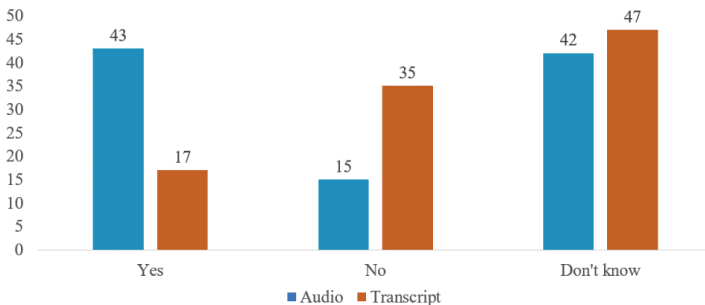


Figure 5: Participant responses to 'Do you think the interviewee is telling the truth?'

interviewee is telling the truth?' in each of the audio and transcript conditions. Chi-square significance testing showed there was a significant effect of modality (audio vs transcript) on participants' perceptions of whether the interviewee was telling the truth ( $\chi^2$  (df = 2,  $N = 250$ ) = 23.82,  $p < 0.001$ ). The results show that the interviewee was more likely to be perceived to be telling the truth when participants heard the audio compared with when they read the transcript, with a similar number of participants using the 'don't know' option in both modality conditions.

Participants in the study were also asked 'in your opinion, and based only on what you have heard/read, should the interviewee be found Guilty or Not Guilty of being complicit in the murder?'. Chi-square significance testing showed the effect of modality (audio vs transcript) on participants' perceptions of whether the interviewee should be found guilty fell short of statistical significance ( $\chi^2$  (df = 1,  $N = 250$ ) = 3.37,  $p = 0.07$ ). There were 16 participants across the experiment (4 in the audio condition and 12 in the transcript condition) who judged that the interviewee was not telling the truth, but also that they should be found not guilty of being complicit in the murder. This perhaps reflects the higher burden of proof required to give a guilty verdict compared with judging that somebody is not being truthful.

#### **Influence of pauses on perceptions of the interviewee**

Our second research question was whether the presence and/or representation of pauses in both the audio and transcripts would affect participants' perceptions of the interviewee. This expands on the suggestion made by Deamer et al. (2022) that future studies in this area should explicitly address the effect that the representation of specific linguistic features have on perceptions of interviewees. This was not addressed by Deamer et al. (2022) and therefore represents a significant addition to the previous study.

The experiment contained two different audio versions of the interview; the original version used in Deamer et al. (2022), and a second version which had all mid-sentence pauses longer than 200ms removed and between-sentence pauses reduced to 200ms. This facilitated an assessment of whether the presence or absence of pauses in the audio recording would affect perceptions of the interviewee. Statistical analysis of the numerically ranked traits was performed using Wilcoxon (audio) and Kruskal-Wallis (transcript) tests, and statistical analysis of the categorically rated traits was conducted using chi-square tests.

The analysis comparing the two audio conditions showed no significant differences between participants' perceptions of the interviewee for any of the traits, as displayed in Table 3. Additionally, there were no significant differences between the two audio conditions on perceptions of whether the interviewee was telling the truth ( $\chi^2$  (df = 2,  $N = 100$ ) = 2.07,  $p = 0.35$ ), or on judgements of whether the

Table 3: Statistical analysis of ratings traits between the two audio conditions

Trait	<i>W</i> =	<i>p</i> -value
Angry	1014	0.08
Fearful	1441	0.17
Plausible	1129.5	0.38
Panicked	1149.5	0.48
Sincere	1165.5	0.54
Calm	1326	0.59
Contempt	1323	0.60
Agitated	1319	0.63
Credible	1187.5	0.65
Surprised	1193	0.69
Happy	1209.5	0.69
Aggressive	1210	0.71
Sad	1301.5	0.71
Cooperative	1295.5	0.74
Defensive	1209.5	0.77
Nervous	1287	0.79
Assertive	1274	0.87
Disgusted	1227	0.87
Relaxed	1271	0.88
Friendly	1266	0.91
Anxious	1246.5	0.98

interviewee should be found guilty of being complicit in the murder ( $\chi^2$  (df = 2,  $N = 100$ ) = 0.208,  $p = 0.65$ ).

The experimental design also tested the effect that the representation of pauses in the transcripts would have on readers' perceptions of the interviewee. In the three versions of the transcript, pauses were either omitted completely (Transcript 1), represented with timings in '(0.2 sec)' format (Transcript 2), or represented as '...' (Transcript 3). Where applicable, a transcription key was provided to ensure that participants were aware of the meaning of the representations of the pauses within the transcripts. Table 4 details the results of the statistical comparisons between participant ratings for the numerically ranked traits in the three transcript conditions.

The results in Table 4 show significant differences across the three transcripts in perceptions of four traits: *contempt*, *aggression*, *assertiveness* and *defensiveness*. To determine the source of the significant effects, post-hoc pairwise comparison tests were conducted on the significant traits, with *p*-values adjusted using the Benjamini-Hochberg procedure. The output of this analysis is displayed in Table 5.

The results of this analysis shows that in each case, there was a significant difference between the transcript with no pauses (Transcript 1) and the transcript with

Table 4: Statistical analysis of ratings traits between the three transcript conditions

Trait	$KW \chi^2 =$	$p$ -value
Contempt	12.44	<b>0.002**</b>
Aggressive	6.79	<b>0.03*</b>
Assertive	6.90	<b>0.03*</b>
Defensive	6.22	<b>0.04*</b>
Anxious	4.65	0.09
Surprised	4.63	0.10
Agitated	3.55	0.17
Sad	3.27	0.19
Disgusted	3.24	0.20
Angry	3.20	0.20
Relaxed	0.79	0.47
Panicked	0.75	0.69
Calm	0.52	0.77
Happy	0.48	0.79
Sincere	0.37	0.83
Nervous	0.23	0.89
Plausible	0.15	0.92
Cooperative	0.14	0.93
Credible	0.09	0.95
Friendly	0.09	0.96
Fearful	0.08	0.96

Table 5: Post-hoc statistical analysis of ratings traits between the three transcript conditions

Trait	Transcript 1 ~ Transcript 2	Transcript 1 ~ Transcript 3	Transcript 2 ~ Transcript 3
Contempt	<b>0.003**</b>	0.06	<b>0.046*</b>
Aggressive	<b>0.01*</b>	0.12	0.31
Assertive	<b>0.009**</b>	0.22	0.16
Defensive	<b>0.01*</b>	0.4	0.12

timed pauses (Transcript 2). There was also a significant difference in *contempt* ratings between Transcript 2 and Transcript 3, although it is not clear why the change in pause representation would create a difference in this specific trait. To explore this further, Figure 6 shows percentage distributions for each number on the rating scale (1, 2, 3, 4, 5) to illustrate the differences between responses within the transcript conditions for these significant traits.

The plots in Figure 6 show the differences between ratings in the three transcript conditions for the significant effects in Table 3. Figure 6 shows that there was a general tendency for higher ratings to be assigned when the pauses were omitted from the transcripts. The trend in the data was for the interviewee to be rated as more *aggressive*, more *assertive* and more *contemptuous* when pauses

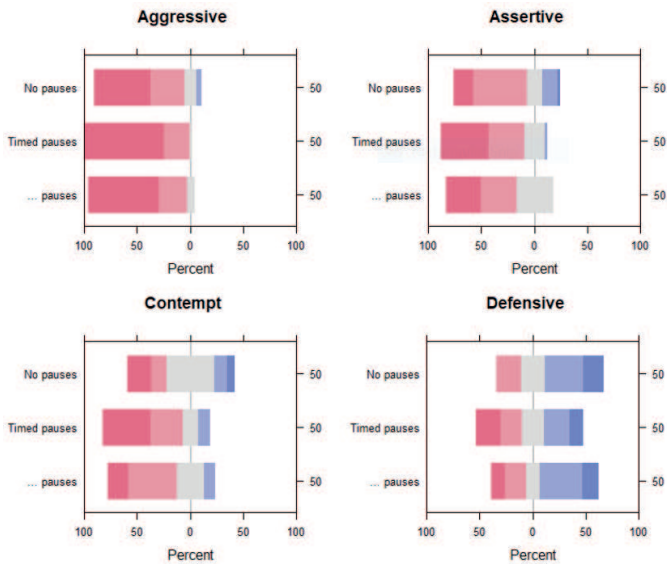


Figure 6: Percentages of numerical ratings assigned by participants for significantly different traits between the transcript conditions (1= not at all [TRAIT], 5 = very much [TRAIT]).

were omitted from the transcripts compared to when they were included. This pattern was less clear for *defensive* ratings, although there were no ratings of 1 (the lowest rating for defensiveness) in the condition where pauses were omitted, compared with the other two conditions.

In contrast to the differences between the audio and written conditions, there were no significant differences between the three transcript conditions on perceptions of whether the interviewee was telling the truth ( $\chi^2$  (df=1, N = 150) = 2.98,  $p = 0.56$ ). There was also no significant effect of pause representation on judgements of whether the interviewee should be found guilty of being complicit in the murder ( $\chi^2$  (df=1, N = 150) = 0.51,  $p = 0.77$ ).

One further question that can be addressed in this section is whether participants attended to the pausing within the audio and transcripts, or whether other features more strongly influenced judgements. One way to address this question is to examine the qualitative responses to the questions ‘What is it about the language used and/or how it is said that led you to your conclusion about how credible/plausible/sincere the interviewee is?’ and ‘What is it about the language used and/or how it is said that led you to your conclusion about whether or not what the interviewee is saying is true?’

Across the experiment, there were 1000 comments provided by participants in response to the four questions detailed above. Table 6 shows the number of mentions of either ‘pauses’ or ‘hesitation’ within the responses. The mention of

hesitation here only refers to responses which either mentioned this generally, or in the context of gaps or pauses in the speech. Mentions of vocalised hesitation markers such as ‘um’ or ‘er’ were not included. The responses are separated into three categories of pause mention. These are ‘no mention’, where the response did not mention pauses or hesitation, ‘absence mentioned’, where the participant actively mentioned the lack or absence of pauses or hesitation, and ‘presence mentioned’, where the participant actively mentioned the presence of pauses or hesitation.

The data in Table 6 show that the majority of responses to all four questions did not mention pauses or hesitation. Where pauses were mentioned by participants, they most frequently noted the presence, rather than the absence, of pausing or hesitation. There was also variability between the questions, with more noting of pauses when participants were asked what aspects of language influenced their judgement of whether the interviewee was telling the truth, compared to judgements of how credible, plausible and sincere the interviewee was. Table 7 shows the breakdown of the mentions of both the presence and absence of pauses in each condition in the experiment. The data in Table 7 are collated responses from all four questions.

The data in Table 7 show several key trends. Although the overall number of mentions of the absence of pauses were low, pause absence was most frequently noted by participants who heard the audio where the pauses had been removed. This would have been expected prior to the experiment. Unsurprisingly, in the transcript condition where no pauses were represented, no participant mentioned the absence of pauses. Instead, the opposite pattern was seen, with nine

Table 6: Frequencies of participants’ mentions of pauses and hesitation in qualitative responses

	Credibility	Plausibility	Sincerity	Truth
No mention	225	241	229	195
Absence mentioned	1	1	3	5
Presence mentioned	24	8	18	50
Total	250	250	250	250

Table 7: Frequencies of participants’ mentions of pauses and hesitation in each experimental condition

	Experimental condition				
	Original audio	Audio with pauses removed	Transcript without pauses	Transcript with timed (X.X sec) pauses	Transcript with ‘...’ pauses
Absence mentioned	2	8	0	0	0
Presence mentioned	7	1	9	43	38

participants mentioning pauses/hesitations despite there being no overt marking of these phenomena in the transcript.

The differences between the audio and transcript conditions are further emphasised when the mentions of pauses being present are analysed. The data in Table 7 show that pauses were far more likely to be mentioned as influencing perceptions of the interviewee by participants who read one of the transcripts which contained pauses, than by those who heard the audio which also contained the same pauses. This highlights an interesting and potentially problematic issue for transcript production, in that representing pauses in the transcripts brings the transcript closer to the original audio, but in doing so, this creates a perceptual difference between the two modalities.

The qualitative responses also show that different participants had opposing views about the meaning and influence of pauses in the context of whether the interviewee was *sincere*, *credible*, *plausible* and/or *telling the truth*. This is illustrated by the three examples below, all of which were in response to the question ‘What is it about the language used and/or how it is said that led you to your conclusion about whether or not what the interviewee is saying is true?’ and taken from participants who read a transcript of the interview (Transcript 3, ‘... pauses’).

- (1) Because of their disbelief and the pauses in their speech, to me this truly shows that they were caught in shock by the act committed by the person and were surprised that this could happen in such a short time like smoking a cigarette.
- (2) She is incoherent and a lot of unusual pauses which suggests she is thinking how to answer them to her benefit.
- (3) There were a lot of pauses and unfinished sentences. This could indicate lying or shock/distress.

The participant in (1), above, judged that the interviewee was telling the truth, the participant in (2) said that the interviewee was not telling the truth, and the participant in (3) said that they did not know whether the interviewee was telling the truth. Noticeably, in each example, the pauses were influential in these judgements. This illustrates that, for different people, the same representation of the same linguistic feature led the participants to arrive at three different conclusions about whether the interviewee was telling the truth or not.

The qualitative responses also highlight the wider judgements that people are willing to infer from transcripts. The examples in (4), (5), (6) and (7) below illustrate this.

- (4) It was the way he was saying it, there was no element of truth just based from the tone of the voice. (Transcript 1 – No pauses represented)

(5) The language sounds like the person is quite common / chavy [sic] and of lower than average intelligence, but f [sic] low cunning at the same time. (Transcript 1 – No pauses represented)

(6) IE doesn't appear to have the intelligence to be acting a role so convincingly. (Transcript 3 – '...' pauses)

(7) Also why would you ask permission to go round someone's for a fag? I'm assuming she meant a spliff or something like that but didn't want to divulge that information. (Transcript 1 – No pauses represented)

The participant providing the comment in (4) confused the sex of the interviewee, using 'he' rather than 'she', despite having been told that the interviewee was female in the instructions. Furthermore, they infer a 'tone of voice' from the transcript, despite never being exposed to any audio from the interview. Given that 'tone of voice' is an aspect of speech which is both linguistically unclear (see e.g. Tompkinson (2018) for discussion of why this can be a problematic term in speech analysis), and tied specifically to the way in which words are pronounced and other non-verbal aspects of speech, the inference of this from written words is both interesting and potentially problematic. The examples in both (5) and (6) show that judgements of social class and intelligence were inferred from the written transcripts. The participant in (5) describes the interviewee as 'common', 'chavvy' and deems them to have low intelligence, entirely from the transcript. The participant in (6) also judged the interviewee to have low intelligence. In example (7), the participant infers that the interviewee may have hidden information about taking illegal drugs, despite this never being mentioned in the interview.

Finally, the pattern found by Deamer et al. (2022) where some participants were more reluctant to make judgements without the audio or visual information was also seen in these data. Examples of this are shown in (8), (9) and (10), below.

(8) Im [sic] not too sure which is why I chose 'Dont [sic] know', you cant [sic] tell from a transcript whether someone is lying or not. More context is needed.

(9) It is hard to say whether they're being truthful especially since there are no audio and visual cues and it is difficult to know how the words are being delivered. They are left to the interpretation of whomever is writing the transcript.

(10) Listening and watching, rather than simply reading, might help me to form an opinion. It's difficult to ascertain body language etc.

It should also be noted that some of the patterns detailed above were also seen in the responses from participants who heard the audio recordings as opposed to reading a transcript. This is illustrated by the examples below.

(11) I think because it's a woman I'm more inclined to believe her.

(12) I think the West Country accent makes her sound more credible (i.e. less likely to be lying. But the content of the interview sounds a bit suspect.

(13) I don't have enough information to make a judgement as to whether she is telling the truth or not. If I had to guess I would say she is not, but that is just a hunch based on tone of voice rather than evaluating evidence.

(14) I think she is genuinely upset, but difficult to confirm this from audio, without seeing her body language.

In example (11), the participant's decision about whether the interviewee is lying was influenced by the fact that the interviewee was female. In (12), the participant uses the interviewee's accent as a factor in determining credibility, while tone of voice is used as a factor by the participant in (13), who states that they used this ahead of the evidence from the content of the interview. Finally, in (14) the participant mentions the absence of visual information as being a hinderance in the decision-making process.

## Discussion

The two core aims of the research presented in this article were to assess whether the presentation of a police interview in either audio or written format would affect perceptions of the interviewee, and to investigate whether manipulating the representation of pauses within both audio and transcripts would also affect perceptions of the interviewee.

The findings from this study strongly support those in Deamer et al. (2022) that the mode in which a police interview is presented can significantly affect perceptions of the interviewee. Significant differences were found across judgements of the interviewee between participants who heard an audio version of the interview compared with those who read the transcript. The interviewee in this study was judged as being significantly less *credible*, *plausible*, *sincere*, *cooperative*, *calm*, *friendly* and *relaxed*, as well as significantly more *agitated*, *nervous*, *surprised* and *panicked* by participants who read a transcript compared with those who heard an audio version of the interview. There was also a significant difference between the two modalities in judgements of how truthful the interviewee was, with the interviewee significantly more likely to be judged as not being truthful if the person making the judgement read a transcript as opposed to listening to the audio

recording. There is, of course, no expectation that the direction and degree of the differences found for judgements of this particular interviewee would hold for different interviewees from varying demographic backgrounds, in different contexts discussing different topics. However, the level of disparity between judgements based on audio and those based on transcript information in this study illustrates the perceptual instability that is introduced by changing the modality in which the interview is presented to someone tasked with evaluating the evidence.

In the context of ROTI transcripts being treated as ‘unproblematic copies’ of audio recordings, these results are alarming. Although the experimental stimuli were generated from a single police interview with one interviewee, the high degree of difference between judgements from participants in the two modality conditions suggests that the perceptual processing of audio and written information in this context are different from one another.

The second research question addressed in this study was whether manipulating the presence and/or representation of pauses in both audio and transcripts would affect participants’ judgements of the interviewee. This expanded the scope of the study beyond Deamer et al. (2022) and is an important consideration in the production of all types of transcripts, including ROTI transcripts, as it is important to know what impact the representation of linguistic features has on those tasked with evaluating information.

The results from this experiment showed no difference in participant judgements between those who heard the original version of the interview recording and those who heard the version where the pauses had been removed. There were some significant differences in responses to the three different versions of the transcripts, but these were markedly smaller in number than those found for the differences between the audio and written versions of the interview. The most marked differences were also between the version of the transcript which had no pauses and the transcript with pauses represented in a timed format. When the qualitative responses were analysed, they showed that participants in the audio condition paid relatively little attention to the presence of pausing, whereas those who were provided with transcripts which contained pauses listed them as an influential factor in their decision-making at a notably higher rate. The absence of pauses was hardly noted by participants in the audio condition where pauses were removed, and not noted at all in the transcript condition where pauses were not overtly represented. Furthermore, the qualitative findings illustrated an instability in perceptual judgements about pauses. Some people listed pauses as a factor which influenced a more favourable judgement of the interviewee, whereas others listed pauses as a factor which contributed to a more negative perception of the interviewee.

The results from this aspect of the study add further weight to the assertion made by Deamer et al. (2022) that the overt representation of linguistic features within transcripts can act as a perceptual prompt to the reader, encouraging them to pay more attention and place more judgemental weight on features which would have been comparably less noticeable in the equivalent audio. For example, it could be argued that the visual representation of pauses in transcripts makes the speech appear more disjointed than it sounds, and this is why the pauses were commented on more by those who read the transcripts than heard the audio. This creates tension between the perception and representation of linguistic features such as pauses within transcripts. By accurately detailing pausing information within transcripts, the transcriber brings the information in the transcript closer to that of the audio recording, and the transcript could therefore be viewed as a more accurate representation of that audio. However, the results of this experiment show that the act of doing this can create a perceptual difference between the audio and written versions of the interviewee. The findings suggest that people pay closer attention to the represented linguistic features in the transcript than they otherwise would in the audio, and these linguistic features can then be used to make consequential judgements about the interviewee. This problem is further compounded by the findings that the indicating of pauses in the transcripts has contrasting perceptual effects for different people, creating a further layer of divergence. Somewhat encouragingly, the results of this study found no significant differences between the two different ways of representing pauses (‘...’ and ‘(X.X sec)’) for all traits apart from *contempt* ratings, suggesting that *how* pauses are represented has less impact on readers of transcripts than *whether* they were represented.

The qualitative evaluations also showed that there were many other social traits that influenced participants’ perceptions of the interviewee. These included emotional traits, the speaker’s accent and gender, and the interviewee’s ‘tone of voice’. This was despite the fact that these features were not explicitly manipulated within the experimental design of the study, nor were they overtly represented in the transcripts. This illustrates two particular challenges in relation to speech perception and transcript production. The first challenge surrounds the complex and holistic nature of person perception, both from transcripts and audio information. Our results lend support to the idea that person perception is highly complex and multi-faceted, and crucially that this holds for the perception of individuals from written versions of interviews as well as from audio stimuli.

The results from this study lend support to the view that a transcript alone cannot accurately replicate the original audio recording without introducing the risk of creating differences in perceptions of the interviewee, no matter how ‘accurate’ the transcript is. This would directly support the view of Fraser (2022: 11), who

explicitly argues that it is impossible for a transcript to ever accurately replicate all of the information contained within the corresponding audio recording. This feeds into the idea that users of transcripts are tasked with ‘recontextualising’ the information within them (Komter 2019; Fraser 2022). Fraser (2022: 11) highlights the potential for transcripts to mislead readers if they are not used in conjunction with audio recordings, and therefore it could be argued that one way to navigate this in the context of ROTI transcripts would be a move towards playing the original audio in courtrooms, with the function of the ROTI transcript being to assist with comprehension. While this is almost certainly the safest approach for any transcript presented in court, a specific issue of practicality remains for ROTI transcripts, in that they are multifunctional and serve different purposes to forensic transcripts, and this in turn impacts on the how information is recorded and presented. Fraser (2020, 2022) calls for transcription to be treated as a dedicated branch of linguistics, and we would argue that the results of this study further illustrate how engaging with the perceptual effects that information in transcripts can have on listeners should be a key component of study in this area. Indeed, Fraser (2020) argues that a key problem with transcription is how transcripts are evaluated by lawyers and judges, which would link to the argument for greater consideration of the effects that reading a transcript can have on people.

Future experimental research in this area could test the effect of providing a combination of audio and written information on perceptions of interviewees, the relative perceptual weight of the two modalities, and whether or not participants are capable of ignoring one modality if instructed to do so (Mileva et al. 2018; Neil et al. 2021). Secondly, there is a further challenge around whether social information about speakers, such as accent, should or could be represented within a transcript. The identification and accurate description of speakers’ accents is a particular challenge given research which has highlighted variable performance in the ability of listeners without specialist training in linguistics and phonetics to identify accurately certain accents in recorded speech (Tompkinson and Watt 2018; Braber et al. 2022). It seems clear that accent is one attribute that has the potential to influence listeners’ judgements of speakers in audio recordings in legally relevant contexts, following research which links accent judgement to perceptions of traits such as guilt (Dixon et al. 2002; Dixon and Mahoney 2004). Regional accent is also one of the many attributes that is inherently lost when audio is converted to written text. It is also difficult to envisage how accurate accent information could be captured within transcripts without resorting to problematic ‘eye-dialect’ spellings or a phonetic notation system which requires specialist knowledge. Both of these options would almost certainly introduce more problems than they would solve. Future work in this area should also focus on a wider variety of interviewees and interview situations, which would allow

an assessment of the relative stability of the findings in this study, given that our focus here was only on one interviewee. We see this as an essential next step to broadening the scope of the research presented in this article. Finally, we would also argue that work in this area must extend beyond the vacuum of experimental research. Direct engagement with both the users and producers of ROTIs and other transcripts generated in legal contexts will be essential in ensuring a balanced, appropriate and linguistically sound solution to this problem. This is something we are actively working towards through the 'For the Record' project, an ongoing collaboration between the Aston Institute for Forensic Linguistics and a regional UK police force (Richardson et al. 2022).

## Conclusion

There are three key findings from this study. Firstly, the results should serve to urge severe caution over the treatment of audio recordings and transcripts of police interviews as 'unproblematic copies' of one another (Haworth 2018). The differences in participants' judgements between the audio and transcript versions of the interview show that the presentation of information in one modality cannot, and should not, be assumed to be perceptually equivalent to the other. Secondly, although the interviewee was generally perceived more favourably in the audio in this experiment, the directionality of any effects should not be assumed for all interviews. Instead, the key outcome from this study was that the presentation of information in different modalities created the potential for perceptual instability and differences in judgements. While different interviews, or interviewees, could produce different results in different directions, the presence of difference rather than the directionality is the important finding from this study. Finally, the results also lend support to the idea that the perceptual effects created by the representation of linguistic features should be an important consideration when assessing the form, function and overall accuracy of any transcript. Understanding if and how linguistic features such as pauses should be integrated into police interview transcripts is a key consideration moving forward, given that they have the potential to serve as evidence in courtrooms.

## Acknowledgements

This research was funded as part of an award to the Aston Institute for Forensic Linguistics by Research England's Expanding Excellence in England (E3) fund. We would like to thank the two anonymous reviewers and the editors of IJSL who provided insightful and thoughtful feedback which undoubtedly helped to improve the overall quality of this article.

### About the authors

Dr James Tompkinson is a Research Fellow in the Centre for Spoken Interaction in Legal Contexts (SILC) within the Aston Institute for Forensic Linguistics. His research focuses on topics which are relevant to both forensic phonetics and forensic linguistics. These include transcription, speaker comparison methods, regional variation in English dialects, perceptual judgements based on speech information, and spoken and written threats.

Dr Kate Haworth is the Director of the research centre for Spoken Interaction in Legal Contexts (SILC), part of Aston's Institute for Forensic Linguistics. Prior to her academic career she was a barrister and practised civil and criminal law. Her research interests include all aspects of language and communication in legal contexts, especially spoken interaction and the use of language data as evidence, focusing particularly on investigative interviews.

Dr Felicity Deamer is a Lecturer in the Aston Institute for Forensic Linguistics. Her current research uses language analysis to shed light on issues that emerge at the intersection between law and psychiatry. This work includes analysis of a wide range of linguistic data including evidential expert psychiatric reports submitted as evidence, police interviews with people experiencing serious mental illness and bodycam footage from secure mental health wards.

Dr Emma Richardson is a Lecturer in Language and Social Interaction at Loughborough University. Her research interests centre on improving access to criminal justice for 'vulnerable' (as defined by law) victims and witnesses of crimes such as kidnap, domestic violence and sexual offence. She uses conversation analysis to examine how these crimes are reported and progressed.

### Note

1. This is not an extensive profile of the speaker's voice but is instead based on a combination of holistic listening and knowledge of the case which was known to the researchers at the start of the experiment.

### References

- Allport, G. W. and Cantril, H. (1934) Judging personality from voice. *Journal of Social Psychology* 5(1): 37–55. <https://doi.org/10.1080/00224545.1934.9921582>
- Belin, P., Bestelmeyer, P. E., Latinus, M. and Watson, R. (2011) Understanding voice perception. *British Journal of Psychology* 102: 711–725. <https://doi.org/10.1111/j.2044-8295.2011.02041.x>
- Braber, N., Smith, H., Wright, D., Hardy, A. and Robson, J. (2022) Assessing the specificity and accuracy of accent judgments by lay listeners. *Language and Speech*, 00238309221101560. <https://doi.org/10.1177/00238309221101560>
- Carlson, K. A. and Russo, J. E. (2001) Biased interpretation of evidence by mock jurors. *Journal of Experimental Psychology: Applied* 7(2): 91. <https://doi.org/10.1037/1076-898x.7.2.91>

- Coupland, N. and Bishop, H. (2007) Ideologised values for British accents. *Journal of Sociolinguistics* 11(1): 74–93. <https://doi.org/10.1111/j.1467-9841.2007.00311.x>
- Deamer, F., Richardson, E. and Haworth, K. (2022) For the record: exploring variability in interpretations of police investigative interviews. *Language and Law/Linguagem e Direito*. [https://doi.org/10.21747/21833745/lanlaw/9\\_1a2](https://doi.org/10.21747/21833745/lanlaw/9_1a2)
- Dixon, J. A. and Mahoney, B. (2004). The effect of accent evaluation and evidence on a suspect's perceived guilt and criminality. *Journal of Social Psychology* 144(1): 63–73. <https://doi.org/10.3200/socp.144.1.63-73>
- Dixon, J. A., Mahoney, B. and Cocks, R. (2002) Accents of guilt? Effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology* 21(2): 162–168. <https://doi.org/10.1177/02627x02021002004>
- Duez, D. (1982) Silent and non-silent pauses in three speech styles. *Language and Speech* 25(1): 11–28. <https://doi.org/10.1177/002383098202500102>
- Duez, D. (1985) Perception of silent pauses in continuous speech. *Language and speech* 28(4): 377–389. <https://doi.org/10.1177/002383098502800403>
- Ekman, P. (1992) An argument for basic emotions. *Cognition and Emotion* 6: 169–200. <https://doi.org/10.1080/02699939208411068>
- Erickson, B., Lind, E. A., Johnson, B. C. and O'Barr, W. M. (1978) Speech style and impression formation in a court setting: the effects of 'powerful' and 'powerless' speech. *Journal of Experimental Social Psychology* 14(3): 266–279. [https://doi.org/10.1016/0022-1031\(78\)90015-x](https://doi.org/10.1016/0022-1031(78)90015-x)
- Fox Tree, J. E. (2002) Interpreting pauses and *ums* at turn exchanges. *Discourse Processes* 34(1): 37–55. [https://doi.org/10.1207/s15326950dp3401\\_2](https://doi.org/10.1207/s15326950dp3401_2)
- Fraser, H. (2020) Forensic transcription: the case for transcription as a dedicated branch of linguistic science. In M. Coulthard, A. May and R. Sousa-Silva (eds) *The Routledge Handbook of Forensic Linguistics* 416–431. Abingdon: Routledge.
- Fraser, H. (2022) A framework for deciding how to create and evaluate transcripts for forensic and other purposes. *Frontiers in Communication* 7: 898410. <https://doi.org/10.3389/fcomm.2022.898410>
- Giles, H. (1970) Evaluative reactions to accents. *Educational Review* 22: 211–227. <https://doi.org/10.1080/0013191700220301>
- Haworth, K. (2018) Tapes, transcripts and trials: the routine contamination of police interview evidence. *International Journal of Evidence and Proof* 22(4): 428–450. <https://doi.org/10.1177/1365712718798656>
- Hepburn, A. and Bolden, G. B. (2013) The conversation analytic approach to transcription. In J. Sidnell and T. Stivers (eds) *The Handbook of Conversation Analysis* 57–76. Malden, MA: Blackwell Publishing.
- Jefferson, G. (2004) Glossary of transcript symbols with an introduction. In G. H. Lerner (ed.) *Conversation Analysis: Studies from the First Generation* 13–31. Amsterdam: John Benjamins.
- K.L.E.E. Photography. (2015) Becky Watts murder: 3rd police interview (long version). <https://www.youtube.com/watch?v=hN4fYTXObcg>.

- Kendall, T. (2007) Listening to silence: interpretation and transcription of pause in deposition. In M. T. Turell, M. Spassova and J. Cicres (eds) *Proceedings of the Second European IAFL Conference on Forensic Linguistics / Language and the Law* 323–332. Barcelona, Spain: IULA, Documenta Universitaria.
- Kendall, T. (2013) *Speech Rate, Pause and Sociolinguistic Variation: Studies in Corpus Sociophonetics*. Dordrecht: Springer.
- Komter, M. (2019). *The suspect's statement: Talk and text in the criminal process* (Vol. 33). Cambridge: Cambridge University Press.
- Laver, J. (1994). *Principles of phonetics*. New York, NY: Cambridge University Press.
- Lundholm Fors, K. (2015) *Production and Perception of Pauses in Speech*. Doctoral dissertation, Department of Philosophy, Linguistics, and Theory of Science, University of Gothenburg.
- McAndrew, F. T. and De Jonge, C. R. (2011) Electronic person perception: what do we infer about people from the style of their e-mail messages? *Social Psychological and Personality Science* 2(4): 403–407. <https://doi.org/10.1177/1948550610393988>
- McDougall, K. and Duckworth, M. (2017) Profiling fluency: an analysis of individual variation in disfluencies in adult males. *Speech Communication* 95: 16–27. <https://doi.org/10.1016/j.specom.2017.10.001>
- Mileva, M., Tompkinson, J. A., Watt, D. and Burton, A. M. (2018) Audiovisual integration in social evaluation. *Journal of Experimental Psychology: Human Perception and Performance* 44(1): 128–138. <https://doi.org/10.1037/xhp0000439>
- Murphy, N. A. (2007) Appearing smart: the impression management of intelligence, person perception accuracy, and behavior in social interaction. *Personality and Social Psychology Bulletin* 33(3): 325–339. <https://doi.org/10.1177/0146167206294871>
- Neil, G. J., Higham, P. A. and Fox, S. (2021) Can you trust what you hear? Concurrent misinformation affects recall memory and judgments of guilt. *Journal of Experimental Psychology: General* 150(9): 1741. <https://doi.org/10.1037/xge0001023>
- Ohala, J. J. (1984) An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41(1): 1–16. <https://doi.org/10.1159/000261706>
- Oosterhof, N. N. and Todorov, A. (2008) The functional basis of face evaluation. *Proceedings of the National Academy of Sciences* 105(32): 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- R Core Team (2022) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Richardson, E., Hamann, M., Tompkinson, J., Haworth, K. and Deamer, F. (2023, in press) Understanding the role of transcription in evidential consistency of police interview records in England and Wales. *Language in Society*.
- Richardson, E., Haworth, K. and Deamer, F. (2022) For the record: questioning transcription processes in legal contexts. *Applied Linguistics* 43(4): 677–697. <https://doi.org/10.1093/applin/amac005>
- Romero-Rivas, C., Morgan, C. and Collier, T. (2022) Accentism on trial: categorization/stereotyping and implicit biases predict harsher sentences for foreign-accented defendants. *Journal of Language and Social Psychology* 41(2): 191–208. <https://doi.org/10.1177/0261927x211022785>

- Tompkinson, J. (2018) Assessing the influence of phonetic variation on the perception of spoken threats. PhD thesis, University of York.
- Tompkinson, J. and Watt, D. (2018) Assessing the abilities of phonetically untrained listeners to determine pitch and speaker accent in unfamiliar voices. *Language and Law/Linguagem e Direito* 5(1): 19–37.
- Tsantani, M. S., Belin, P., Paterson, H. M. and McAleer, P. (2016) Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception* 45(8): 946–963. <https://doi.org/10.1177/0301006616643675>
- Walker, A. G. (1986) Context, transcripts and appellate readers. *Justice Quarterly* 3(4): 409–427. <https://doi.org/10.1080/07418828600089041>
- Watt, D., Kelly, S. and Llamas, C. (2013) Inference of threat from neutrally-worded utterances in familiar and unfamiliar languages. *York Papers in Linguistics* 13: 99–120.
- Wennerstrom, A. and Siegel, A. F. (2003) Keeping the floor in multiparty conversations: intonation, syntax, and pause. *Discourse Processes* 36(2): 77–107. [https://doi.org/10.1207/s15326950dp3602\\_1](https://doi.org/10.1207/s15326950dp3602_1)
- Willis, J. and Todorov, A. (2006) First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science* 17(7): 592–598. <https://doi.org/10.1111/j.1467-9280.2006.01750.x>
- Zuckerman, M. and Driver, R. E. (1989) What sounds beautiful is good: the vocal attractiveness stereotype. *Journal of Nonverbal Behavior* 13(2): 67–82. <https://doi.org/10.1007/bf00990791>