Language and Law / Linguagem e Direito, *Vol. 11*(2), 2024, pp. 55-78
DOI: doi.org/10.21747/21833745/lanlaw/11_2a2

The effect of linguistic detail in police interview transcripts on perceptions of an interviewee

James Tompkinson [1] and Kate Haworth [2]

[1] *Department of Language and Linguistic Science, University of York*, james.tompkinson@york.ac.uk
[2] *Aston Institute for Forensic Linguistics, Aston University*, k.haworth@aston.ac.uk

Abstract
The question of what constitutes a "linguistically accurate" transcript is highly relevant for transcribers of police interview recordings. Does the inclusion of aspects of speech such as pauses, emphasis and laughter improve the quality of an interview transcript? Or does it make transcripts more difficult to interpret, and affect how interviewees are perceived? In this article, we explore whether the provision of a more linguistically detailed transcript made perceptions of an interviewee more aligned with the corresponding audio recording. Results showed mixed effects, and differences depending on the trait being perceived and the interview content. Our work suggests that the inclusion of additional linguistic detail in transcripts does not systematically make perceptions of interviewees more aligned with audio recordings, although there may be some scope for including the most salient features in transcripts. Our research illustrates the connected nature of speech perception and person perception in the context of evidential records of spoken interaction.

Resumo
A questão de saber o que constitui uma transcrição "linguisticamente exata" é altamente relevante para os transcritores de gravações de interrogatórios policiais. Será que a inclusão de traços do discurso oral como as pausas, a ênfase e o riso melhora a qualidade da transcrição de um interrogatório? Ou será que torna as transcrições mais difíceis de interpretar e afeta a perceção dos interrogados? Neste artigo, exploramos se o acesso a uma transcrição mais detalhada do ponto de vista linguístico torna as perceções de um entrevistado mais alinhadas com a gravação áudio correspondente.

Os resultados revelaram efeitos diversos e diferenças consoante o traço do discurso que estava a ser percecionado e o conteúdo da entrevista. No nosso trabalho verificamos que a inclusão de detalhes linguísticos adicionais nas transcrições não torna sistematicamente as perceções dos entrevistados mais alinhadas com as gravações áudio. Ainda assim, é possível haver alguma margem para a inclusão das caraterísticas mais salientes do discurso oral nas transcrições. A nossa investigação demonstra ainda a interligação da perceção do discurso e da perceção da pessoa no contexto de registos probatórios da interação oral.

1. Introduction

Transcripts are a fundamental and widely used tool within legal systems around the world, and the process of transcription within the legal context has received a great deal of attention from both within and beyond the field of linguistics. A range of transcription contexts have been examined by linguists, from transcripts of covert recordings produced by expert phoneticians (Fraser, 2022; French & Fraser, 2018) to courtroom transcripts (Eades, 1996) and transcripts of police interviews (Harrington, 2024; Haworth, 2018; Haworth, Tompkinson, Richardson, Deamer, & Hamann, 2023; Richardson, Hamann, Tompkinson, Haworth, & Deamer, 2023; Tompkinson, Haworth, Deamer, & Richardson, 2023). The wide-ranging use, variety of contexts, and applications of transcripts leads Fraser (2022) to call for transcription to be treated as a dedicated branch of linguistics. At the most basic level, irrespective of its function or format, a transcript is a written representation of speech. In the absence of an audio recording, transcripts are one method of creating a permanent record of an otherwise momentary event, although a large proportion of transcripts are produced from either an audio or video recording. This makes transcripts different from other records of spoken interaction, such as notes taken by a police officer which may then be converted into a summarised written statement. In this article, we aim to contribute research around the question of whether, and how, police interview transcripts should represent information about 'additional' aspects of speech such as pauses, overlapping talk, laughter and emphasis. We specifically build on the research presented in Deamer, Richardson, Basu, and Haworth (2022) and Tompkinson et al. (2023) to illustrate the effect of excluding or including this type of information on social trait judgements of an interviewee. We focus particularly on the situation in England and Wales, where police interviews with suspects are audio recorded and then a subsequent transcript is produced from the audio recording (Haworth, 2018), rather than other jurisdictions where the term 'transcript' may be used as a catch-all term for any record of spoken interaction. The research presented in this article aims to contribute to the wider debate around the best way to produce transcripts of police interviews.

2. Backgroud

2.1. The transcription of police interviews

In England and Wales, there is a legal mandate for all police interviews with suspects to be recorded, following the introduction of the Police and Criminal Evidence Act 1984. In addition to the original audio or video recording, it is common practice for Record of Taped Interview (ROTI) or Record of Video Interview (ROVI) transcripts to be produced, with these transcripts used throughout the case for both investigative and evidential purposes (Haworth, 2018). The process is different for interviews with witnesses (see Rock (2020) for a detailed description of how witness interviews are conducted in the UK), and in this article we will therefore concentrate solely on the process for suspect interviews. Our focus is primarily the English and Welsh legal jurisdiction, although the findings could be applicable to other countries and legal systems. We acknowledge, however, that different legal systems might have different processes in place for capturing and recording police interview evidence. For example, Fraser (2022) discusses the police interview transcription process in the Australian legal system, which has some similarities to and some differences from the process in England and Wales.

In England and Wales, ROTI transcripts are predominantly produced by staff employed within individual police forces. Fraser (2022, p. 10) states that police interview transcribers frequently have "contextual understanding of police and legal processes in general, and sometimes of specific cases". However, ROTI transcribers are not routinely provided with linguistic training, nor do they have advanced legal knowledge (Haworth, 2018). ROTI transcripts also frequently contain summarised information (Haworth, n.d.; Richardson et al., 2023), with some police forces displaying preference for verbatim transcripts and others producing transcripts with some speech captured verbatim and some speech summarised (Tompkinson et al., 2022)[1]. This could lead to a question around whether 'transcript' is an appropriate name for ROTI and ROVI documents, if a transcript is to be viewed as an attempt to represent speech in a verbatim manner. A legitimate future question for linguistic research in this area is whether there is a need to separate the terms used for completely verbatim transcripts and those produced with a mix of verbatim and summarised information, given that a summarised record contains the 'voice' of the transcriber. In this article, we continue to use the term 'transcript' as a catch-all description of these documents, which serve as written records of spoken interaction for evidential purposes, while also acknowledging the difference between the two main forms that ROTI and ROVI transcripts take.

Haworth (2018) identifies that transcripts of police interview recordings with suspects are frequently treated as being a direct copy of, or substitute for, the original audio. However, there have been repeated warnings from linguists in recent years that this is a flawed assumption (Haworth, 2018; Fraser, 2022; Tompkinson et al., 2023), and that speech and writing are fundamentally different mediums which cannot be treated interchangeably. Haworth (2018) asserts that the routine practice of taking audio-recorded police interviews, converting them into transcripts, and then having a legal official such

[1]Tompkinson, J., Haworth, K., & Richardson, E. (2022). *For the record: assessing force-level variation in the transcription of police-suspect interviews in England and Wales.* Conference of the International Investigative Interviewing Research Group, Winchester.

as a barrister read the transcript aloud in court, is a problematic process and a form of "routine contamination" of evidence in the legal system.

Haworth (2018) lists six recommendations for the improvement of police interview transcripts, which include the recommendation that the practice of reading transcripts aloud in courtrooms should be abandoned, and a recommendation that people who are required to assess evidence in courtrooms (such as juries or other triers of fact) should be exposed to the original audio recording rather than being left to solely rely on a transcript without the accompanying audio recording. The research presented in this article aims to further investigate another of the proposed recommendations set out by Haworth (2018); that there should be a standard code of practice introduced for the production of police interview transcripts. Haworth (2018, p. 446) states that "this should include a set of standard transcription conventions, to cover features such as overlaps, pauses, and any areas of uncertainty". Haworth (2018) further asserts that more research is required in this area, specifically with respect to understanding which features could be reliably included and represented within a transcript. However, at the time of writing, there is no standardised national guidance for how ROTI transcripts should be produced or formatted in England and Wales (Tompkinson et al., 2023).

The research presented in this article was conducted as part of a wider research project exploring issues around the production of ROTI transcripts. For the Record is a research project led by the Aston Institute for Forensic Linguistics; an overarching summary of the project can be found in Haworth et al. (2023), with the broad aim of the project described as applying "insights from linguistics to improve evidential consistency in police interview transcripts" (Haworth et al., 2023). In addition to research into how interview records are currently produced (Haworth, n.d.; Richardson et al., 2023), the project also questions how well ROTI transcripts serve their current purpose as both investigative and evidential records, and asks what linguistic analysis can offer to improve current transcription processes. It is this last question that the research presented in this article aims to address in relation to the provision of linguistic detail in a transcript.

2.2. Producing and standardising transcripts

When attempting to define what a transcript is, Harrington (2024) explains that as a minimum, a transcript should provide details of who is talking and a verbatim record of the speech. Fraser (2022) warns that although a transcript is a written document, separating the process of transcribing from the process of writing is critical. This issue is further explored by Leemann, Perkins, Buker, and Foulkes (2024, p. 130), who highlight the difference between 'real speech' and reported speech in writing, stating that "fluent speech does not resemble the syntactically perfect utterances you might read as reported speech in a novel or as lines in a script for film or theatre." Fraser (2022, p. 2) warns that no transcript can accurately capture every aspect of an audio recording of someone speaking. This links to Bucholtz's 2000 assertion that a transcript can never be neutral, with the process of transcription inevitably requiring the transcriber to make a series of choices about what to include, what to omit, and how to represent aspects of speech in written form. This point is further emphasised by Leemann et al. (2024, p. 129), who describe the process of transcription as a "subjective event". Linked to this, both Fraser (2022) and Richardson et al. (2023) state that there can never be such a thing

as the transcript, only a transcript. This is an important consideration and a necessary reflection on the limitations of the process of transcription, but it nevertheless poses a challenge for the production of consistent evidential transcripts by institutions such as police forces, where there is a clear desirability for a lack of inter-transcriber variation in the production of written records.

Harrington (2024, p. 22) provides a call for more standardisation in transcription practices, arguing that "standardisation is one step towards the ultimate goal of better, more accurate, and more impartial transcripts being presented to juries". However, problems such as what kind of standards transcribers should adhere to, how these standards should be set and achieved, and how a balance could be struck to avoid the risk of standards and guidelines becoming too restrictive and therefore hindering the process of transcript production, remain. This issue is raised by Haworth et al. (2023), who highlight the importance of creating a robust research base upon which any recommendations for standardisation can be made.

2.3. What should be included in a transcript?

The form that any transcript takes will be highly dependent on its function and purpose. For example, a transcript of an interaction between a medical practitioner and a patient differs drastically in form and function from a transcript of an indistinct piece of audio which is used as evidence as part of a criminal case. However, there are some issues which will be applicable to transcripts produced in a wide range of settings. In a discussion of how expert phoneticians produce transcripts of indistinct or difficult audio recordings, Leemann et al. (2024) highlight a range of problems that a transcriber might encounter, including how aspects of speech such as pausing, overlap, non-standard grammatical constructions and dialect terms are dealt with. They also crucially highlight the importance of considering the effect that the inclusion, omission, and representation of such features can have on a reader, questioning whether including representations of this linguistic detail could confuse linguistically untrained readers of transcripts (Leemann et al., 2024, p. 130). This issue has also been explored for police interview transcripts, with Richardson et al. (2023) pointing out that in an analysis of a small corpus of ROTI transcripts, there was inconsistency in the way that these kinds of features were represented. Richardson et al. (2023, p. 28) state that "police transcribers are left to use their personal judgement on when to include HOW, in addition to WHAT, was said", and highlight the issues that a reader might face if they are confronted with a transcript which contains unclear or inconsistent representations of aspects of speech. Summarising this issue, Gibbons (2003, p. 30) argues that a transcript must be able to be easily understood by a reader in order to successfully fulfil its primary purpose as a representation of spoken language.

Richardson et al. (2023) highlight the possibility that transcribers could be trained to more consistently represent aspects of speech such as pausing, overlap, and disfluency features in their transcripts, while simultaneously cautioning that this should not happen without explicit consideration of the effect that this could have on readers. This idea is commented on in detail by Fraser (2022), who outlines a range of cautions around the use of Conversation Analysis (CA)-style representations of speech features in legally relevant transcripts. These highlighted issues include difficulties in training transcribers to consistently use CA-style representations, and the possibility that the overt represen-

tation of such features could actually mislead readers and cause them to misinterpret information.

The research presented in this article provides some initial experimental testing to explore the various arguments that have been made in relation to the provision of linguistic detail in transcripts of police interview recordings. In doing so, the work builds directly on the previous research presented in Deamer et al. (2022) and Tompkinson et al. (2023). Using stimuli from a single police interview, these studies showed that there were differences between perceptions of the interviewee depending on whether the interview was presented as a written transcript or as an audio recording.

Additionally, the study in Tompkinson et al. (2023) showed that the overt representation of a specific linguistic feature, silent pauses, resulted in an increased degree of perceptual instability on the part of experimental participants. Perhaps most important was the observation that although accurately marking pauses in transcripts theoretically brought the transcript closer to the original audio, this overt marking created some additional perceptual differences between the two modalities. This creates a specific problem for those tasked with producing police interview transcripts. One the one hand, the inclusion of additional linguistic detail, such as marking pauses, emotion, overlap and emphasis, could be seen as making a transcript more 'accurate', as it reduces the difference between the transcript and the original audio recording through the overt representation of linguistic features. On the other hand, if there is the possibility that the inclusion of such detail will enhance the perceptual distance between audio recordings and corresponding transcripts, then it could be argued that this kind of information should not be included. To more comprehensively address this issue, this article aims to extend the scope of the studies presented in both Deamer et al. (2022) and Tompkinson et al. (2023) and specifically examine how the inclusion and exclusion of additional linguistic features within transcripts can affect how listeners perceive an interviewee.

In their experiment, Deamer et al. (2022) opted to include representations of pausing, emphasis, overlapping speech and emotion markers within the transcript stimulus. The logic for doing so was to compare what the authors describe as a "best possible transcript" (Deamer et al., 2022, p. 29) to the audio recording. However, the study did not then compare these linguistically detailed transcripts with an orthographic transcript which may more closely represent the kinds of police interview transcripts produced by ROTI transcribers. This could also have affected the results within the Deamer et al. (2022) study. For example, the results showed no significant difference between the audio and transcript conditions in perceptions of *sadness*, but it could be assumed that this was because both the audio and transcript contained the relevant emotional markers which would indicate sadness. Indeed, in their call for a follow-up study, Deamer et al. (2022) identified the need to include a plain or orthographic transcript condition to broaden the scope of their findings. This three-way comparison between audio recordings, orthographic transcripts and transcripts containing some additional level of linguistic detail was addressed in Tompkinson et al. (2023), but only in relation to silent pauses. This again limited the scope of the previous research on this topic as other linguistic features were not considered.

Another limiting factor to the research presented in Deamer et al. (2022) and Tompkinson et al. (2023) was that both studies used a single clip from one publicly available

police interview recording. While this has the benefit of providing the high level of control required for exploratory research, it means that the findings are somewhat limited in scope. On this issue, Tompkinson et al. (2023, p. 47) highlight that "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". The two studies also do not assess whether there is inter-speaker consistency in perceptual judgements of police interview interactions. This is an important consideration which we address in this article, by exploring the relative stability of judgements of an interviewee in two separate audio clips and corresponding transcripts. By doing this, we aim to assess the relative contributions of speech and speaker to judgements of an interviewee.

2.3.1. Research Questions

As previously outlined, the research in this article specifically addresses the highlighted limitations of the work presented in Deamer et al. (2022) and Tompkinson et al. (2023). In this study, we aim to provide more detail in relation to the following questions:

1. What are the differences between orthographic transcripts, linguistically detailed transcripts and audio recordings of police interviewees in relation to social trait perception?
2. To what extent do differences in social trait perception between transcripts and audio recordings hold constant for different sections of a single interview recording?

2.3.2. Methodology

The stimuli for this study were taken from the West Yorkshire Regional English Database (WYRED), which contains mock police interview recordings as one of the speech elicitation tasks. We opted to use simulated police interview recordings from a publicly accessible research database for this article given the lack of available authentic police interview data for experimental work, to minimise the ethical challenges that are involved with working with real police interview data, and to increase overall researcher control of the data. There are, of course, limitations to working with mock police interview recordings which need to be acknowledged here. The main limitation is that the interactions do not represent the high-stakes and potentially high-stress environment of a real police interview. It is difficult to know the exact degree to which this was a problem for our study, but we considered it a worthwhile trade-off for the benefits that using mock recordings provided.

The full design and composition of the WYRED database is set out by Gold, Ross, and Earnshaw (2018). In the mock police interview task, participants were asked a series of questions about a fictitious crime, with a map used as an aid to elicit certain responses (Gold et al., 2018, p. 2749). The interview we selected consisted of an interaction between a male interviewee and a real life female police officer. We extracted two sections of speech from different parts of the interview. For Experiment 1, we used a sample of speech which was 2 minutes and 26 seconds in length, and for Experiment 2, we extracted a sample which was 1 minute and 59 seconds in length. We considered this to provide enough information upon which participants could evaluate the interviewee, but also be sufficiently short to minimise the risk of participants losing focus during the

experiment. One limitation of this approach is that listeners did not get the full context that would have been provided by a longer extract, but we judged that providing participants with extended audio clips could compromise the experimental design and the level of participant engagement with the study. For each section of audio, we created two versions of the transcript of the interview interaction. The first version was a plain orthographic transcript which contained only the words that were spoken and basic punctuation markers. The second was a more linguistically detailed transcript which specifically marked *pauses, emphasis, overlapping speech, whispered speech* and *laughter*. We aimed to mark these as accurately as possible with respect to the corresponding audio recording, using symbols which were drawn from both Conversation Analysis and previous analysis of representations of these aspects of speech in ROTI transcripts (see Richardson, Haworth, & Deamer, 2022 for a more detailed discussion of these issues). A transcription key was also provided at the top of each transcript so that readers knew what the various symbols represented. We also opted to represent these features in the speech of both the interviewer and the interviewee to avoid creating differences between representations of the two speakers. Again, however, we acknowledge this was another choice that may have affected overall perceptions of the interviewee in our experiment. This issue also relates to a broader point, linking back to Bucholtz's (2000) cautions about transcription choices. It is important to acknowledge that the style, format and choices that we made about representations could differ between transcribers, meaning our version is only one possible representation of the audio recordings. Table 1 shows the representations of the different features and the number of times each feature was marked in the transcripts used for Experiment 1 and Experiment 2.

Feature	Representation	Number of occurrences in Experiment 1 transcript	Number of occurrences in Experiment 2 transcript
Pause longer than 0.5 seconds	**(X.X sec)**	36	23
Pause shorter than 0.5 seconds	**(.)**	6	7
Emphasis	**Underlined speech**	18	18
Overlapping speech	**[Words spoken in overlap]**	6	8
Whispered speech	**((whisper))**	0	1
Laughter	**((laughs))**	1	1

Table 1. Representations of features in the detailed transcripts, and the number of occurrences of each feature in the transcripts used for this experiment

To further exemplify the differences between the respective transcripts, examples of the contrasting versions (plain orthographic and linguistically detailed) are displayed in

Figures 1 and 2. Figure 1 shows a section from a plain orthographic transcript, while Figure 2 shows the corresponding detailed transcript.

```
36      IR:   Okay. Could you explain then? We've got CCTV footage of
37            your car with you driving and it looks like a person is in
38            the passenger seat. Would you like to think again if- what
39            that could possibly have been? Or who?

40      IE:   When was this?

41      IR:   On Thursday evening leaving work.

42      IE:   Leaving work. I'm not too sure, I think I might have
43            stopped off or dropped someone off from work.

44      IR:   Okay so maybe a short journey with someone in the car?

45      IE:   Yeah yeah.
```

Figure 1. Example of a section of a plain orthographic transcript

```
49      IR:   Okay. Could you explain then. We've got CCTV footage of
50            your car (.) with you driving and it looks like a person
51            in the passenger seat. (1.25 sec). Would you like to think
52            again if- (0.5 sec) what that could possibly have been?
53            (0.8 sec) Or who?

54            (2.9 sec)

55      IE:   When was this?

56      IR:   On Thursday evening leaving work.

57      IE:   ((whisper)) Leaving work. (7.8 sec). I'm not too sure. I
58            think I might have stopped off or dropped someone off from
59            work.

60      IR:   Okay. So maybe a short journey [with someone in the car]?

61      IE:   [Yeah yeah].
```

Figure 2. Example of a section of a linguistically detailed transcript

In each experiment, we conducted a perceptual judgement task similar to the designs adopted by Deamer et al. (2022) and Tompkinson et al. (2023). A total of 300 participants were recruited via Prolific (https://www.prolific.com/) to take part in the experiments. 150 participants took part in Experiment 1, and 150 participants took part in Experiment 2. In each experiment, participants were provided with either the audio, the plain orthographic transcript, or the linguistically detailed transcript of the interview clip, and then asked a series of questions about the interviewee. Participants were

UK residents and UK nationals, between the ages of 18 and 70, and equally split between male and female raters. Participants were instructed to either listen to the audio or read the transcript they had been provided with, and provide a judgement of how *sincere, plausible, credible, relaxed, anxious, fearful, disgusted, surprised, happy, angry, sad, contemptuous, agitated, calm, panicked, friendly, cooperative, aggressive, defensive, assertive* and *nervous* the interviewee was. These were the same descriptors used by Deamer et al. (2022) and are based on Ekman's 1992 universal emotion categorisation system. These ratings were provided using a 1-5 Likert-style scale, with 1 representing "not at all…" and 5 representing "very…". Participants were also asked to state whether they thought the interviewee was telling the truth, with *"yes"*, *"no"* and *"don't'* know" possible outcomes. No participant was permitted to assess multiple conditions in the experiments. All data was collected using the JISC Online Surveys experimental platform. The study received Aston University ethical approval, and participants were paid for their participation in the research study at the standard approved Prolific rate.

2.3.3. Results

Experiment 1

Statistical analysis was performed using R software (R Core Team, 2024). Main effect p- values for the numerically judged traits in the experiment were calculated using Kruskal- Wallis significance testing. Post-hoc analysis was conducted using Benjamini-Hochberg adjusted Dunn tests, in order to determine the source and direction of differences within the data.

Statistical analysis of the responses to the question *"do you think the interviewee is telling the truth"* was conducted using a chi-square test. The output of this testing showed that there was no significant main effect of condition (audio, plain orthographic transcript, linguistically detailed transcript) on judgements of whether the interviewee was being truthful (x^2 = 7.50, df = 4, p = 0.11). However, post-hoc analysis of the comparisons between each of the conditions did show a significant difference between responses in the audio condition and those in the linguistically detailed transcript condition (x^2 = 7.17, df = 2, p = 0.03). The number of responses in each category for the question about whether the interviewee was telling the truth is shown in Figure 3, below. The data illustrates that although the number of respondents who judged the interviewee to be telling the truth was low in all conditions, there was a more notable difference between responses of 'no' and 'don't know'. Twice as many participants in the audio condition (n=21) answered 'don't know' when compared to participants in the linguistically detailed transcript condition (n=10). This suggests that the most negative judgements of whether the interviewee was telling the truth came from participants who read the linguistically detailed version of the transcript.

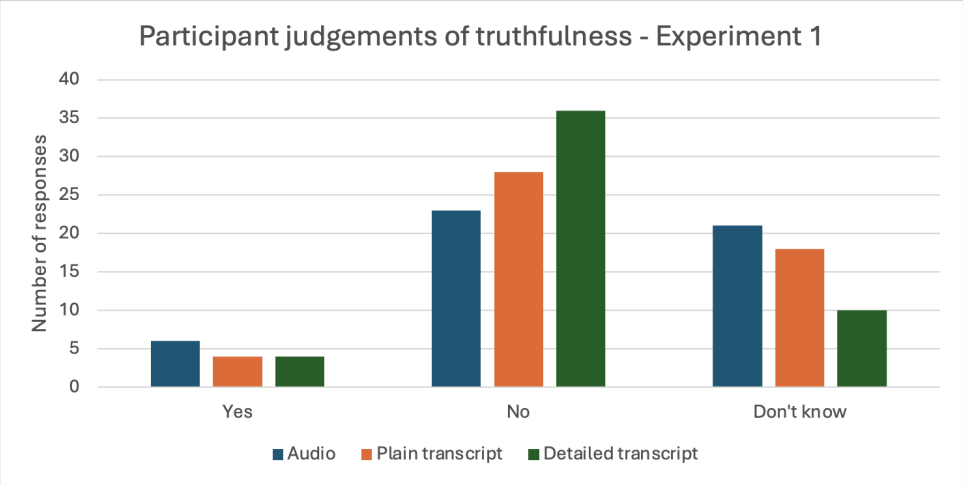

Figure 3. Participant responses to the question "do you think the interviewee is telling the truth?" in Experiment 1

Table 2, below, details the results of the main effect significance testing for the numerically rated traits in Experiment 1.

Trait	*KW χ2 =*	*df=*	*p=*
Angry	14.4	2	**<0.001*** **
Fearful	14.3	2	**<0.001*** **
Panicked	13.49	2	**0.001****
Plausible	13.45	2	**0.001****
Sincere	12.81	2	**0.002****
Anxious	12.43	2	**0.002****
Nervous	11.79	2	**0.003****
Sad	10.32	2	**0.006****
Credible	9.43	2	**0.009****
Relaxed	8.82	2	**0.01***
Calm	7.35	2	**0.03***
Aggressive	6.46	2	**0.04***
Agitated	6.06	2	**0.048***
Surprised	5.92	2	0.052
Cooperative	5.68	2	0.058
Contempt	4.95	2	0.08
Friendly	3.84	2	0.15
Disgusted	3.72	2	0.16
Defensive	2.22	2	0.33
Happy	0.97	2	0.62
Assertive	0.31	2	0.86

Table 2. Main-effect significance testing for Experiment 1

The results in Table 2 show that for all traits apart from *surprised, cooperative, contempt, friendly, disgusted, defensive, happy* and *assertive*, there was a significant effect of modality on participants' perceptions of the interviewee. However, this analysis does not show where differences in the data occur. Table 3, below, shows the p-values from the post-hoc pairwise comparison testing and illustrates where significant differences

existed between each of the modality conditions in the experiment which showed a significant effect in Table 2.

Trait	*Audio ~ Orthographic transcript*	*Audio ~ Detailed transcript*	*Orthographic ~ Detailed transcript*
Angry	**0.006****	**<0.001*****	0.15
Fearful	0.25	**<0.001*****	**0.003****
Panicked	0.25	**<0.001*****	**0.004****
Plausible	0.14	**0.01***	**<0.001*****
Sincere	0.46	**0.002****	**0.003****
Anxious	0.47	**0.003****	**0.002****
Nervous	0.47	**0.004****	**0.003****
Sad	0.35	**0.008****	**0.005****
Credible	0.38	**0.008****	**0.01***
Relaxed	0.43	**0.01***	**0.01***
Calm	0.15	**0.01***	0.07
Aggressive	0.09	**0.02***	0.17
Agitated	0.30	**0.03***	0.05

Table 3. Post-hoc testing of significant effects in Experiment 1

The data in Table 3 shows that there was a significant difference between perceptions of the interviewee in the audio and detailed transcript condition for every trait which had a significant effect in Table 2. This shows that for this interview section, providing more linguistic detail in the transcript did not result in perceptions of the interviewee being more closely aligned with the corresponding audio recording. For perceptions of how *angry* the interviewee was, there was no significant difference between the two versions of the transcript, but a significant difference between both versions of the transcript and the audio recording. This result aligns with the conclusion reached by Deamer et al. (2022) that a difference in modality (written or spoken) can create differences in perceptual judgements of an interviewee. This difference is illustrated in Figure 4, which shows the perceptual judgements of anger in all three conditions and the distribution of ratings provided by participants for each experimental condition. Although most participants rated the interviewee as not being particularly angry in all three conditions, there were far more ratings of 1 (the lowest possible rating) from participants who heard the audio recording compared with those who read either transcript.

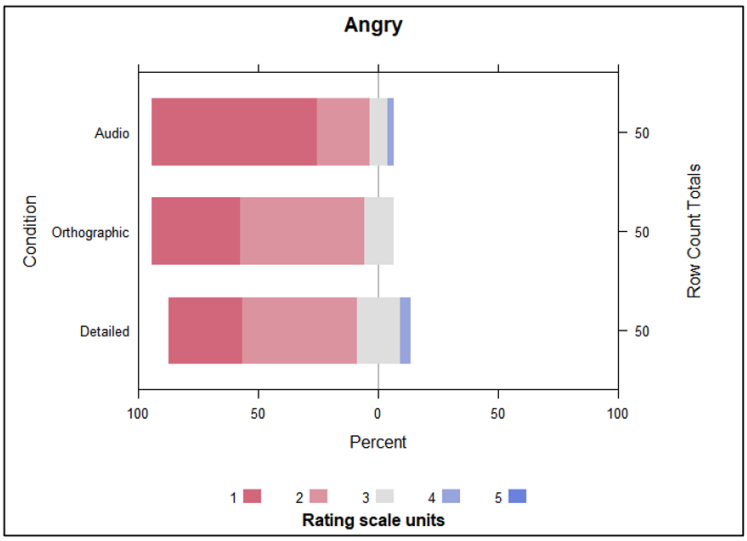

Figure 4. Listener judgements of how angry the interviewee was in each condition

Perhaps more notably, for judgements of how *fearful, panicked, plausible, sincere, anxious, nervous, sad, credible* and *relaxed* the interviewee was, the provision of linguistic detail in the transcript made perceptions of the interviewee significantly different to both the audio and plain orthographic transcript, but there was no significant difference between the audio condition and orthographic transcript condition. This means that perceptions of the interviewee were aligned in the plain transcript and audio conditions, but the provision of linguistic detail created more disparate judgements. These effects are illustrated in Figure 5, below.

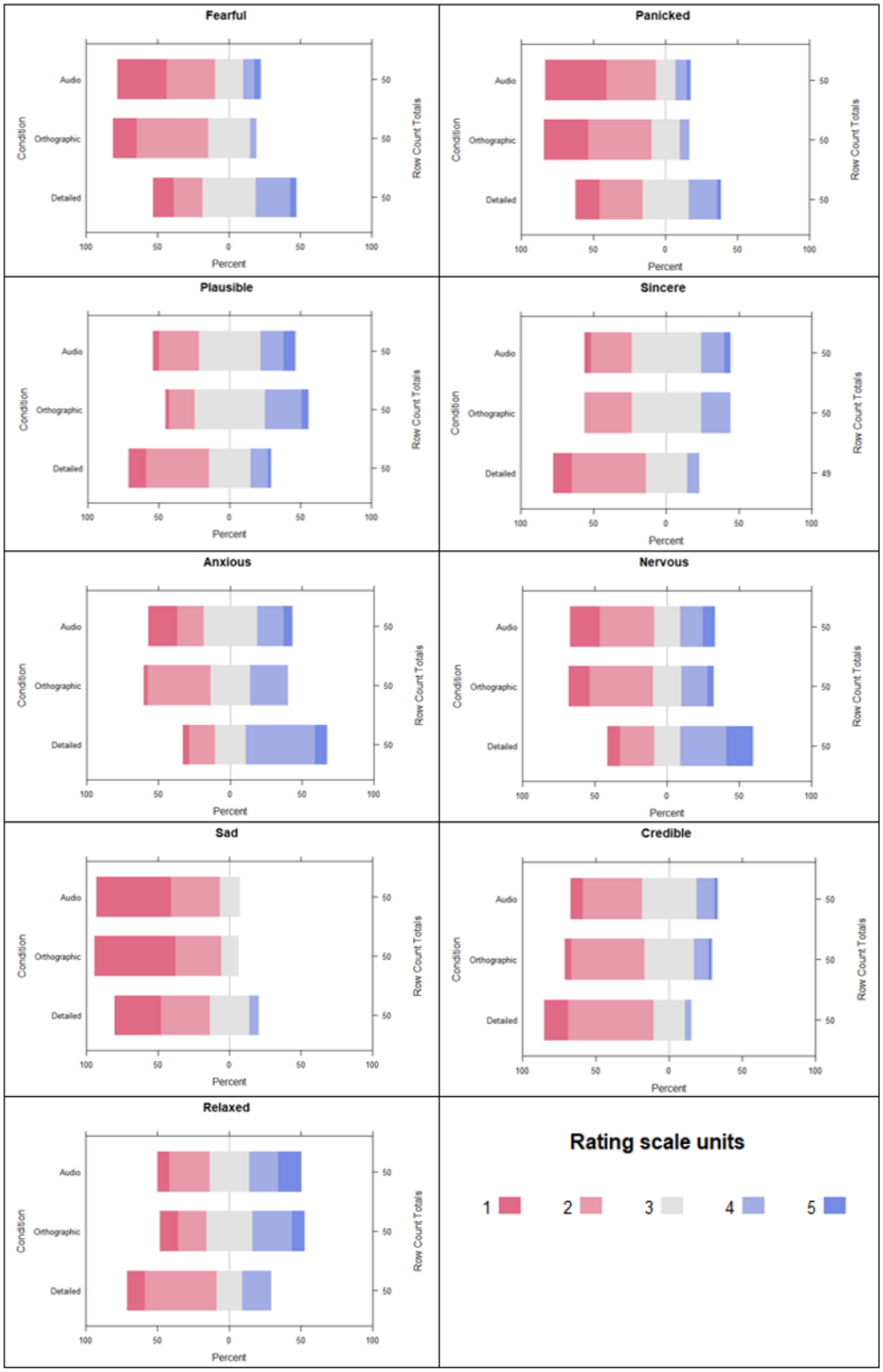

Figure 5. Listener judgements of how *fearful*, *panicked*, *plausible*, *sincere*, *anxious*, *nervous*, *sad*, *credible* and *relaxed* the interviewee was in each condition

Finally, for ratings of how *calm, aggressive* and *agitated* the interviewee was, the results showed a significant difference between the detailed transcript and audio conditions, but no other significant differences. Again, this illustrates that the provision of linguistic detail in the transcript created perceptual divergence from the audio rather than the alignment that might have been predicted if it is assumed that the provision of linguistic detail brings a transcript 'closer' to the corresponding audio recording. These effects are shown in Figure 6.

Another consistent effect across the judgement data for Experiment 1 was in the direction of the differences between the three conditions. For every significant trait shown in Figures 4-6, the inclusion of linguistic detail in the transcript created a more negative perception of the interviewee compared to the audio recording. The interviewee was judged to be more *angry, fearful, panicked, anxious, nervous, sad, aggressive* and *agitated* in the detailed transcript condition, and less *calm, relaxed, plausible* and *sincere*. This illustrates that for this interviewee in this section of the interview, the provision of linguistic detail not only created perceptual differences between the transcript and the audio recording, but that these differences were disadvantageous to the interviewee.

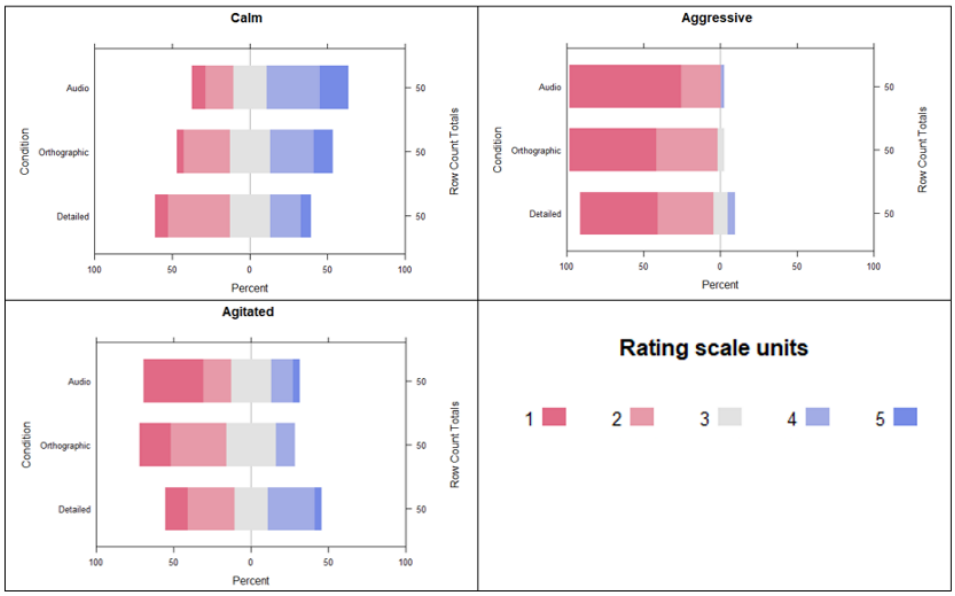

Figure 6. Listener judgements of how *calm, aggressive* and *agitated* the interviewee was in each condition

Experiment 2

The research design in Experiment 2 mirrored that of Experiment 1 but used a different section of the same interview. This experiment was designed to test whether the findings from Experiment 1 would be repeated, or whether a different section of the same interview involving the same speakers would produce different results. The same experimental design was used for both experiments, with the same statistical testing procedure used for data analysis.

As in Experiment 1, there was no significant main effect of condition on judgements of interviewee truthfulness for Experiment 2 (x^2 = 2.213, df = 4, p = 0.70). The number of responses in each category for assessments of truthfulness is shown in Figure 7, below.

The data in Figure 4 highlights that across all conditions, participants predominantly judged the interview not to be telling the truth, with minimal differences between responses in the individual conditions. Post-hoc analysis showed no significant differences between any of the individual conditions for this question in Experiment 2.

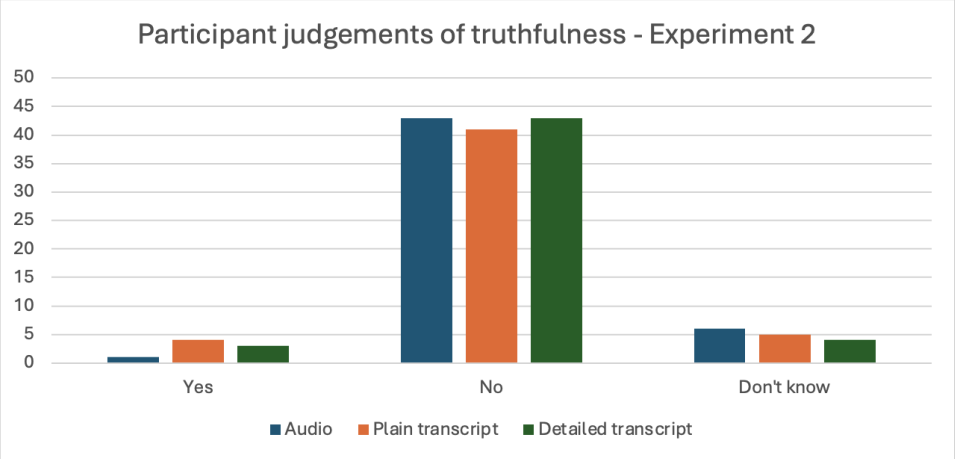

Figure 7. Participant responses to the question "do you think the interviewee is telling the truth?" in Experiment 2

Table 4, below, details the results of the main effect significance testing for the numerically rated traits in Experiment 2.

Trait	*KW χ2 =*	*df=*	*p=*
Aggressive	24.11	2	**<0.001*** **
Calm	15.47	2	**<0.001*** **
Agitated	13.09	2	**0.001** **
Defensive	11.86	2	**0.003** **
Angry	7.75	2	**0.02***
Relaxed	7.35	2	**0.03***
Disgusted	6.94	2	**0.03***
Surprised	4.72	2	0.09
Cooperative	4.14	2	0.12
Anxious	4.06	2	0.13
Fearful	3.73	2	0.16
Happy	3.58	2	0.17
Plausible	3.16	2	0.21
Contempt	3.02	2	0.22
Panicked	2.72	2	0.26
Assertive	2.31	2	0.31
Credible	2.20	2	0.33
Sad	1.68	2	0.43
Sincere	1.60	2	0.45
Nervous	1.24	2	0.53
Friendly	1.07	2	0.58

Table 4. Main-effect significance testing for Experiment 2

The results in Table 4 show that there were comparably fewer significant differences between response conditions in Experiment 2 compared Experiment 1. There was a

significant effect of condition on judgements of how *aggressive, calm, agitated, defensive, angry* and *relaxed* the interviewee was. All other traits showed no significant effect of condition on judgements of the interviewee. Table 5, below, shows the p-values from the post-hoc pairwise comparison testing for traits which showed a significant main effect in Table 4.

Trait	*Audio ~ Orthographic transcript*	*Audio ~ Detailed transcript*	*Orthographic ~ Detailed transcript*
Aggressive	**<0.001*** **	0.31	**<0.001*** **
Calm	**<0.001*** **	**0.01***	0.06
Agitated	**<0.001*** **	0.10	**0.02***
Defensive	**0.002***	0.28	**0.006***
Angry	**0.02***	0.41	**0.02***
Relaxed	**0.04***	**0.01***	0.26
Disgusted	**0.02***	0.48	**0.03***

Table 5. Post-hoc testing of significant effects in Experiment 2

The most notable feature about the results in Table 5 is that they show different patterns to the results from Experiment 1. In Experiment 2, there were only two traits which displayed a significant difference in participant judgements between the audio and the linguistically detailed transcript conditions. For judgements of how *calm* and *relaxed* the interviewee was, there was a significant difference between the audio condition and both transcript conditions, but no significant difference between judgements in the two transcript conditions. These effects are shown in Figure 8, below. Figure 8 shows that the interviewee was judged to be more *calm* and more *relaxed* in the audio condition compared to the two transcript conditions. This aligns with the direction of difference shown in Experiment 1, where the interviewee was perceived more favourably in the audio condition than either of the transcript conditions.

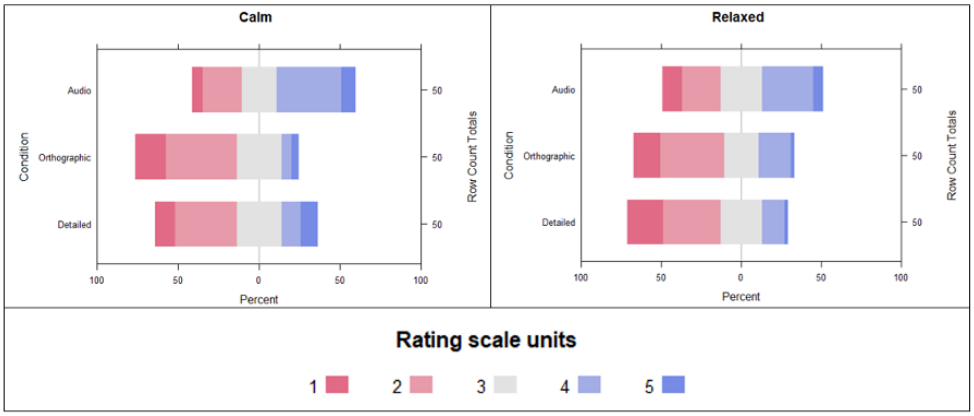

Figure 8. Listener judgements of how *calm* and *relaxed* the interviewee was in each condition

For judgements of how *aggressive, agitated, defensive, angry* and *disgusted* the interviewee was, the differences in judgements patterned in the direction that might be expected if the provision of linguistic detail made the audio recording and linguistically detailed transcript more alike. For these traits, there was no significant difference between the audio recording and the linguistically detailed transcript, a significant difference between the audio recording and the plain orthographic transcript, and a signif-

icant difference between the linguistically detailed transcript and the plan orthographic transcript. These effects are illustrated in Figure 9, below.

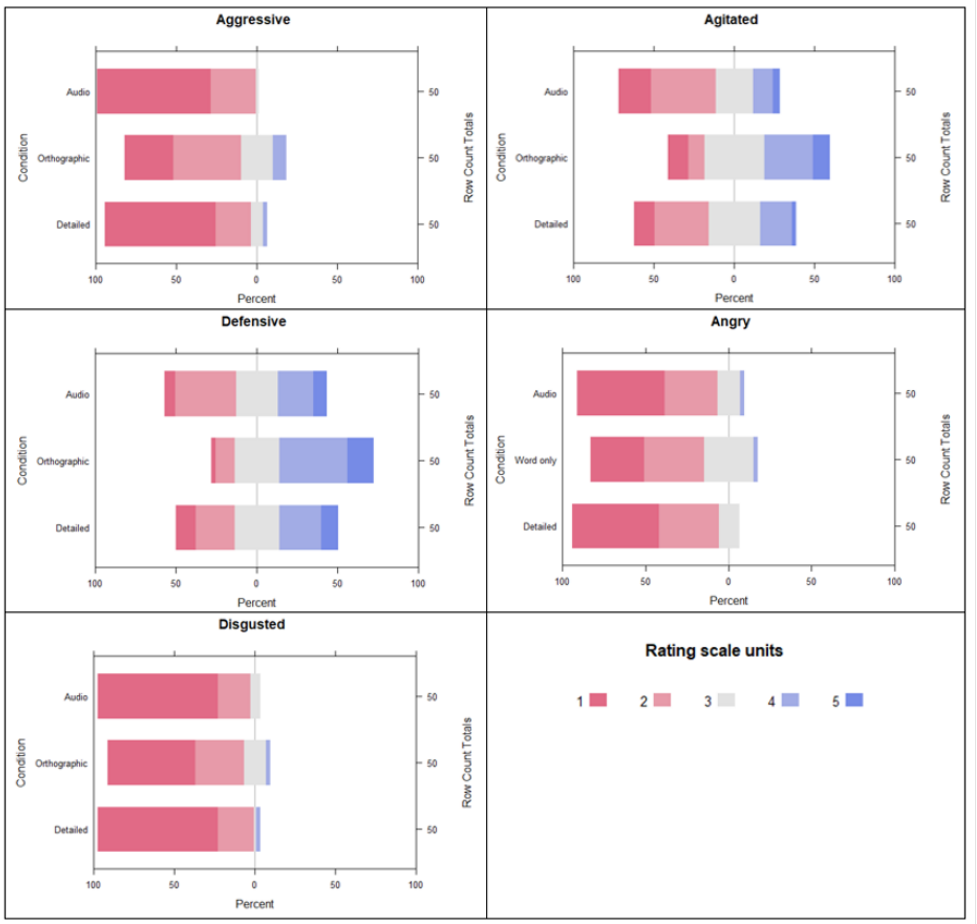

Figure 9. Listener judgements of how *aggressive, agitated* and *relaxed* the interviewee was in each condition

Similarly to Experiment 1 and the other significant effects from Experiment 2, the significant differences between the audio recording and the plain orthographic transcript all resulted in a more negative perception of the interviewee in the written modality. The interviewee was perceived to be more *aggressive, agitated, defensive, angry* and *disgusted* in the plain orthographic transcript condition compared to the audio condition. However, unlike in Experiment 1, the provision of the extra linguistic detail in Experiment 2 mitigated the effect and brought perceptions of the interviewee for these traits closer to the corresponding audio recording.

Qualitative analysis of judgements in Experiment 1 and Experiment 2

It could be considered somewhat puzzling that two sections of the same interview, involving the same speakers, could produce such divergence in the patterning of perceptual judgements. There could be many reasons for the differences between Experiments 1 and 2, including random variation and perceptual instability in judgements of interviewees from audio recordings and transcripts, an interaction between the verbal content and the non-verbal information in the two interview clips, or differences in the degree to which participants used the non-verbal aspects of the interviews to inform their de-

cisions about what they were reading or hearing. Tompkinson et al. (2023) illustrated how participants were much more likely to specifically attend to, and base decisions on, pausing behaviour when pauses were marked in a transcript compared to being present in an audio recording. This led Tompkinson et al. (2023, p. 45) to argue that overtly representing linguistic features in transcripts can cause readers to pay more attention to aspects of speech that would have gone relatively unnoticed in the corresponding audio recordings.

Linking this to the results in the current experiments, it is possible that the degree to which participants noticed the features of speech in the audio recordings could explain why there was a difference between the results in the experiments. In both experiments, participants were asked to provide free-text responses to explain which aspects of the interviewee's language use influenced their judgements of sincerity, plausibility, credibility and truthfulness. Analysis of these responses provides an insight into which language features were influencing participants' judgements across the different conditions. Table 6, below, shows the number of times each of the features that were overtly marked in the detailed transcripts (*pausing, emphasis, overlapping speech, laughter* and *whispered speech*), were mentioned in the free-text responses for the different conditions in each experiment.

	Experiment 1				**Experiment 2**		
	Audio	**Detailed transcript**	**Plain transcript**		**Audio**	**Detailed transcript**	**Plain transcript**
Pausing	7	39	1		34	37	3
Emphasis	3	3	1		3	8	1
Overlap/ interruption	2	2	0		1	7	10
Whisper	0	0	0		0	1	0
Laughter	0	0	0		0	0	0

Table 6. Number of mentions of different linguistic features in participants' free-text responses

The data in Table 6 shows a clear difference between Experiment 1 and Experiment 2 with regards to the number of mentions of pauses as an influencing factor on listeners' judgements of truthfulness, sincerity, credibility and/or plausibility. In Experiment 1, participants in the detailed transcript condition attended to the presence of pauses far more frequently (n=39) than participants who heard the corresponding audio recording (n=7). However, in Experiment 2, a similar number of participants in both the audio (n=34) and detailed transcript (n=37) conditions mentioned pausing as having an influence on their judgements. This difference could potentially explain why the provision of extra linguistic detail in Experiment 1 created more perceptual divergence in listener judgements, whereas in Experiment 2, there was a greater alignment in judgements between the audio and detailed transcript conditions. Analysis of the overall length of the pauses in the speech of the interviewee across the two experiments also showed that in Experiment 1, there were no pauses that were longer than two seconds, whereas in Experiment 2, there were three examples of longer pauses of 2.9 seconds, 5.8 seconds, and 7.9 seconds respectively. This raises the possibility that listeners only attended to the

longer pauses in Experiment 2, and that these were the pauses that were more noticeable and influential on listener judgements. The placement of these pauses within the overall interaction could have also been important in shaping listeners' views. Figure 10 shows the three pauses discussed above, all highlighted in yellow. Two of the pauses come between a question and the answer (lines 57 and 67), and another comes within the middle of an answer given by the interviewee (line 60). Although it is impossible to predict or infer the effect of this on every listener, there was no comparable discourse structure in the extract of audio used for Experiment 1. The specific effect of this kind of linguistic issue could be tested in future research on this topic.

52 IR: Okay. Could you explain then. We've got CCTV footage of
53 your car (.) with you driving and it looks like a person
54 in the passenger seat. (1.25 sec). Would you like to think
55 again if- (0.5 sec) what that could possibly have been?
56 (0.8 sec) Or who?
57 (2.9 sec)
58 IE: When was this?
59 IR: On Thursday evening leaving work.
60 IE: ((*whisper*)) Leaving work. (7.8 sec). I'm not too sure. I
61 think I might have stopped off or dropped someone off from
62 work.
63 IR: Okay. So maybe a short journey [with someone in the car]?
64 IE: [Yeah yeah].
65 (0.8 sec)
66 IR: Can you think who that might have been maybe?
67 (5.7 sec)
68 IE: I think it was the new- the new lass at work.

Figure 10. Extract from Experiment 2 showing the longer pauses in the transcript

3. Discussion

Our first research question focussed on whether there would be differences between orthographic transcripts, linguistically detailed transcripts, and audio recordings of police interviewees in relation to social trait perception. The results in both experiments showed that there was at least the potential for a change in the format of the police interview evidence to affect perceptions of the interviewee, with differences shown across a range of traits in both interviews. This aligns with the findings of Deamer et al. (2022) and Tompkinson et al. (2023), which both observed differences in listener and reader judgements of police interview evidence depending on whether the perceiver was provided with a transcript of the interview or an audio recording.

However, in this study we have illustrated an additional issue. Our results indicate that there is at least the potential for the overt marking of additional linguistic features using a standardised CA-style system to make listener perceptions more divergent from the corresponding audio recording when compared to a plain orthographic transcript. This was particularly true in Experiment 1, where listeners paid minimal attention to the marked linguistic features in the audio condition, and where judgements were less favourable for the interviewee when participants were presented with a linguistically detailed transcript. This is clearly an undesirable outcome for police interview transcripts, and suggests that marking features in a transcript that might not be clearly perceived in the corresponding audio recording could be a problematic practice, even if those features are marked accurately. The results from Experiment 1 would support the proposition by Fraser (2022) that the provision of CA-style representations in legally relevant transcripts could serve to mislead readers and cause them to misinterpret information within a transcript.

However, the findings from Experiment 2 suggest that when listeners attend to the features that are marked in the transcript in the corresponding audio recording, then perceptions of the interviewee in both conditions can be more closely aligned, and crucially different from judgements made from a plain orthographic transcript. This adds a level of complexity to the discussion of the findings from Experiment 1, suggesting that more research is needed to better understand how listeners respond to different features and their realisations in transcripts. It appears that there may be a 'tipping point', at which the representation of non¬verbal features in transcripts becomes more of a hinderance than a help with respect to social trait perception. If this were the case, the transcripts used in Experiments 1 and 2 would sit on either side of this tipping point. This is particularly interesting for the interview clips used in this experiment as they were both taken from recordings of the same speaker, showing that it is possible for two sections of the same interview with the same speaker to produce different perceptual results. This relates to our second research question and suggests a lack of automaticity in assuming that the same speaker will always be perceived similarly by listeners of audio recorded interviews or readers of transcripts.

4. Conclusion

The research in this article has attempted to provide some empirical testing to contribute to the wider debate about whether it is beneficial for transcripts of police interviews to overtly mark aspects of speech such as pausing, overlapping speech, emphasis and other non-verbal cues such as laughter and whispering. By adopting an approach centred around social trait perception, the research in this study has illustrated how the different ways in which police interview evidence can be presented can contribute to judgements of the interviewee.

Although the research is small in scope, using two separate interview extracts from the same speaker, we have aimed to illustrate some of the complexities around the effect that overtly marking linguistic information in transcripts can have on how interviewees are perceived by listeners of audio recordings and readers of transcripts. Our results suggest that caution is required, and more research needed, before recommendations are made that non-verbal cues should be marked in police interview transcripts. The results of this study indicate that the inclusion of additional linguistic detail in transcripts did

not systematically make perceptions of interviewees more aligned with corresponding audio recordings, but also that there could be some features for some interviews which are worthy of representing. However, how this could be done systematically remains a significant challenge, and even if some features are worth representing in transcripts, it is still not clear *how* they should be marked and whether this kind of standardisation could ever be achieved in practice. This issue could be further complicated by the rapid development and potential appetite for the use of automatic transcription systems for the creation of police interview transcripts (Harrington, 2023). This could be a focus for further work in this area, but it would appear to be a significant potential source of variation in transcript production. There is also a wider range of potentially relevant 'additional' features which could be present in police interview recordings which are not considered in the current study as they did not appear in the audio recordings, such as crying or shouting. Again, this issue could be further explored in additional research in this area. Finally, future research could also explore the effect that the provision of video recordings of police interview interactions has on perceptions of interviewees, in comparison to audio recordings and transcripts. More broadly, the research in this article illustrates the connected nature of speech perception and person perception, and highlights some potential complexities regarding perceptions of interviewees in this important and legally relevant context. Given these complexities, we would argue that the initial recommendation by Haworth (2018), that people who are required to assess police interview evidence at any stage of the legal process should be exposed to the original audio recording rather than being left to rely solely on a transcript, is a cautious but sound basis on which to use this important form of evidence.

References

Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics*, *32*(10), 1439–1465. Retrieved 2025-08-13, from https://linkinghub.elsevier.com/retrieve/pii/S0378216699000946 doi: 10.1016/S0378-2166(99)00094-6

Deamer, F., Richardson, E., Basu, N., & Haworth, K. (2022). For the Record: Exploring variability in interpretations of police investigative interviews. *Language and Law=Linguagem e Direito*, *9*(1), 25–46. Retrieved 2025-08-13, from https://ojs.letras.up.pt/index.php/LLLD/article/view/12825/11681 doi: 10.21747/21833745/lanlaw/9_1a2

Eades, D. (1996). Verbatim courtroom transcripts and discourse analysis. In H. Kniffka (Ed.), *Recent developments in forensic linguistic* (pp. 241–254). Frankfurt: Lang.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3-4), 169–200. Retrieved 2025-08-13, from https://www.tandfonline.com/doi/full/10.1080/02699939208411068 doi: 10.1080/02699939208411068

Fraser, H. (2022). A Framework for Deciding How to Create and Evaluate Transcripts for Forensic and Other Purposes. *Frontiers in Communication*, *7*, 898410. Retrieved 2025-08-13, from https://www.frontiersin.org/articles/10.3389/fcomm.2022.898410/full doi: 10.3389/fcomm.2022.898410

French, P., & Fraser, H. (2018). Why "Ad Hoc Experts" should not Provide Transcripts of Indistinct Audio, and a Better Approach. *Criminal Law Journal*, *42*(5), 298–302.

Gibbons, J. (2003). *Forensic Linguistics: an introduction to language in the justice system.* Oxford: Blackwell.

Gold, E., Ross, S., & Earnshaw, K. (2018). The 'West Yorkshire Regional English Database': Investigations into the generalizability of reference populations for forensic speaker comparison casework. In *Interspeech 2018: Speech Research for Emerging Markets in Multilingual Societies* (pp. 2748–2752).

Harrington, L. (2023). Incorporating automatic speech recognition methods into the transcription of police-suspect interviews: factors affecting automatic performance. *Frontiers in Communication*, *8*, 1165233. Retrieved 2025-08-13, from https://www.frontiersin.org/articles/10.3389/fcomm.2023.1165233/full doi: 10.3389/fcomm.2023.1165233

Harrington, L. (2024). *Towards improving transcripts of audio recordings in the criminal justice system* (PhD thesis). University of York.

Haworth, K. (n.d.). Process, procedure and professional practice in the creation of police interview evidence. In T. Grieshofer & K. Haworth (Eds.), *Language and Justice: Communication in Legal Practice.* Cambridge: CUP.

Haworth, K. (2018). Tapes, transcripts and trials: The routine contamination of police interview evidence. *The International Journal of Evidence & Proof*, *22*(4), 428–450. Retrieved 2025-01-28, from https://journals.sagepub.com/doi/10.1177/1365712718798656 doi: 10.1177/1365712718798656

Haworth, K., Tompkinson, J., Richardson, E., Deamer, F., & Hamann, M. (2023). "For the Record": applying linguistics to improve evidential consistency in police investigative interview records. *Frontiers in Communication*, *8*, 1178516. Retrieved 2025-08-13, from https://www.frontiersin.org/articles/10.3389/fcomm.2023.1178516/full doi: 10.3389/fcomm.2023.1178516

Leemann, A., Perkins, R., Buker, G. S., & Foulkes, P. (2024). *An Introduction to Forensic*

Phonetics and Forensic Linguistics. London New York: Routledge, Taylor & Francis Group.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org

Richardson, E., Hamann, M., Tompkinson, J., Haworth, K., & Deamer, F. (2023). Understanding the role of transcription in evidential consistency of police interview records in England and Wales. *Language in Society*, 1–32. Retrieved 2025-01-28, from https://www.cambridge.org/core/product/identifier/S004740452300060X/type/journal_article doi: 10.1017/S004740452300060X

Richardson, E., Haworth, K., & Deamer, F. (2022). For the Record: Questioning Transcription Processes in Legal Contexts. *Applied Linguistics*, *43*(4), 677–697. Retrieved 2024-06-08, from https://academic.oup.com/applij/article/43/4/677/6524571 doi: 10.1093/applin/amac005

Rock, F. (2020). Witnesses and Suspects in Interviews: Collecting Oral Evidence: The Police, the Public and the Written Word. In M. Coulthard & A. May (Eds.), *The Routledge handbook of forensic linguistics* (pp. 112–126). London and New York: Routledge.

Tompkinson, J., Haworth, K., Deamer, F., & Richardson, E. (2023). Perceptual instability in police interview records. *The International Journal of Speech, Language and the Law*, *30*(1), 22–51. Retrieved 2025-08-13, from https://utppublishing.com/doi/10.1558/ijsll.24565 doi: 10.1558/ijsll.24565