

The power and persistence of contextual priming: more risks in using police transcripts to aid jurors' perception of poor quality covert recordings*

*By Helen Fraser** and
Independent researcher*

*Bruce Stevenson***
Senior Lecturer, School of Behavioural Cognitive and Social
Sciences, University of New England, Australia*

Abstract A poor quality covert recording from an Australian murder case, along with the police transcript used in the trial but later shown to be inaccurate, are used to explore general issues regarding this increasingly common type of evidence. Two experiments were run, in which participants heard an excerpt from the audio, first with no transcript, then with suggested and alternative

* Thanks are due to the 'father' in this case, and his family, for allowing the audio and transcripts to be used in these experiments, and also to participants in both experiments and in a series of pilots.

** Email: helen@helenfraser.com.au.

*** Email: bstevens@une.edu.au.

transcripts. In Experiment 1, they were given no contextual information, while Experiment 2 started with a background story about the case and the issue the recording was intended to resolve. Results indicate that background knowledge of a case can dramatically increase listeners' acceptance of a police transcript, even when the transcript is manifestly inaccurate. It is suggested that such contextual priming may affect not just juries but others involved with the trial, and recommended that police transcripts be treated with more caution than is currently common with Australia's 'ad hoc expert' rules.

Keywords Forensic transcription; Forensic phonetics; Priming; Cognitive bias; Digital evidence; Indistinct audio

Covert recordings are used as evidence in increasing numbers of criminal cases, often providing useful information not available by other means. One problem is that, due to the manner in which they are obtained, their quality is often very poor—to the extent that, without prior knowledge of the contents, few if any words can be clearly identified.

For this reason, it is common, in Australian and other jurisdictions, for covert recordings to be transcribed by those with relevant background knowledge, typically (but not only) by detectives investigating the case. Of course, transcribing poor quality audio, even for someone with background knowledge, requires repeated listening. In Australian and other jurisdictions, the law grants the status of so-called 'ad hoc expert' to those who have listened many times to a recording. This status allows detectives to provide their transcripts in court, as an aid to the jury's perception of the audio.¹

Unfortunately, police transcripts are frequently inaccurate, incomplete or otherwise misleading.² To prevent potentially adverse effects on jurors' interpretation of the audio, judges in cases involving police transcripts are required to caution the jury to use the transcript only as an aid, relying on their own ears to decide what is actually said in the recording.³

Is this a sufficient safeguard? Phonetic science would suggest not.⁴ Well-established evidence demonstrates that seeing an inaccurate transcript can influence

1 See, e.g., Evidence Act 1995 (NSW), s. 79(1).

2 M. Coulthard, 'Some Forensic Applications of Descriptive Linguistics' (2005) 9 *VEREDAS—Revista de Estudios Lingüísticos* 9; P. French and P. Harrison, 'Investigative and Evidential Applications of Forensic Speech Science' in A. Heaton-Armstrong, E. Shepherd, G. Gudjonsson and D. Wolchover (eds.), *Witness Testimony: Psychological, Investigative and Evidential Perspectives* (Oxford University Press: Oxford, 2006) 247; R. Shuy, *Language Crimes: The Use and Abuse of Language Evidence in the Courtroom* (Blackwell: Oxford, 1993).

3 See, e.g., J. Wood, *Report 136: Jury Directions* (NSW Law Reform Commission: Sydney, 2012) sections 6.52–6.56.

listeners' perception of audio they would otherwise interpret quite differently, or not be able to interpret at all. This effect is usually known as 'priming' in this context, but is closely related to well-known psychological concepts such as 'suggestibility' and 'cognitive bias'.⁵ Importantly, listeners typically have no awareness that they have been primed by a transcript, confidently believing they simply heard what was 'there to be heard'.

Further, recent research⁶ confirms that, due to the persistent power of priming, juries, once exposed to a particular transcript of poor quality audio, may be unable to disregard its interpretation and re-evaluate the audio from a neutral perspective. Again, this is in line with a phenomenon well known to psychology as the 'continuing influence effect'.⁷

The present article goes beyond priming by a transcript, to consider contextual priming, i.e. the effect on interpretation of poor quality audio of listeners' knowledge or assumptions about the context in which a recording was made.

It reports two new experiments, both using a poor quality covert recording previously presented as evidence in an Australian murder case, along with the police transcript provided at the trial to aid the jury in hearing what was said in the recording, and later shown to be inaccurate (see below). Of course, the intention of the experiments was not to reach any conclusions about this particular case. Rather, their aim was to use these materials, typical of those appearing in criminal cases in Australia and elsewhere, to explore general issues.

In both experiments, an online survey tool (Qualtrics) was used to allow participants to play an excerpt from the audio, first with no transcript, then with the police transcript and an alternative transcript. Experiment 1, in which participants listened to the audio with no contextual information, demonstrates that the police transcript is not only inaccurate but implausible. Experiment 2, which started by giving a story about the case and the issue the recording was intended to resolve, demonstrates that contextual knowledge can have a dramatic and lasting

4 See H. Fraser, 'Issues in Transcription: Factors Affecting the Reliability of Transcripts as Evidence in Legal Cases' (2003) 10 *International Journal of Speech Language and the Law* 203.

5 A. Ridley, F. Gabbert and D. La Rooy (eds.), *Suggestibility in Legal Contexts* (Wiley-Blackwell: 2013); D. Kahneman, *Thinking, Fast and Slow* (Farrar Straus Giroux: New York, 2011).

6 H. Fraser, B. Stevenson and T. Marks, 'Interpretation of a Crisis Call: Persistence of a Primed Perception of a Disputed Utterance' (2011) 18 *International Journal of Speech Language and the Law* 261.

7 U. Ecker, S. Lewandowsky and D. Tang, 'Explicit Warnings Reduce But Do Not Eliminate the Continued Influence of Misinformation' (2010) 38 *Memory & Cognition* 1087.

priming effect in its own right, and can also increase and prolong the priming effect even of a manifestly inaccurate transcript. Further, it confirms⁸ that such priming can influence listeners' opinions about guilt.

Discussion considers how it so frequently happens that inaccurate transcripts are accepted as sufficiently reliable to admit as an aid to the jury's perception of indistinct audio, draws attention to problems with the concept of 'ad hoc expert', and advocates the need for transcripts to be properly evaluated by a genuine expert in relevant branches of phonetic science before admission, and preferably before circulation to defence and prosecution teams.

1. Background

The case concerned a father, son and grandfather from a single family. One night the son visited the grandfather, drank some beer with him, and then shot him. The grandfather died the next day of his wounds. After several weeks, police arrested the son, who eventually confessed to the murder, claiming the grandfather had repeatedly humiliated and abused him. A few days later, the father was arrested. A conversation between father and son, recorded by a covert listening device, was alleged by police to include a confession showing the father had been an accomplice in the murder. No phonetics expert was consulted at the time, by either prosecution or defence. However, the audio and transcript were provided to the first author after the trial was concluded.

The experiments reported here used the extremely poor quality covert recording actually played in court. This had been 'enhanced' by an audio engineer. Despite some effort, it has not been possible to find out the exact nature of the 'enhancing'. However, the detective responsible for the recording testified in court that the process had made 'no difference' to the auditory quality of the recording.

The 'enhanced' audio was in the form of a series of 38 1-minute tracks, all of very low amplitude, in 44kHz AIFF format (re-recorded from VHS videotape). For present purposes, the tracks were concatenated, the amplitude normalised (making them sound considerably louder than previously), and the file format changed to 22kHz MP3 (to facilitate deployment over the internet).

As a preliminary to the experiments, three people (graduate students in law) were asked to listen to the 38-minute recording right through, on studio quality

⁸ Cf. Fraser *et al.*, above n. 6.

equipment with professional headphones, but with no background information beyond the fact that the material related to a murder case. They reported being unable to make out more than a few random words, to have no idea what the conversation was about, and to have had great difficulty concentrating on the audio for its entire duration. It is worth emphasising that audio of this quality is quite typical of covert recordings admitted as evidence in criminal trials in Australia and elsewhere.

Several pilot experiments were then run, using 6-minute and 4-minute excerpts from the recording. However, it proved difficult to find participants who could persevere through multiple listenings to excerpts of this duration. The experiments reported here therefore used a 14-second excerpt (described below). Readers wishing to hear the excerpt before continuing (recommended) can do so at <http://forensictranscription.com.au/audio>.

2. Experiment 1 (Transcripts with no context)

2.1 Participants

Experiment 1 had 56 participants (after exclusion of a small number who reported prior exposure to the material). They were recruited via academic and personal email lists, and randomly divided (by the survey software) into two groups of 30 and 26. As shown in Appendix 1, they were diverse across a range of variables, including age, gender, education and occupation. There is no statistically significant difference between the two groups on any demographic characteristic measured.

2.2 Materials

2.2.1 Audio

The 14-second excerpt used in the experiments was chosen for the key role that a phrase it allegedly contained played in the trial. Like the entire recording, this excerpt was of extremely poor quality. Voices can be heard, in which one man (acknowledged to be the father in the case) appears to speak five separate phrases in a whispered, urgent tone, and another man (the son) responds several times with ‘mm’. However, few words can be heard with any confidence. The entire excerpt, and indeed most of the full recording, would likely have been evaluated by a phonetics expert, if called, as ‘untranscribable’—i.e. while interpretations might be offered, none can be verified with sufficient confidence for use as evidence in a trial.⁹

⁹ See Fraser *et al.*, above n. 6.

The audio was presented to participants with a special player which facilitated repeated listening to all or any chosen part of the excerpt.

2.2.2 Transcripts

The experiment used two transcripts of the excerpt, focusing on the father's speech. The first was the police transcript, as tendered in court, which interpreted the five phrases as shown in Table 1, Transcript 1. Phrase 3 was particularly important in the trial.

Table 1. The two transcripts used in the experiments

Phrase	Transcript 1	Transcript 2
1.	I'm not saying you're going in ... fucking ... you know	I'm not saying you're going in ... fucking ... you know
2.	just in case mate	just in case mate
3.	at the start we made a pact (the pact phrase, or PACT)	it's fucking payback (the payback phrase, or PAYBACK)
4.	in it till the end	in it till the end
5.	you know what I mean	you know what I mean

Transcript 1 can readily be demonstrated to be unreliable. While a few words can be given some limited credence (notably Phrase 5, *You know what I mean*), none can be verified with any confidence, and some (notably Phrase 3, *At the start we made a pact*) are contradicted by what can be discerned of the acoustics, using auditory and visual analysis (see Figure 1). For example, the overall rhythm of the phrase does not fit the alleged phrase; the vowel alleged to be that of 'at' has a high front quality; the 's' alleged to be the start of 'start' appears more closely associated with the preceding syllable; the vowel interpreted as the /a/ of 'start' is too short; what should be the /m/ of 'made' sounds like an aspirated oral stop; the consonant at the start of the word alleged to be 'pact' appears unaspirated.

The experiment also used an alternative interpretation of the audio, shown as Transcript 2 in Table 1. This differs from Transcript 1 only in Phrase 3. It is important in understanding the results of the experiments to be clear that Phrase 3 is not suggested as an accurate interpretation of the audio. It was simply made up for these experiments as a plausible, but unverifiable, distractor from Transcript 1.

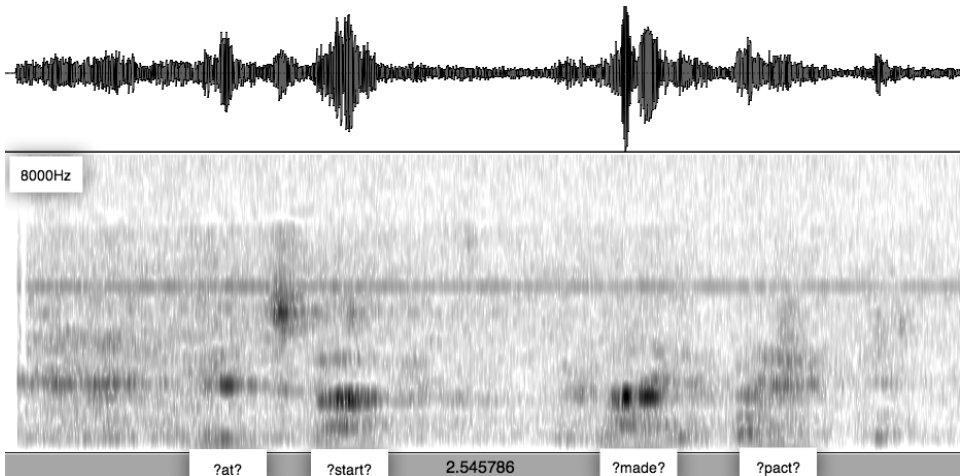


Figure 1. Spectrogram of Phrase 3, alleged to be 'At the start we made a pact'

2.3 Procedure

Participants first read an information sheet and clicked 'consent' to proceed. They then tested their computer audio, and learned how to use the audio player and timer by listening to a list of words and answering the question 'what word do you hear at Time x?'.

They were then given the experimental audio, with the following information.

Here is the audio from the case. It features a whispered conversation between two men. One mostly just says 'Mm'. We are interested in what the other man says. The recording is 14 seconds in duration. You may play it as often as you wish. Remember you can move to a specific time by clicking in the grey timeline and noting the time on the left.

On the next screen we will give you a transcript to evaluate. First, we would like to know what you can hear unaided. Please note any words you hear in the boxes below, next to the appropriate time. The times are arranged to roughly match the phrasing of the speech. You can use ? for any parts you are not sure of. If you cannot make out any words for a section, just leave it blank. We are particularly interested in what you hear from 6.5 to 9 seconds.

Having read the information and listened to the audio, participants were invited to write what they thought was said in each phrase into text boxes set up to show the approximate timing of each phrase (see Figure 2). Here and for all questions,

0 to 4 seconds	<input style="width: 100%; height: 30px;" type="text"/>
4 to 6.5 seconds	<input style="width: 100%; height: 30px;" type="text"/>
6.5 to 9 seconds	<input style="width: 100%; height: 30px;" type="text"/>
9 to 11 seconds	<input style="width: 100%; height: 30px;" type="text"/>
11 to 14 seconds	<input style="width: 100%; height: 30px;" type="text"/>

Figure 2. *Text boxes used to collect participants' interpretations of the audio*

participants were also given text boxes for free-form comments on any aspect of their experience.

Participants were then provided with a suggested transcript, and asked to listen again to the audio (as often as they wished), state whether they agreed with each phrase of the transcript, and, if they disagreed, to enter what they heard into text boxes like those in Figure 2.

Finally, they were given an alternative transcript, with attention drawn to the fact that the only difference was in Phrase 3. Again they were asked to listen to the audio as often as they wished, state whether they agreed with each phrase, and if not, to enter what they now heard into text boxes.

Group 1 saw Transcript 1 first, as their suggested transcript, then Transcript 2, as their alternative transcript. Group 2 saw Transcript 2 first, then Transcript 1 (see Table 2).

At the end, demographic data were collected, along with information about participants' hearing, headphone use, etc. (see Appendix 1).

2.4 Analysis

Participants' interpretations of the crucial phrase (3) were categorised as in Table 3, based on their agree/disagree responses, their proffered transcriptions, and any free-form comments.

Table 2. Design of Experiment 1

	Group 1	Group 2
Open condition	No transcript	No transcript
Suggested transcript	Transcript 1 (with PACT phrase)	Transcript 2 (with PAYBACK phrase)
Alternative transcript	Transcript 2 (with PAYBACK phrase)	Transcript 1 (with PACT phrase)

Table 3. Categorisation of participants' interpretations of Phrase 3

Response category	Criteria for inclusion in category
PACT	The exact 'pact' phrase
PACT SIM	Anything similar to the 'pact' phrase, either in terms of overall rhythm, or by virtue of including the word <i>pact</i> , or a similar-sounding word, such as <i>pack</i> , <i>impact</i> or <i>backed</i>
PAYBACK	The exact 'payback' phrase
PAYBACK SIM	Anything similar to the 'payback' phrase, either in terms of overall rhythm, or by virtue of including the word <i>payback</i> or a similar-sounding word or phrase, such as <i>playback</i> , <i>layin' back</i> , <i>pay it back</i> , <i>pay you back</i>
OTHER	Something different from either PACT or PAYBACK Some examples: <i>it's that plain bag there</i> ; <i>it's fucked bloody bad</i> ; <i>it's not reliable</i> ; <i>it's back, try your best</i> ; <i>it's about the main advantage</i> ; <i>it's back I begged</i> ; <i>yeah yeah it's fine</i> Note that OTHER was used if any words or parts of words at all were offered, not only for complete phrases
NOTHING	No words or part-words at all were offered

2.5 Results (Experiment 1)

2.5.1 The non-crucial phrases

It is useful to consider first the non-crucial phrases (1, 2, 4 and 5). Here data from both groups are combined, since both had the same experience of these parts of

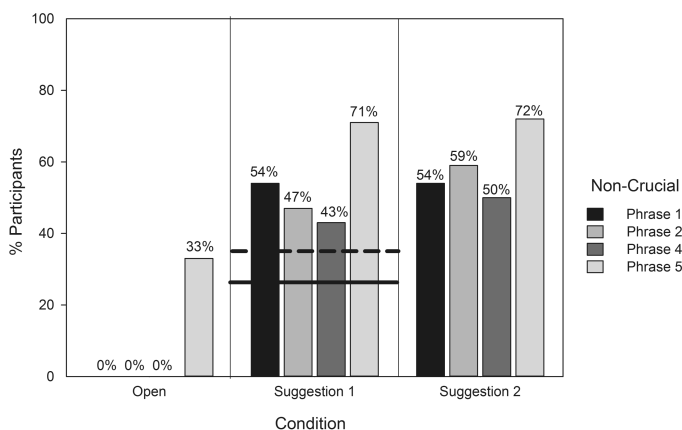


Figure 3. Percentage of Expt 1 participants who heard each non-crucial phrase exactly as in transcript under each condition. Phrase 3 results are shown with horizontal lines (solid: Group 1, PACT, 27%; dashed: Group 2, PAYBACK, 35%)

the audio, which they heard once in Open condition (with no transcript), then twice with transcripts differing only in Phrase 3.

Figure 3 presents the percentage of all participants who agreed with each phrase of the transcript in Open condition, and upon first and second suggestion of the transcript.

As can be seen, Phrases 1, 2 and 4 were heard by no one in Open condition. However, on first suggestion of the transcript, around half the participants claimed to hear the words presented for each phrase. This is a typical ‘priming’ effect for hard-to-hear audio: material that seems uninterpretable on its own suddenly becomes ‘clear’ to many listeners when a transcript seems to ‘aid’ perception—even if, as in this case, the accuracy of the transcript is highly questionable. For this reason, these phrases provide a useful baseline with which to compare results for the two transcripts of the crucial phrase (3) when they were presented as each group’s Suggestion 1. Phrase 3 results are shown in Figure 3 with horizontal lines (solid and dashed), and discussed below.

A similar priming effect is observed for Phrase 5, *You know what I mean*. It starts from a significantly higher base than the other phrases (33% compared to 0%, Cochran $Q(3) = 57.0, p < .001$), and reaches a significantly higher level (71% compared to 54%, $Q(1) = 5.00, p < .05$), reflecting the greater plausibility of the transcript of this phrase. However, it shows a similar significant change in numbers hearing the phrase without a transcript (33%) to hearing it on first suggestion of the transcript (71%) ($Q(1) = 38.3, p < .001$).

While differences between no transcript and first suggestion are significant for all phrases, differences between first suggestion and second suggestion are significant only for Phrase 2 ($Q(1) = 5.14, p < .05$), with a 12% increase in the identification of the content proposed for Phrase 2 by the transcript following a second exposure.

2.5.2 The crucial phrase (3)

2.5.2.1 Open condition

Figure 4 shows how many participants offered interpretations of Phrase 3 in each response category (recall Table 3) in Open condition (i.e. with no transcript). At this stage, conditions for both groups are the same, so data are combined.

Note that no participants offered the ‘pact’ phrase as an interpretation of Phrase 3, and none suggested any similar phrase (PACT SIM = 0). By contrast, PAYBACK was offered by two participants (4%), while similar phrases (PAYBACK SIM) were suggested by 39%. Nearly half the responses were categorised as OTHER, i.e. something unlike either PACT or PAYBACK, with a wide range of interpretations offered (see Table 3), while 13% of participants were unable to decipher any words at all.

2.5.2.2 Agreement with Suggestion and Alternative

Having interpreted the audio in Open condition, participants were shown their group’s suggested and alternative transcripts (which differed only in the crucial

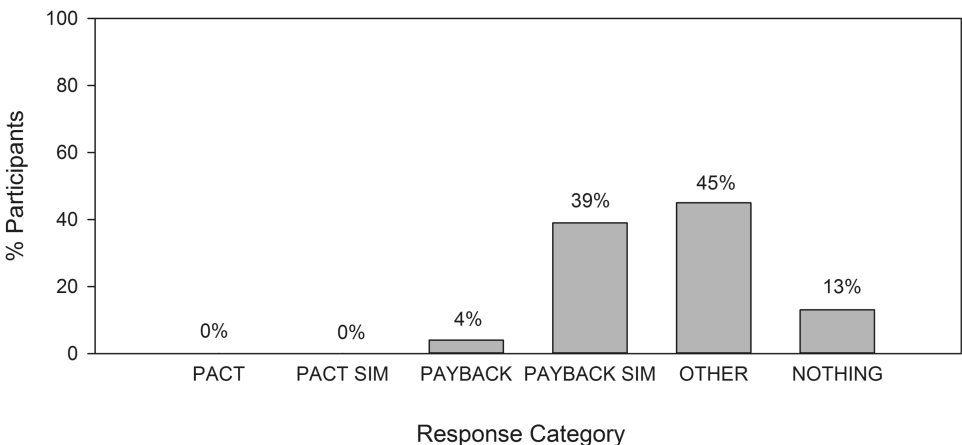


Figure 4. Interpretation of Phrase 3 in Expt 1 ‘Open’ condition (no context and no transcript)

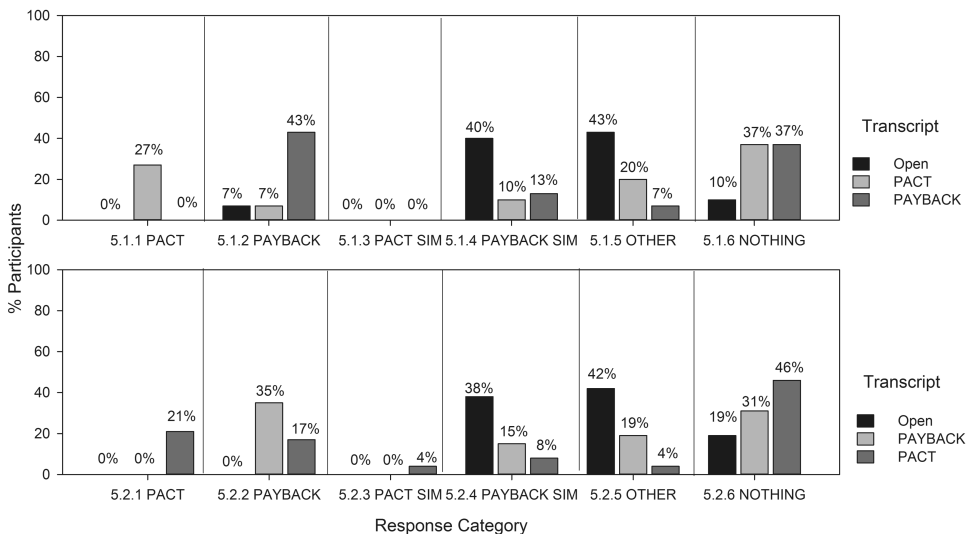


Figure 5. Percentage of Phrase 3 responses in each category under each condition (Expt 1). Upper panel (5.1 Group 1), PACT was presented first; lower panel (5.2 Group 2), PAYBACK was presented first

phrase (3); see Tables 1 and 2). Figures 5.1 and 5.2 show the results. To start, we focus on how many simply agreed with PACT and PAYBACK when each was presented, either as suggestion or as alternative (Panels 5.1.1, 5.1.2, 5.2.1 and 5.2.2).

In Group 1, who received PACT first, 27% agreed to that transcript, while in Group 2, who received PAYBACK first, 35% agreed with their transcript. These are statistically similar. However, while PAYBACK responses (dashed line, Figure 3) are statistically similar to responses for Phrases 1, 2 and 4, PACT responses (solid line, Figure 3) are significantly lower than Phrase 1 ($Q(1) = 8.00, p < .01$) and Phrase 2 ($Q(1) = 4.50, p < .05$), and similar only to Phrase 4, ‘in it till the end’. This can be taken as a first indication of the relative implausibility of the ‘pact’ transcript.

When shown their alternative transcript, 43% of Group 1 accepted PAYBACK having previously seen PACT, while only 21% of Group 2 accepted PACT having previously seen PAYBACK (Panels 5.1.2 and 5.2.1). The apparent preference for PAYBACK does not quite reach statistical significance here ($.10 > p > .05$), but gains weight from consideration of the effect of the alternative transcript on the first suggestion.

2.5.2.3 Effect of Alternative on Suggestion

Participants who did not accept their suggested and alternative transcripts were invited to provide their own interpretation in a comments box (Figure 2), and these were categorised as described in Table 3. This allows us to compare what happened to each group's suggested transcript when their alternative was presented.

In Group 1 (Panel 5.1.1), none of the few who first accepted the suggestion of PACT indicated that they still preferred this interpretation after their alternative, PAYBACK, was offered (27% to 0%, McNemar $\chi^2(1) = 6.13, p < .05$).

In Group 2 (Panel 5.2.1), by contrast, around half of the 35% who initially accepted PAYBACK said they still preferred this interpretation even after PACT was offered. This makes the number of PAYBACK responses statistically similar regardless of which transcript was shown (35% to 17%, McNemar $\chi^2(1) = 2.29, p > .10$), and indeed statistically similar to the PACT responses when the 'pact' transcript was shown (17% and 21%, $Q(1) = 0.11, p > .10$). This further confirms the implausibility of the 'pact' phrase as an interpretation of Phrase 3.

2.5.2.4 Disagreement with Suggestion and Alternative

Panels 5.1.3–5.1.6 and 5.2.3–5.2.6 show interpretations other than PACT and PAYBACK given by participants for Phrase 3. The most important observation is that, apart from the relatively small numbers (Panels 5.1.1 and 5.2.1) who agreed to the exact PACT transcript when it was presented as suggestion (27%) or alternative (21%), virtually no one, in either group or any condition, ever heard anything remotely like the 'pact' phrase. The only response to receive the PACT SIM coding (after PACT had been shown as an alternative; Panel 5.2.3) was *it's like ? we made a pack*.

By contrast, numerous participants heard something generally similar to the 'payback' phrase (PAYBACK SIM; Panels 5.1.4 and 5.2.4). This was heard most often (38% and 40%) in Open condition, with responses dropping significantly upon presentation of either transcript (Group 1 $Q(1) = 7.36, p < .05$; Group 2 $Q(1) = 4.50, p < .05$). However, a substantial proportion of both groups (8% and 15%) continued to offer an interpretation similar to the 'payback' phrase no matter which transcript was presented.

A similar pattern is seen for OTHER responses, which start above 40% in Open condition, and drop significantly across each succeeding condition (Group 1 $Q(2) = 11.63, p < .01$; Group 2 $Q(2) = 15.20, p < .001$).

Together, the falling patterns of PAYBACK SIM and OTHER responses suggest that participants may have been distracted from their own interpretations by seeing the transcripts.

Finally, we can notice the difference in NOTHING responses depending on which transcript was seen first. These are low in Open condition, suggesting that participants are seeking to put some interpretation on the audio. For Group 1 (Panel 5.1.6), NOTHING responses then increase significantly ($Q(1) = 6.40, p < .05$) between Open condition (10%) and PACT suggestion (37%). This perhaps reflects an attitude of ‘I don’t know what it is but I am sure it is not PACT’ (a comment actually made by a number of participants). NOTHING responses then remain similar for the PAYBACK alternative.

For Group 2 (Panel 5.2.6), who saw PAYBACK first, NOTHING responses are statistically similar from Open condition to first suggestion, but trend significantly upwards from Open condition to PACT alternative ($Q(2) = 3.86, p < .05$)—again suggesting that it is rejection of the PACT transcript that is influencing NOTHING responses.

2.6 Discussion (Experiment 1)

The key observation from Experiment 1 is that ‘At the start we made a pact’ is not only an inaccurate but also an implausible transcription of Phrase 3. No one except the original detective transcriber has ever heard the ‘pact’ phrase, or anything remotely like it, without it having been explicitly suggested.

Even when it is suggested, the priming effect of PACT is remarkably low (though far from non-existent), and remarkably non-persistent, with even those few who initially accepted it readily abandoning it when an alternative was suggested.

The relative effectiveness of priming with PAYBACK, and relatively high PAYBACK and PAYBACK SIM responses in all conditions (even in the face of priming with PACT) indicates that this is a generally more plausible interpretation than PACT. That is not an indication that PAYBACK is an accurate transcription, but does add weight to the extreme implausibility of PACT.

In short, Experiment 1 shows that the police transcript of Phrase 3 is so implausible that just listening carefully to the relevant portion of the audio, with no presuppositions and a critical attitude engendered by exposure to an alternative, is enough to demonstrate its inaccuracy to most listeners.

This raises the question of why the manifest unreliability of the police transcript was not picked up before or during the trial by either prosecution (who have the duty of fairness in relation to evidence) or defence (who have the duty of protecting the accused person from false allegations). Some answers to this and other questions may be suggested by the results of Experiment 2, which is a closer approximation to how hard-to-hear covert recordings are experienced in real legal cases.

3. Experiment 2 (Transcripts with context)

3.1 Participants

Participants for Experiment 2 were recruited via academic and personal email lists, and also via websites discussing issues in forensic transcription, including mention (in general terms, not giving specific interpretations of the audio) of the results of Experiment 1. In total, 97 people reporting no previous exposure to the audio did the experiment. Fifteen were excluded due to incomplete data, leaving a total of 82, in a single group. As shown in Appendix 1, this group was similar in demographic composition to the two groups of Experiment 1, except that Experiment 2 participants were younger by 5.7 years ($t(135) = 2.31, p = .02$).

3.2 Materials

The same audio and the same two transcripts were used as in Experiment 1.

3.3 Procedure

The procedure was generally similar to that of Experiment 1, but the conditions under which the audio was presented were different. The intention was to reflect more closely the manner in which covert recordings are usually experienced by juries in criminal trials, where an interpretation is provided in the opening statements, long before the audio is actually heard.

Before hearing the audio for the first time, Experiment 2 participants were heavily primed with a story similar to the one given in **Background** above, but with the following additional information (roughly similar to the actual case).

Police allege this conversation refers to a pact made before the murder, in which the son agreed to kill the grandfather at a time the father had a clear alibi, and the father agreed to share the proceeds of the grandfather's will with him. If arrested, the son was to claim provocation, and serve a short sentence, getting his share of the money when he was released. Existence of such a pact of course would make the father an accomplice to murder, at least as guilty as the son

who pulled the trigger. However, the father denies the allegation, and claims the conversation shows him encouraging his son to turn himself in to police, and making a pact to stand by him no matter what. The listening device recording is crucial evidence in deciding which side is telling the truth. This experiment asks for your opinion as to what kind of pact is discussed in the recording.

Having read this story, participants were asked for a preliminary opinion as to the father's guilt, and then further primed with the following contextual information about the recording.

The full conversation is half an hour in duration. Here you hear just the 14 seconds in which the father discusses the pact. The recording is of very poor quality. You might not hear anything much except the word PACT, but try to make out the sentences in which the word appears, as these contain evidence as to whether the father is guilty or innocent.

Participants were then enabled to listen to the audio, under each of the four sequential conditions shown in Table 4. For each condition, they were able to listen to the audio as often as they wished. From 'Doubt' onwards, they were also able to play a shorter extract (2.5 seconds) containing just Phrase 3, in isolation.

When given a transcript, they were asked to state whether the transcript helped them hear its suggested phrase, and, if not, to say what they heard instead (see Figure 2). At the very end, they were asked to re-rate their impression of the father's guilt.

3.4 Analysis

Responses were coded in the same way as for Experiment 1 (see Table 2).

3.5 Results (Experiment 2)

For Experiment 2, we focus on results for Phrase 3 only.

3.5.1 Open condition

Figure 6 shows the distribution of participants' interpretations of Phrase 3 before they saw any transcript (compare Experiment 1 results in Figure 4).

It is notable that, in Experiment 2 as in Experiment 1, no one heard the exact phrase of the police transcript before it was explicitly suggested, despite the heavy contextual priming. However, the contextual priming caused nearly a third of

Experiment 2 participants to give PACT SIM responses—as opposed to zero participants in Experiment 1, where no background story was given.

Table 4. Conditions of Experiment 2 (all participants in one group)

Open	No transcript
Suggestion	Transcript 1
Doubt	No transcript, but the following information: At this point, the defence call in a phonetics expert, who testifies that the police transcript is inaccurate. In fact, the expert states the word PACT is not spoken at any time in this recording. As well as giving technical evidence, the expert invites you to listen just to the part alleged to be ‘At the start we made a pact’, in isolation from the rest, suggesting this might help you hear for yourself that PACT is not spoken. Here you can listen to the full 14 seconds, and to just the crucial phrase that the police say contains the word PACT.
Alternative	After listening to the phonetics expert, an alternative interpretation of the crucial phrase is offered. Perhaps the father said ‘It’s fucking payback’. Listen again to evaluate that suggestion.

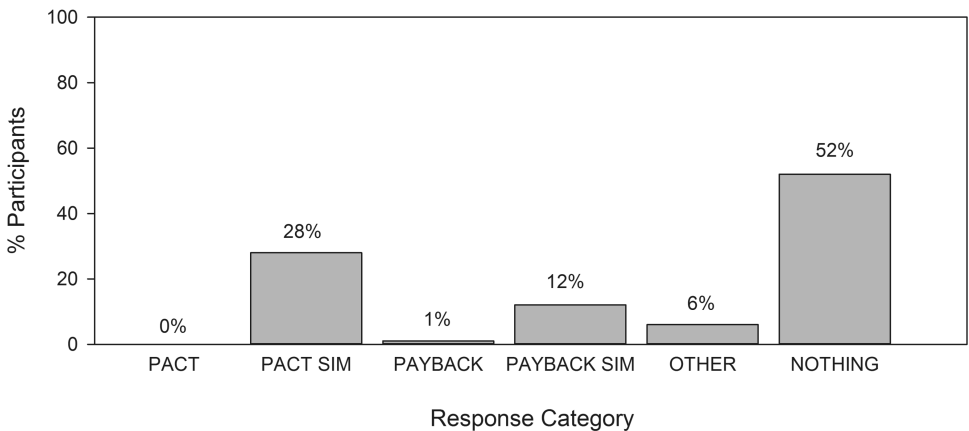


Figure 6. Interpretation of Phrase 3 in Expt 2 'Open' condition (context but no transcript)

It is worth mentioning that these PACT SIM responses were overwhelmingly given on the basis of the specific word ‘pact’ being provided, rather than other criteria listed in Table 3. Indeed, many participants indicated that they could hear ‘pact’ without being able to give any other words to make a phrase. Several commented that they might not have heard the word ‘pact’ without the prior suggestion in the story, but nonetheless indicated that they did indeed hear it (even though we know from phonetic analysis that this word was not spoken, and from Experiment 1, that it was never heard without being suggested).

Also notable is the far lower number in Experiment 2 than Experiment 1 of PAYBACK SIM ($\chi^2(1) = 26.4, p < .001$) and OTHER ($\chi^2(1) = 14.5, p < .001$) responses, and the far higher number of NOTHING responses ($\chi^2(1) = 20.2, p < .001$).

It seems that the contextual priming with a story about a pact persuaded participants to hear Phrase 3 either as something containing the word ‘pact’, or as nothing, and dissuaded them from attempting to interpret it in a neutral manner.

3.5.2 Agreement with Suggestion and Alternative

After listening to the audio in Open condition, participants were shown Transcripts 1 and 2 under conditions summarised in Table 4. Figure 7 shows the results. This can be compared with Experiment 1 results in Figures 5.1 and 5.2 to see the effect of knowing the background context on interpretation of this crucial phrase.

Again we start by considering the number of participants who agreed with the PACT and PAYBACK transcripts when they were presented (Panels 7.1 and 7.2). Note that twice as many participants agreed to the PACT transcript when it was first suggested in Experiment 2 (55%) as in Experiment 1 (27%, $\chi^2(1) = 7.01, p < .01$).

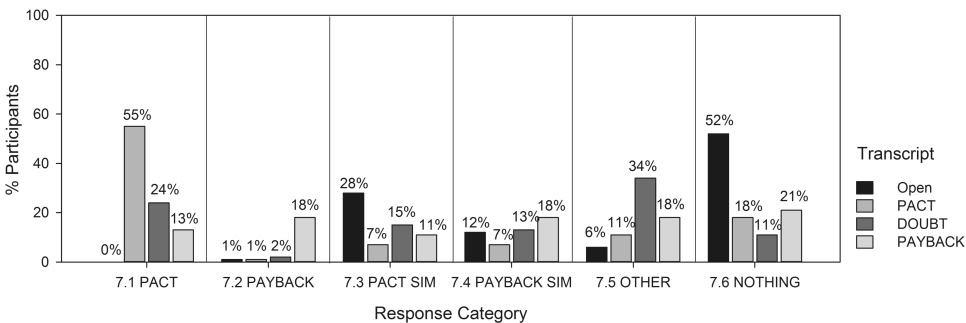


Figure 7. Percentage of responses in each category under each condition (Expt 2).

Having an expert cast doubt on the inaccurate PACT transcript, and listening to Phrase 3 in isolation, significantly lowered acceptance of PACT ($Q(1) = 18.8, p < .001$), but only to 24%—meaning that, despite the fictional expert’s caution, a quarter of participants still (erroneously) heard the exact ‘pact’ phrase in the ‘Doubt’ condition. Note that this is approximately the same proportion as heard PACT on first suggestion in Experiment 1, and very much higher than the 0% in either experiment who heard PACT in Open condition. In other words, the expert’s caution was ineffective in ‘resetting’ listeners’ perception.

When the alternative PAYBACK transcript was offered, the number accepting PACT again dropped significantly ($Q(1) = 7.36, p < .05$), but only to 13%. This is still very much higher than in Experiment 1, where, with no background story to push ‘pact’, all participants who had previously accepted PACT rejected it when the PAYBACK alternative was presented.

Finally, Panel 7.2 shows that in Experiment 2, relatively few participants heard PAYBACK before it was suggested, or even accepted PAYBACK when it was given as an alternative (18%, as opposed to 43% in Experiment 1 (Panel 5.1.2), $\chi^2(1) = 7.35, p < .01$)).

3.5.3 Disagreement with Suggestion and Alternative

As in Experiment 1, participants who disagreed with the transcripts as they were presented were invited to offer their own interpretations of what they heard. Panels 7.3–7.6 offer a breakdown of these other responses for Experiment 2, for comparison with results for Experiment 1 (Panels 5.1.3–5.1.6 and 5.2.3–5.2.6).

It will be noted that, overall, PACT SIM was a far more frequent response in Experiment 2 than in Experiment 1 (where it was only ever heard once). Its most frequent occurrence was in Open condition, where 28% of participants were led by the contextual priming to believe they heard the word ‘pact’ (even if they could identify no other words). PACT SIM responses then decreased significantly to 7% ($Q(1) = 15.2, p < .001$) when Transcript 1 was suggested—driven of course by the large numbers now accepting the exact ‘pact’ phrase—and remained at a similar level throughout the rest of the experiment ($Q(2) = 3.38, p > .05$).

PAYBACK SIM and OTHER responses were relatively low, not just in Open condition (see Section 3.5.1), but throughout Experiment 2—except for a significant increase in OTHER responses from 11% to 34% in the Doubt condition ($Q(1) = 17.2, p < .001$) followed by immediate decline in the Payback condition ($Q(1) = 11.3, p < .01$). It seems that participants who were influenced by the expert briefly cast about for their own interpretation, then accepted PAYBACK when it

was suggested. Again this differs markedly from Experiment 1, where, with no contextual priming, PAYBACK SIM and OTHER were the most frequent responses in Open condition, each heard by well over a third of participants, then dropped as suggested and alternative transcripts affected perception. This suggests that, in Experiment 2, the relatively open-minded PAYBACK SIM and OTHER responses were suppressed by the contextual priming by the ‘pact’ story.

That view is supported by consideration of NOTHING responses, which also show a reversal of the trend seen in Experiment 1. In Experiment 1, almost all participants offered some interpretation of Phrase 3 in Open condition, with the number offering NOTHING then gradually increasing to nearly half by the end of the experiment. In Experiment 2, by contrast, NOTHING was the most frequent response (52%) in Open condition, given by almost everyone except the 28% who heard PACT SIM. NOTHING then declined significantly ($Q(1) = 23.1, p < .001$) to 18% when PACT was suggested, and remained low for the rest of the experiment ($Q(2) = 4.33, p > .05$). In other words, knowing the story of the case inclined participants to keep trying to find an interpretation of this (uninterpretable) audio, rather than accept it as meaningless.

3.5.4 Guilt

At the start of Experiment 2, and again at the end, participants were asked whether they thought the father was guilty. At first, as shown in Figure 8, 80% said they had no opinion, significantly more than those saying either guilty or not guilty ($\chi^2(2) = 83.2, p < .001$). By the end, significantly fewer (40%) said they had no opinion (McNemar $\chi^2(1) = 25.0, p < .001$), while significantly more (29% and 30% respectively) were inclined to think either that he was guilty (McNemar $\chi^2(1) = 5.04, p < .05$), or that he was not guilty (McNemar $\chi^2(1) = 16.0, p < .001$). This meant that by the end, there was no significant difference between No Opinion, Guilty and Not Guilty responses ($\chi^2(2) = 1.78, p > .10$).

This effect is driven in part by a strong correlation between hearing the ‘pact’ phrase and finding the father guilty. Of the 13% who still heard the exact ‘pact’ phrase at the end of the experiment (despite doubt cast by an expert, presentation of the PAYBACK alternative, and an opportunity to listen to Phrase 3 in isolation), nearly three-quarters found the father guilty, significantly more than gave ‘no opinion’ or ‘not guilty’ responses ($\chi^2(1) = 11.59, p < .001$). Importantly, this effect cannot be accounted for by the reverse explanation—that those who already thought the father was guilty at the beginning were more likely to hear Phrase 3 as ‘at the start we made a pact’—since there is no correlation between ‘guilt at start’ and ‘PACT at end’ ($\chi^2(1) = 1.62, p = 0.20$), or indeed between ‘guilt at start’ and ‘guilt

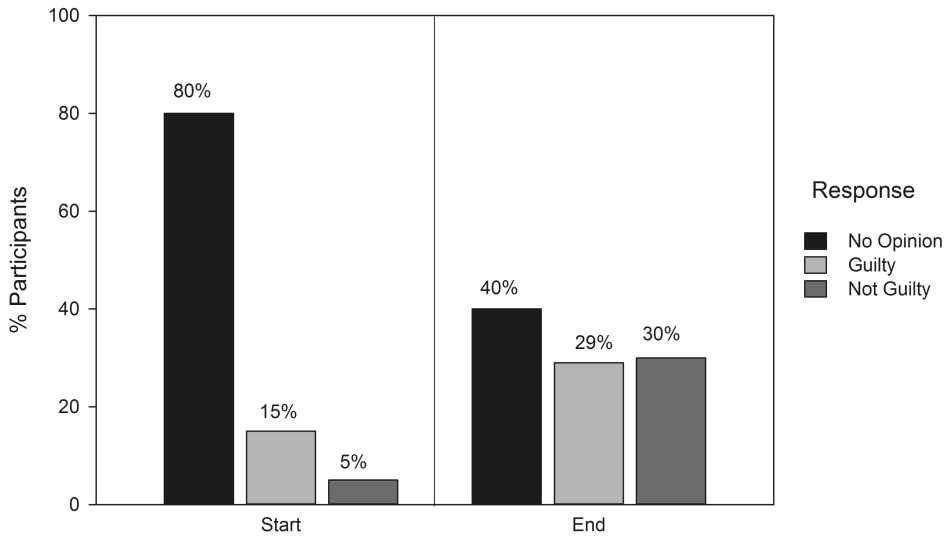


Figure 8. Responses concerning guilt, at start and end of Expt 2

at end' ($\chi^2(1) = 1.62, p = 0.20$). This suggests that participants were using their 'hearing' of *at the start we made a pact* as evidence for the story of the murder, not realising the extent to which their 'hearing' was itself a product of the story of the murder.

However, the effect of the audio on opinions about guilt went far deeper than that. Even participants who rejected the PACT prime were significantly more likely to find the father either guilty or not guilty than to express No Opinion ($\chi^2(2) = 20.3, p < .001$). In other words, listening to this indistinct 14-second extract from a whispered conversation (which is uninterpretable and thus incapable of demonstrating guilt or innocence), in the context of a story about a murder, tended to give participants an opinion as to whether the father was guilty or not guilty (of some poorly specified crime).

3.5.5 Demographics

No statistically significant correlations were found between any demographic characteristic measured and propensity to interpret the audio in a particular way (in either experiment), or to find the father guilty or not guilty (in Experiment 2). However, as in our previous transcription experiment,¹⁰ results given by the small number of Experiment 2 participants who identified as police are notable.

¹⁰ Ibid.

First, seven out of the eight police (88%, compared to 51% non-police) accepted PACT when it was first suggested, while two (25%, as opposed to 12% non-police) still heard PACT at the end—with both of these, and another two police, finding the father guilty. This means that, while all eight police started with ‘no opinion’ regarding guilt, by the end, 50% had moved to considering the father ‘guilty’, as compared with 19% of non-police moving from ‘no opinion’ to ‘guilty’.

Though numbers here are too small for reliable statistics, these findings would appear to be worthy of further exploration, especially in view of the role police transcripts of recordings like this are given in trials in Australia and elsewhere.

3.6 Discussion (Experiment 2)

Experiment 1 demonstrated, first, that the PACT transcript is not just inaccurate but implausible; and, secondly, that despite this, it has the power to prime participants’ perception.

Experiment 2 demonstrated that prior knowledge of a story about the audio dramatically increased the priming effect of the PACT transcript, and made it far less likely that participants would revise their interpretation when offered a more plausible transcript. Importantly, as in our previous experiments,¹¹ the transcript had more subtle effects on the perception even of those who did not fully accept its interpretation.

Further, Experiment 2 demonstrated that seeing a transcript of this uninterpretable audio encouraged participants to form an opinion as to the father’s guilt. Again, this effect was seen even in those who rejected the suggested transcript.

Experiment 2 also highlights problems with the concept that police who ‘listen many times’ can be deemed to be ‘ad hoc experts’ with respect to a covert recording. First, it makes clear that listening many times is no guarantee of hearing accurately. Participants listened to the excerpt numerous times under a range of conditions. However, they reached a wide range of different conclusions as to what was said—though of course each participant was aware only of their own.

Secondly, it suggests that knowing the context has more effect on perception than does listening many times. Participants in both experiments listened approximately equal numbers of times. However, by the end of Experiment 2 (with context) only 21% correctly concluded that Phrase 3 was uninterpretable,

¹¹ Ibid.

compared to 37% (Group 1) and 46% (Group 2) in Experiment 1 (without context). Further, demographic analysis suggested that police may be more likely to be influenced by context than other groups.¹²

These results suggest that, while transcripts by police, or others involved in a case,¹³ may have investigative value, they should not be accepted without evaluation. They also suggest that juries are not appropriate evaluators.

4. General discussion

Juries typically know the story of the case long before hearing the audio, putting them in a position more like that of Experiment 2 than Experiment 1. However, juries do not have the advantage of a special audio player allowing them to listen repeatedly to individual phrases under a range of conditions designed to induce scepticism about the suggested transcript. Nor do they have unlimited time. All participants in these experiments listened to the 14-second excerpt for at least 10 minutes—many for far longer. At that rate it would take around 20 hours to work through the entire 38-minute recording.

As we have seen, even with these advantages, Experiment 2 participants were unable to ‘reset’ their primed perception. This indicates that a caution from the judge is unlikely to be a sufficient safeguard against an inaccurate police transcript affecting a jury’s perception of what is said in a poor quality covert recording, and thus influencing their views regarding the defendant’s guilt.

Taken together, these findings highlight the importance of ensuring that transcripts are confirmed as reliable before being accepted for use as an aid to jurors’ perception of poor quality audio. This raises again the question of how it can happen that transcripts as manifestly inaccurate as the one used in this case (not an isolated incident) can be accepted as reliable evidence in the first place.

It is not only juries who know the story of the case before hearing covert recordings used as evidence. Prosecution and defence teams are well acquainted with the story, and with the issues the covert recording purports to resolve. They also typically know the contents of a transcript before hearing the audio. Under these conditions, our results suggest, they may be likely to ‘hear’ at least some of

12 Of course, there is no suggestion here of any kind of intentional bias on the part of police; priming happens below the level of consciousness.

13 The focus here has been on police, as they are the most commonly used ‘ad hoc experts’. However, the same arguments apply to anyone with background knowledge of a case, especially if they have an interest in the outcome.

the words suggested by the transcript. And this, the results further suggest, may be enough to encourage them to accept the transcript as a whole—especially when they are assured, by the law itself, that the transcript was produced by an ‘ad hoc expert’, and that the jury will be listening critically to the audio, using the transcript only as an aid.

5. Conclusion

Results of these experiments suggest several rather straightforward, and indeed common-sense, conclusions in relation to the presentation of hard-to-hear covert recordings as evidence in criminal trials.

Juries do need a transcript to help them interpret poor quality covert recordings. Without this, there is a high risk of their perception being unwittingly influenced by their knowledge and preconceptions about the case. What is essential is for the transcript to be reliable. Without that, there is a high risk that it will ‘help’ them to hear words that are not objectively there.

For these reasons, transcripts of indistinct covert recordings should be properly evaluated before being provided as an aid to perception. Such evaluation requires both independence and expertise in relevant branches of phonetic, forensic and cognitive science.¹⁴ Leaving it to the jury is not fair on jurors, let alone on the defendant.¹⁵

6. Appendix 1: Demographics

Note: our classification of occupation was not rigorous and should be taken as indicative only. Future research might attempt better demographic analysis of responses.

¹⁴ See French and Harrison, above n. 2.

¹⁵ For those interested, the father in this case was found guilty of murder and is currently serving a 24-year prison sentence.

		Experiment 1 (n = 56)		Experiment 2 (n = 82)
		Group 1 (n = 30)	Group 2 (n = 26)	
Sex	male	43%	27%	48%
	female	53%	62%	52%
Age	teen	3%	4%	1%
	20s	7%	4%	20%
	30s	17%	23%	28%
	40s	23%	4%	19%
	50s	30%	31%	20%
	60s	13%	19%	9%
	70s	3%	4%	4%
Education	postgraduate	50%	65%	50%
	diploma or certificate	7%	12%	18%
	undergraduate	37%	12%	28%
	school	3%	0%	4%
First language	English	83%	85%	87%
	non-English	17%	15%	13%
Hearing	very good	37%	27%	40%
	good	53%	58%	45%
	poor	10%	15%	15%
Headphones	professional	13%	19%	12%
	ordinary	40%	35%	56%
	none	43%	31%	32%
Occupation	academic (includes students)	27%	42%	37%
	legal	30%	12%	9%
	language or linguistics	0%	0%	10%
	other	40%	31%	45%