
Why “Ad Hoc Experts” Should Not Provide Transcripts of Indistinct Forensic Audio, and a Proposal for a Better Approach

Peter French and Helen Fraser*

Indistinct covert audio recordings frequently figure in criminal trials together with transcripts prepared by police officers who have been accorded the status of ad hoc experts on the basis of their prolonged and repeated exposure to the recordings. Drawing on research in linguistic and phonetic science, we explain why such transcripts are highly prone to be unreliable, why they may mislead juries into misinterpreting the contents of the conversations and why current court procedures for mitigating this risk are inadequate. We conclude by outlining a proposal drawn up and endorsed by senior expert linguists for establishing a process whereby reliable transcripts of indistinct covert recordings can be provided for juries.

I. INTRODUCTION: COVERT RECORDINGS AND AD HOC EXPERTS

Covert recordings, captured by “bugging” houses, cars, business premises, etc., are used during the investigation of very many major crimes, providing vital intelligence for police and security personnel. The present article is concerned not with this investigative use, but with the subset of recordings that go on to be used as evidence in Australian criminal trials.

Covert recordings can provide powerful evidence in court, allowing the jury to hear with their own ears criminal conspiracies and admissions that would not be made overtly. A major problem, however, is that, because of the conditions under which they are obtained, the audio is often of very poor quality, to the extent that the jury needs assistance to make out what is being said.

Since the 1987 High Court ruling in *Butera v Director of Public Prosecutions (Vic)*,¹ it has been common practice to provide this assistance in the form of a transcript, usually prepared by investigators from the case, who are given the status of “ad hoc experts” to facilitate admission of their transcripts as opinion evidence.

Of course, the risk is recognised that investigators’ transcripts might not be fully reliable. The law seeks to mitigate this risk via the expectation that the transcripts will be fully checked by defence, who will bring disagreements to the attention of prosecution. Where disputes over specific utterances cannot be resolved directly between the parties, opinions may be sought from experts. As further risk mitigation, the judge is required to direct the jury to the effect that the evidence is the recording, not the transcript. Therefore jurors should be sure to listen to the audio carefully, using the transcript(s) only as an aid, and reach their own opinion as to its contents.

Over the past 30 years, lawyers have become very familiar with these concepts and practices, which are commonly used, and seldom questioned. However, from the perspective of linguistic science, they are highly problematic, known to create actual and potential unfairness or injustice.² This article briefly

* Peter French: Professor of Forensic Speech Science, University of York; Chairman, JP French Associates, England; President, International Association of Forensic Phonetics and Acoustics (IAFPA). Helen Fraser: Adjunct Associate Professor, Linguistics, University of New England, Australia; Director, Forensic Phonetics Australia.

¹ *Butera v DPP (Vic)* (1987) 164 CLR 180; 30 A Crim R 417; see also: *R v Cassar* [1999] NSWSC 436; *Eastman v The Queen* (1997) 76 FCR 9; *R v O’Neill* [2001] VSCA 227.

² H Fraser, “Covert Recordings Used as Evidence in Criminal Trials: Concerns of Australian Linguists” (2018) 30 JOB 53.



explains the nature and cause of the problems, and indicates the direction towards a solution that is being recommended by Australian linguists.³

II. WHY TRANSCRIPTS BY AD HOC EXPERTS ARE USED

Investigators can often make out more than others can in poor quality covert recordings related to their cases. Indeed, with repeated listening, they may be able to transcribe a good deal of the recording. Furthermore, listening with the aid of their transcripts can help others to hear material that previously seemed unintelligible to them.

Observations like these have made it seem reasonable for the law to allow juries to benefit from the assistance of police transcripts, with precautions such as those mentioned above. Indeed, this concept was supported by some phonetics experts testifying in Australian courts during the 1990s. The first author was one of these, but two factors have changed his position. First, a much greater awareness has developed over the past 20 years⁴ of the ways in which background knowledge of and/or involvement in an investigation can give rise to expectations of what *might* be said in a covertly recorded conversation. Where the audio is unclear, these expectations may influence a police transcriber’s perception of what, in fact, was said and lead to errors in transcripts, some of which are unlikely to be picked up by listeners (see Part IV). Second, he has repeatedly had direct experience of cases in which this has actually happened, that is, where police transcripts containing highly incriminating but demonstrably incorrect content have been proffered to the courts.

The use of police transcripts of indistinct audio is now opposed by most informed experts in forensic phonetics and linguistics, locally and internationally. Why has this view not extended to lawyers and judges? There are two main reasons for this. First, detective transcribers’ ability has been explained in terms of their having a form of specialised knowledge, in the sense of *Evidence Act 1995 (NSW) s 79*, derived from listening to the recording “many times”. Second, listeners’ experience of being assisted by a police transcript has been taken as an indication that the transcript is generally accurate. Both of these explanations are incorrect from a scientific perspective.⁵ The next section attempts a very brief overview of some relevant findings of linguistic science regarding the nature of speech and speech perception.

III. THE REAL REASON POLICE CAN MAKE OUT MORE THAN OTHERS IN COVERT RECORDINGS

Listening to speech in everyday circumstances gives the strong impression that it is a sequence of discrete words, each made up of a sequence of sounds (or “phonemes”) articulated with varying degrees of clarity. On this view, perception of speech involves recognising the sounds, putting them together into words, and then interpreting the meaning of the words in context. However, longstanding research in linguistics (the science of language), and phonetics (the branch of linguistics dealing with speech) has shown that this common view is incorrect.

In fact, speech is a continuous stream of sound.⁶ The discrete words and sounds are abstractions created by unconscious cognitive processes. Most importantly, the information provided by the continuous stream of sound is very often, in itself, insufficient to enable listeners to perceive the words and sounds intended by the speaker (in technical terms: the signal under-determines the percept).

³ The term “linguist” refers to a specialist in linguistics, the science of language, one of whose major branches is phonetics, the science of speech.

⁴ H Fraser, “Issues in Transcription: Factors Affecting the Reliability of Transcripts as Evidence in Legal Cases” (2003) 10 IJSL 203; CS Fishman, “Recordings, Transcripts, and Translations as Evidence” (2006) 81 *Washington Law Review* 473; H Fraser and B Stevenson, “The Power and Persistence of Contextual Priming: More Risks in Using Police Transcripts to Aid Jurors’ Perception of Poor Quality Covert Recordings” (2014) 18 IJEP 205.

⁵ H Fraser, “Assisting’ Listeners to Hear Words that Aren’t There: Dangers in Using Police Transcripts of Indistinct Covert Recordings” (2018) 50 AJFS 129.

⁶ P Ladefoged and SF Disner, *Vowels and Consonants: An Introduction to the Sounds of Language* (Wiley-Blackwell, 3rd ed, 2012).

This means that listeners depend greatly on additional information supplied by their knowledge of the context in which the speech occurs.⁷ This is a major reason that computer speech recognition took so long to develop to a level of practical functionality. Recent advances rest on new methods capable of incorporating contextual information into the computer processing, though even now capabilities are very limited in comparison to human speech perception.

Without contextual information, unmonitored conversational speech can be difficult to follow even in a good quality recording. On the other hand, with appropriate contextual information, a poor quality recording can seem relatively clear. A mere hint as to the content of the recording can create expectations that powerfully “prime” listeners’ perception. However, explicit suggestion as to specific words that are likely to be heard (as in a transcript) has a far stronger priming effect.

This means that “indistinct” audio may be best defined, not in terms of its being “hard to hear” (since appropriately primed listeners may hear it perfectly easily and confidently), but rather in terms of its content being perceived differently by listeners primed with different expectations.⁸

The interesting thing is that listeners are typically unaware of the role played by their contextual expectations, ascribing their perception wholly to the sounds they (erroneously) feel are objectively “there to be heard”. This unawareness makes it surprisingly easy, even with good quality audio but especially with poor quality audio, to manipulate listeners’ perception simply by priming them with different expectations as to what they are likely to hear.

This susceptibility of perception to manipulation is sometimes thought of as a kind of bias. However, the term “priming” is preferable as it gives a less negative, and more realistic, impression of a process that helpfully gets the listener ready to hear what is coming next. In the vast majority of cases, listeners’ contextual expectations are reasonable, and priming helps them hear speech efficiently and accurately.

The problem is that, while priming with reliable contextual expectations can be extremely helpful, priming with unreliable contextual expectations can be powerfully misleading. Worse, there is nothing in listeners’ experience to alert them as to which condition applies in any given case. In particular, listeners’ personal confidence in their perception is known to be a very poor indicator of reliability. Many decades of research have demonstrated that listeners’ confidence correlates poorly with their accuracy. In other words, listeners can be highly confident but completely wrong. The only way to be sure of the reliability of their perception is by reference to the “ground truth” as to what was actually said – which of course is the very thing that is lacking in a forensic situation.

For these reasons and more, reliable transcription of poor quality audio is a highly specialised task requiring independence, professional expertise, and, above all, careful management of priming – not by withholding all contextual information, but by ensuring transcribers start with minimal priming, and revealing relevant, reliable information according to a process that is nowadays described under the heading of “sequential unmasking”.⁹ It involves, after the initial unprimed examinations have been completed, iterative disclosure of relevant reliable external and background facts through a series of controlled and carefully documented stages over which the transcript is revised and finalised.

IV. WHY USE OF TRANSCRIPTS BY AD HOC EXPERTS IS PROBLEMATIC

A. Police Transcripts Are Likely to Be Unreliable

With the brief background provided in Part III, it is easy to see that the reason investigators can make out more than others can in recordings related to their cases is not because they have listened “many times”, but because they are primed by contextual expectations deriving from their familiarity with the case and

⁷ D Byrd and TH Mintz, *Discovering Speech, Words and Mind* (Wiley-Blackwell, 2010).

⁸ H Fraser, “Transcription of Indistinct Covert Recordings Used as Evidence in Criminal Trials” in H Selby and I Freckelton (eds), *Expert Evidence* (Thomson Reuters, 2015).

⁹ I Dror et al, “Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making” (2015) 60 *Journal of Forensic Sciences* 1111.

its circumstances. Indeed it is important to acknowledge that, to the extent these contextual expectations are reliable, they can potentially help police to make out specific utterances reliably. For this reason, police should have a role in creating transcripts of indistinct forensic audio. However, they should not have responsibility for the transcription process.

The reason is that police expectations about the content of indistinct covert recordings are far from neutral. This is by no means to suggest that investigators are inappropriately biased. It is merely to recognise that the familiarity with the case that generates the expectations that prime detectives’ perception may include aspects that would not be independently admissible as evidence in court. Since a transcript of indistinct audio is inevitably influenced by the contextual knowledge and assumptions of the transcriber, these aspects can indirectly influence the jury in uncontrollable ways.

B. Listeners Are Unlikely to Notice and Correct All Relevant Errors

The practices intended by the law to mitigate the jury risk of an unreliable police transcript (Part I) are ineffective. The reason is that listeners are strongly but unwittingly primed by the transcript itself. Longstanding research shows that a transcript exerts a strong influence on perception of poor quality recordings, even when that transcript is demonstrably inaccurate, misleading or downright implausible. Most troubling of all are demonstrations that even listeners who reject all or part of a transcript may still be perceptually influenced by it.

This means that checking of police transcripts by the defence, or even by the judge, is unlikely to result in full recognition of the errors it contains. Finally, judicial instructions to the jury to listen carefully using the transcript only as an aid are unrealistic. With a misleading transcript, jurors may earnestly follow the instructions and yet be unwittingly led to a confident but inaccurate interpretation of the content of the audio, which may in turn have a powerful influence on their interpretation of other evidence presented in the trial.¹⁰

V. THE EXTENT OF THE PROBLEM

There are many known cases of police mis-transcribing utterances in poor quality forensic audio. A few colourful examples have become famous in forensic linguistics. For instance, a police transcript “he died after wank off” was shown by expert analysis to be “he died after one cough”;¹¹ “shot a man to kill” was shown to be “showed the man the ticket”.¹²

The reason errors like these are known is that they have been picked up as “disputed utterances” before being admitted as evidence, and the dispute has been resolved by experts in the branch of forensic phonetics known as “disputed utterance resolution”.¹³ These examples might therefore be interpreted as demonstrating that the current risk-mitigation process (Part I) works in the way intended: starting with the “ad hoc” expert’s version, but subjecting it to analysis by a genuine expert if a dispute arises that cannot be resolved between the parties.

However, this view is belied by the fact that there are also known cases of erroneous transcripts passing through multiple levels of checking without dispute. While some of these are caught by the legal process before going to the jury,¹⁴ there are also known cases of thoroughly misleading transcripts being provided

¹⁰ JB Gould et al, *Predicting Erroneous Convictions: A Social Science Approach to Miscarriages of Justice* (National Institute of Justice, Washington DC, 2012).

¹¹ P French and P Harrison, “Investigative and Evidential Applications of Forensic Speech Science” in A Heaton-Armstrong (ed), *Witness Testimony: Psychological, Investigative and Evidential Perspectives* (2006).

¹² M Coulthard, A Johnson and D Wright, *An Introduction to Forensic Linguistics: Language in Evidence* (Routledge, London/New York, 2nd ed, 2017).

¹³ P French, “Analytic Procedures for the Determination of Disputed Utterances” in H Kniffke (ed), *Texte zu Theorie und Praxis forensischer Linguistik* (Niemeyer, Tübingen, 1990).

¹⁴ *R v Dunn* (2012) 15 DCLR(NSW) 144; [2012] NSWDC 192; *R v Vandergulik (No1)* [2008] VSC 407; *R v Hall* [2001] NSWSC 827; *R v Nguon* (2014) 22 DCLR(NSW) 302; [2014] NSWDC 385.

as “assistance” to the jury.¹⁵ While we can never know for sure what a jury makes of evidence, there is no reason to think their perception would not have been misled.

Importantly, while these known cases are useful as examples to demonstrate the possibility of failure in the current risk-mitigation process, they are undoubtedly just the tip of a far larger problem. The current system is not merely subject, like any system, to occasional failures. It is, by its nature, unlikely to deliver a reliable outcome.

This is because the problem is one not just of practice but of principle. Genuine experts agree that “ad hoc expert” is not a useful concept in relation to transcription.¹⁶ Since police transcribers are likely to create unreliable transcripts, and checking the transcripts by lawyers and judges is unlikely to reveal their unreliability, juries are liable to be misled as to the content of covert recordings.

VI. TOWARDS A SOLUTION

Unfortunately, some apparently obvious solutions are inadequate from the perspective of linguistic science. In particular, providing poor quality audio to the jury without any transcript creates substantial problems of its own, in that it is likely that they would make little or nothing of it. Or, worse, while listening under less than ideal courtroom conditions and without the benefit of training, they might well arrive at an overly “rich” or imaginative interpretation of what was said. Juries *do* need the assistance of a transcript in order to ascertain the content of indistinct audio, but the transcript must be reliable.

Another solution that may come to mind consists of continuing to allow police to prepare transcripts but encouraging defence lawyers to make greater use of phonetics experts to check their accuracy. One major problem with this is that in the Australian context there are few genuine specialists in forensic transcription, leaving a vacuum ready to be filled by self-styled experts who may offer less qualified opinions. More importantly, the opinion of even the best expert consulted by the defence is typically put to the court only as a competing alternative to the police version. This is especially problematic where the expert opinion is that relevant parts of the audio are simply too poor for reliable transcription. By the time the expert opinion is given, everyone in court has been primed by the police version, making it difficult or impossible to reverse its influence even if it is demonstrably wrong.

These problems with police transcripts can perhaps be seen by considering that a very poor quality recording is equivalent to a smudged fingerprint. Where a fingerprint is so smudged that experts find the evidence it contains to be inconclusive, we do not invite detectives to offer their personal opinion as to whose print it is, and encourage the jury to make up their own minds. Yet this is what can be done with unintelligible covert recordings.

So, what is the way forward? A full solution to the problems outlined here requires establishing a better interface between the law and genuine expertise in linguistic science so that an evidence-based process can be initiated for producing reliable transcripts of indistinct covert recordings before they enter the trial process. Australian linguists are currently seeking to collaborate with the judiciary to institute such a system, and all the benefits it entails. The specific proposal is that a collaborative consultation should take place towards establishing a system whereby teams of transcribers who are employed in public/police service, but who are independent of, and isolated from, police investigations, can be trained by experts in the skills necessary for interpreting and representing indistinct audio and in the risks inherent in the process, as demonstrated in the research in linguistic science. A document setting out this proposal and signed by 16 prominent linguists is currently under discussion with the judiciary.

¹⁵ *R v Murrell* (2001) 123 A Crim R 54; [2001] NSWCCA 179; H Fraser, “How Interpretation of Indistinct Covert Recordings Can Lead to Wrongful Conviction: A Case Study and Recommendations for Reform” in R Levy et al (eds), *New Directions for Law in Australia: Essays in Contemporary Law Reform* (ANU Press, Canberra, 2017).

¹⁶ For more general problems with the concept of “ad hoc expert”, see G Edmond and M San Roque, “Quasi-Justice: Ad Hoc Expertise and Identification Evidence” (2009) 33 CLL 8.