HELEN FRASER

Forensic Transcription: Legal and scientific perspectives

Audio recordings are often used as forensic evidence in criminal trials. Unfortunately, they are often of very poor quality, meaning the court needs a transcript to be sure of their content. Many jurisdictions allow transcripts to be provided by police. This creates problems that can result in substantial injustice. Phonetic science is needed, but how can it best assist? Many recommend that transcripts should be produced, or evaluated, by experts in acoustic-phonetic analysis. However, this does not necessarily solve all the problems. The present paper argues that this is because forensic transcription is significantly different from established forms of phonetic analysis, and requires not just applying existing knowledge, but developing new knowledge, with a broader view of the evidence needed to ensure a transcript of indistinct audio is reliable.

Keywords: transcription, perception, forensic, evidence, acoustic.

1. *Introduction*

Forensic transcription is the science and practice of transcribing forensic audio. Forensic audio is recorded speech used as evidence in a criminal trial. It comes in various forms, but the most common is a covert recording – conversation captured secretly, typically via a hidden listening device legally deployed on behalf of police. Covert recordings can provide powerful evidence in a trial, allowing the court to hear speakers making admissions they would not be prepared to make openly. A major problem, however, is that the need for secrecy makes it very difficult to control the recording conditions. As a result, the audio is often 'indistinct' (an informal term used by lawyers to describe audio affected by factors such as overlapping speech, variable microphone distance, background noise, line interference, etc).

Before reading on, readers might like to access two short examples of real forensic audio, which will be discussed throughout this paper. These are available at forensictranscription.net.au/audio: the 4-second sample under 'Interpretation of a crisis call'; and the 14-second sample under 'The pact experiments' (bottom of page). These examples demonstrate the problem with indistinct forensic audio: most listeners find them unintelligible without assistance. Contextual information sometimes helps. The second sample above, for example, comes from a murder trial in which the outcome hinged on the nature of a pact between the speaker (the defendant in the current trial) and a murderer (already convicted in a previous trial). If the pact was an agreement to commit the murder jointly, the defendant was an 'accessory before the fact', equally guilty of murder. However, if the pact was an

agreement to conceal the murder, the defendant was an 'accessory after the fact', a serious crime but not nearly so serious as murder.

This and other contextual information enabled a transcript to be produced which assists many listeners to hear the words 'at the start we made a pact' (Fraser, 2018). Indeed, the jury appears to have heard these words, and taken them to mean that the pact was made before the murder, as they returned a verdict that the defendant was guilty of murder, and he was sentenced to thirty years in prison. The problem is that, in this case as in others, the transcript was wrong, raising the serious possibility that the verdict may also have been wrong.

The present paper starts by examining the problem and showing that it rests ultimately in deep-seated misconceptions in the law about the nature of speech and the processes involved in its perception and transcription. It then turns to solutions and considers how phonetic science can help create a better process. A key argument is that forensic transcription requires more than just providing acoustic evidence to support or refute a suggested transcript of indistinct audio. Solving the problems effectively needs a broader evidence-based process, that requires phonetic science not just to apply existing knowledge but to develop new knowledge.

The paper is based on a plenary presentation summarising a series of previous publications (see references) which provide extensive background on all the points discussed. While the focus is on the Australian trial process, which involves a jury in an adversarial system similar to that used in the United Kingdom, some of the discussion may be relevant in other jurisdictions.

2. *Problems with forensic transcription*

2.1 Transcripts are provided by police

The 'pact' example above shows the value of having a transcript of indistinct forensic audio. Now we must consider how the transcript is created and evaluated. In the 'pact' trial, as in many others, the transcript was provided by a detective investigating the case. This is often found surprising by outsiders, but it is long-established practice in the law, justified via a number of concepts which I gradually came to understand over a decade of casework experience, summarised briefly here (for a detailed account see Fraser, 2020b).

I first became aware of police transcripts via a case in the late 1990s. I was asked by the defence to transcribe an extremely indistinct recording. The audio was of such poor quality that I had to hand in a transcript with many gaps. I was then asked to review an existing transcript that showed several utterances containing the word 'heroin'. When I checked the relevant sections of audio, I found no phonetic evidence at all for the word 'heroin'. My evidence to this effect helped obtain a 'not guilty' verdict. The defence were pleased, but I was troubled. I had only shown that the word 'heroin' had not been spoken. This did not mean the speakers were not discussing drugs. I felt my evidence had been used as a 'gotcha' to undermine the

prosecution case – and wondered who had provided them with a transcript so bad that it left them open to this kind of opposition.

This was when I first learned that transcripts of indistinct audio are usually produced by police investigating the case. I was surprised, as it seemed evident to me that police transcripts might not be fully reliable. However, when I asked 'Why would you use police to produce a transcript?', the answer came quickly: 'Why wouldn't you use police – they are the ones who can hear what is said?'. Indeed, it is true that investigators can often make out more of the content than others can. According to the law, this ability stems from their having listened to the audio 'many times', giving them the status of 'ad hoc expert' (French, Fraser, 2018; Fraser, 2021).

Of course, from the perspective of phonetic science, listening many times is not the real reason for investigators' apparent ability, which actually stems from their access to contextual information about the case. In an effort to counter this legal misconception, I published an article explaining the concept of *contextual priming* (Fraser, 2003). Contextual priming is the phenomenon whereby listeners with relevant background information may be able to interpret indistinct audio that is unintelligible to listeners who do not know its context.

The problem is that, while contextual information can be very helpful, it is a double-edged sword. Priming with *reliable* contextual information can sometimes help listeners hear accurately (though note that the common inference that this means those with reliable contextual information automatically hear accurately is certainly not true). Importantly, however, priming with *unreliable* contextual information can easily cause listeners to hear confidently but inaccurately. Of course, not all contextual information available to police can be confirmed as reliable (testing the reliability of that information is one function of the trial process). This (combined with officers' lack of training in transcription) means that police transcripts are often inaccurate to some degree – though of course that should not be taken to suggest they have a deliberate intention to mislead, as priming occurs without conscious awareness.

From a legal perspective, lawyers explained to me, this linguistic background was interesting but not at all troubling. They assured me that the law fully understands that police transcripts might contain errors. For this reason, the judge is obliged to instruct the jury carefully that the evidence is not the transcript, but the audio: they should listen carefully and reach their own opinion, using the transcript only as assistance.

Again, from the perspective of phonetic science this is unrealistic. With indistinct audio, a transcript does much more than 'assist' listeners' perception. It provides *textual priming* that strongly influences their perception in a lasting way. And again, while this can be beneficial if the transcript is reliable, it can be highly misleading if the transcript is unreliable (for a quick, accessible introduction to textual priming, see Burridge, 2017).

It is worth pausing to note that priming is not the same as bias. One difference is that priming cannot be managed simply by withholding the priming information: as we have seen, without priming, indistinct audio is often unintelligible. Another is that priming affects everyone and cannot be controlled by an effort of will. This

is shown by the common experience of mishearing song lyrics. English comedian Peter Kay is a master of inducing hilarious mishearings of pop songs, simply by playing a lyric (e.g., 'just let me state for the record') with a misleading suggestion ('just let me staple the vicar'). Importantly, the basis of the humour is that listeners hear the ridiculous words, even though they know they can't possibly be true.

2.2 Transcripts inevitably influence juries' perception, even if inaccurate

Around this time (2009) the 'crisis call' sample (referred to in the Introduction above) came into the public domain after being discussed in a murder trial. This enabled an experiment that provided a dramatic demonstration of the phenomenon of textual priming, in a way that was directly relevant to the law (Fraser, Stevenson & Marks, 2011).

At first, none of the 190 participants heard the incriminating phrase 'I shot the prick'. However, after it was suggested, about a third of them heard this exact phrase, with others clearly influenced by it. Further, about half of those who heard the phrase did not change their mind even after being told that experts on both sides agreed that the phrase was inaccurate. Importantly, these participants were more likely to give a 'verdict' that the speaker was guilty.

To linguists, this seemed like compelling evidence that a judicial instruction that the jury should use the transcript 'only as assistance' is unrealistic. Lawyers, however, remained unmoved. They explained that the law understood that juries could be 'suggestible': that was why, should the defence raise any doubt about a police transcript, the judge would listen personally to ensure that potentially misleading errors were corrected before it was provided to the jury.

Once more, this shows a serious misconception about speech and its perception (recently further confirmed by Fraser, Kinoshita, 2021).

The law operates on the principle that careful, responsible listeners like lawyers, and especially judges, are immune to the influence of an inaccurate transcript. However, that is not correct. Priming does not affect only 'suggestible' listeners. It is a necessary and unavoidable feature of human speech perception. Without meaning any disrespect to judges, from the point of view of phonetic science, they are no less likely than anyone else to be influenced by the powerful textual priming of an inaccurate transcript. It seemed to me it was only a matter of time until legal procedures based on these erroneous concepts created substantial injustice. And indeed I did not have long to wait.

2.3 Inaccurate transcripts influence lawyers and judges too

The 'pact' case (mentioned above) came to me in 2011 (for a full account of this case and the issues it raises, see Fraser, 2018). I reviewed the audio and transcript and readily demonstrated that the phrase 'at the start we made a pact' (which had been crucial in achieving the guilty verdict) was not only inaccurate, but implausible. It was not supported at all by either the segmental or the suprasegmental characteristics of the extremely indistinct whispered utterance. Nevertheless, it had passed all the

careful checks by the defence and the judge that lawyers had assured me mitigated any risks associated with providing police transcripts as assistance to juries.

This gave an opportunity to use the 'pact' audio to provide a dramatic and relevant demonstration of the flaws in the legal concepts outlined above. A new experiment (Fraser, Stevenson, 2014) showed, first, that in the absence of contextual information, no one heard anything remotely like the alleged phrase before it was suggested, and that, even after it was suggested, it had a weak, though significant, priming effect. This confirms the implausibility of the police transcript.

Importantly, however, the second part of the experiment showed that when the audio was played in the context of a story similar to that of the actual trial, the transcript had a powerful priming effect, with a majority of participants accepting the (inaccurate) incriminating phrase 'at the start we made a pact'. Further, many were clearly influenced by the phrase when they gave their 'verdict' regarding the speaker's guilt – seemingly unaware of the powerfully circular manner in which their understanding of the story had influenced their acceptance of the transcript, and then their acceptance of the transcript confirmed their understanding of the story.

Of course, in a trial everyone inevitably knows the contextual 'story'. Reviewing the pact trial, for example, it was clear that the lawyers and judge had been strongly influenced by the inaccurate police transcript – while nevertheless believing they were simply hearing words that were objectively 'there to be heard'. This is a well-known phenomenon. Though phonetic science has known for many decades that perception is not a simple 'bottom-up' process of recognising phonemes and putting them together to form words (Fraser, Loakes, 2020), these findings have not yet permeated the 'educated common knowledge' upon which the law is based (Fraser, 2018). The false belief that careful, responsible listeners hear 'what is there to be heard' exposes judges and others to being unwittingly misled by inaccurate transcripts – which they then allow to 'assist' the jury.

2.4 Potential for serious injustice

The main intention of the 'pact' experiments had been to raise a general concern that legal procedures for protecting juries from misleading transcripts were ineffective. However, I could not ignore the effect the transcript had had on the trial itself. As explained earlier, the outcome hinged on the nature of the pact (to commit murder or conceal murder). However, while the trial presented a variety of circumstantial evidence to suggest it was a pact to commit murder, the only 'direct' evidence that there had been any pact at all was the utterance 'at the start we made a pact' – which I had now shown was never actually spoken.

Normally this would have been grounds for appeal against the guilty verdict, but by the time I was consulted on the case, the trial was long over and all opportunities for appeal had been used (unsuccessfully) on other issues. The only option was an application to review the conviction. To assist the defence in making this application, I provided a detailed report, demonstrating that the transcript was certainly inaccurate – but nevertheless highly likely to have influenced the jury.

The application was rejected. This was expected – acceptance of such applications is extremely rare in Australia. However, the reasons for the rejection were concerning. The response did not engage with my arguments at all (for details, see Fraser, 2018). It simply asserted that the trial had been conducted in compliance with all legal requirements. This was true. The trial judge had (i) checked that the police transcriber had listened many times, (ii) sought the views of the defence (who, though objecting to the transcript, had not been able to provide a more plausible alternative), (iii) listened personally to be sure the transcript had no potential to mislead, and, most importantly, (iv) instructed the jury that the evidence was the audio and they should use the transcript only as assistance.

For these reasons, the rejection concluded, my evidence would have made no difference to the verdict. At best it might have given the jury another opinion to consider, along with the detective's. But this would not have changed their interpretation, since the fact that all legal procedures had been followed properly ensured it had been fully open to them to reach their own conclusion regarding the content of the audio.

The latter point, arguably, was true: it was open to the jury to reach their *own* opinion regarding the content of the audio. However, it really was not open to them to reach an *accurate* opinion regarding the content of the audio, as my report had shown.

It was also true, certainly, that all legal procedures had been followed properly. What the rejection missed was the significance of this fact. If a demonstrably misleading transcript had been allowed to 'assist' a jury *despite all legal procedures having been followed properly* – surely there must be some problem with the legal procedures.

3. *The turning point*

The rejection of the application to review the 'pact' conviction, by making explicit the misconceptions that had been concerning me for several years, created a turning point in my thinking about forensic transcription.

Up until this stage, I, like other phoneticians, had been recommending that, before being given to juries, police transcripts should be evaluated by experts in phonetic science. However, experiences in other cases had caused me to change that recommendation. The misconceptions explicitly expressed in the rejection confirmed my growing sense that the problem arose not directly from the fact that police provide the transcripts (though that is not good), but from the fact that legal procedures are founded on substantial misconceptions, embedded in the law, about the nature of speech and its perception and transcription. On that view, simply insisting on involving an expert will not solve the problem.

For one thing, the courts are not always good at recognising appropriate expertise. In Australia, there are few well-qualified experts in phonetic science, and a considerable proportion of those choose not to do case work. This leaves a vacuum readily filled by those with insufficient qualifications, or qualifications in subjects that seem to lawyers to be similar to phonetic science but are actually very different (e.g., dialect coaches, speech pathologists, scholars of specific languages, and audio engineers).

More importantly, legal procedures mean that even a well-qualified expert's advice is not always used optimally (for more, see Fraser, 2021). In fact, I had to concede that the rejection's conclusion that my evidence would have made no difference to the 'pact' verdict was probably right. Not just police transcripts but also expert transcripts are evaluated by lawyers and judges, and ultimately by the jury. This means a police transcript may still be provided to 'assist' the jury even if a well-qualified expert *on their own side* has shown it to be wrong. In fact, police transcripts are routinely privileged over an expert's (see discussion in Fraser, 2021).

For these and other reasons (see Fraser, 2020b), it seemed that rather than opposing police transcripts in individual cases, a better role for experts in phonetic science was to help the law devise better procedures, capable of ensuring that juries are always and only assisted by demonstrably reliable transcripts.

Putting these issues together with a range of other problems that had emerged regarding the legal handling of covert recordings (notably concerning possibilities for improving forensic audio via 'enhancing' – see Fraser, 2019) prompted a Call to Action. This was a 2017 letter endorsed by all four Australian linguistics associations and sent to the Council of Chief Justices, seeking review and reform of the legal handling of covert recordings used as evidence in criminal trials. In 2019, a judicial working party met with a group of linguists, and representatives from police and public prosecution departments across the country. After hearing extended argument and discussion, the judges acknowledged that the linguists' concerns were worthy of investigation – and the following year the University of Melbourne established the Research Hub for Language in Forensic Evidence (Fraser, 2020c).

The Hub has two main goals. The first is to work with the judiciary and appropriate law reform bodies to prevent the routine admission of police transcripts that is currently allowed, and develop procedures for presentation of reliable transcripts. The second goal, and the focus of the present paper, is to develop evidence-based methods that enable demonstrably reliable transcripts to be provided to the court right from the start of a trial. This is clearly a task for phonetic science. However, it requires recognition of some special characteristics that make forensic transcription very different, in important and interesting ways, from 'normal' phonetic analysis (Fraser, 2020b).

4. *How forensic transcription is different*

4.1 Unknown content

The most obvious feature of indistinct forensic audio is its 'indistinctness' – i.e., the fact that it is hard to understand when heard 'cold' (i.e., with no contextual or textual priming). However, 'indistinct' is a relative description. A recording that is completely unintelligible to those who do not know the context and/or the content can seem quite clear to those who do (Fraser, 2020a; Lange, Thomas, Dana & Dawes, 2011). This indeed is the very principle by which a transcript assists perception. The problem is, as outlined above, that a transcript can 'assist' listeners

even if it is inaccurate. Providing a reliable transcript requires knowing the content reliably – and of course the whole point of forensic transcription is that the content is not known reliably (as least not to those in authority). That is the very reason that the law asks the jury to determine the content, with the assistance of the transcript.

4.2 Insufficient internal context

It is well known in phonetic science that almost all recorded conversational speech is 'indistinct' in the sense that individual words and phonemes cannot be determined purely 'bottom-up', i.e., from acoustic information only, with no reference to context (for a full discussion with many references, see Fraser, Loakes, 2020). The reason listeners do not usually notice the indistinctness of words and segments is that their contextual knowledge gives them the top-down information they need to hear the words and phonemes with confidence.

However, while this is well known, it is easy to lose sight of its significance in discussing forensic transcription. In the presentation on which this paper is based, I provided a demonstration by playing three tokens of the word 'year' excised from a longer utterance. The audience could not guess what the words were, nor identify any of their phonemes, nor even categorise them as 'the same word' – and a spectrogram offered no help. I then played the full utterance in which the three tokens were embedded. They were immediately and unambiguously recognisable as three repetitions of the 'same' word 'year', pronounced differently due to their different contextual positions (after the words 'first', 'second' and 'third', respectively; and in syntactic positions that created different intonation). Taken as a whole, the audio was a fair quality recording of relatively clear speech. It was only because the words had first been played out of context that it was possible to observe that each *on its own* was totally unintelligible.

This shows the powerful but unnoticed role of contextual priming in determining content, even for experts. With a relatively clear recording, such as the one used for the demonstration, internal context (surrounding words heard within the recording) is usually sufficient. However, with forensic audio, internal context is often unavailable or insufficient (due to indistinctness). In these cases, perception must rely heavily on external context (listeners' knowledge or assumptions about the circumstances in which the recording was made).

4.3 Uncertain external context

To see the crucial role of external context, recall the experience of mis-hearing song lyrics. It is important to recognise exactly what it is that creates the humour in Peter Kay's examples (mentioned above). It is not merely the fact that he suggests ridiculous words. Suggesting any old ridiculous words would not be funny at all. What causes the hilarity is the fact that Peter Kay's carefully chosen suggestions cause our ears to hear the ridiculous words *even though we know for certain that they can't possibly be right*.

Crucially, our certainty that the words can't possibly be right comes *not from the audio itself, but from external information* – we may know the true lyric (the

content), or if we don't, we know that a romantic pop song is unlikely to contain a line like 'just let me staple the vicar' (the context). So, Peter Kay's humour shows not one, but two key things: first, the power of an inaccurate transcript to induce erroneous perception; and second, the power of reliable external information to override erroneous perception.

Unfortunately, this kind of reliable external information is precisely what is lacking with forensic audio (in fact, as discussed, the audio is typically being used to *establish* the external context). This means there is no corrective for inaccurate perception induced by a misleading transcript. When listeners are given a suggestion like 'I shot the prick' or 'at the start we made a pact', for example, their ears seem to 'hear' those words in just the same way as Peter Kay's audiences seem to hear 'just let me staple the vicar'. However, far from considering the suggestion ridiculous, they accept it so confidently that they use it as the basis from which to evaluate other evidence. This kind of circular reasoning is a serious but unacknowledged problem in trials that use indistinct covert recordings as evidence, as seen in the 'pact' example discussed earlier.

For these reasons and more, it is essential for forensic transcripts to be produced via an accountable, evidence-based method. This is widely assumed in phonetic science to mean providing evidence based on acoustic-phonetic analysis. However, while there is certainly a role for acoustic-phonetic analysis, it is not always enough in itself. In fact, the next section argues that the forensic situation has a number of characteristics that make it significantly different from other situations in which phoneticians analyse speech recordings. Fulfilling its needs requires phonetic science to develop evidence-based methods for creating reliable transcripts that take a broader view of the kind of 'evidence' that is needed to ensure a transcript is reliable.

5. *A new task for phonetic science*

The crucial characteristic of forensic audio, as we have seen, is not that it is indistinct when heard out of context, but that neither the content nor the context is known with certainty. This makes forensic transcription very different from transcription done by phonetic scientists in 'normal' situations.

In most research situations, even if the audio is indistinct, the content is known to researchers. They might have created or chosen specific material to be recorded, so as to test a scientific hypothesis. Or, if the recording is of free-flowing conversation with indistinct sections – the context is known with sufficient certainty to allow the content to be determined reliably.

Acoustic analysis of audio with known content over the past 70 years or more has been extremely valuable, allowing phonetic science to establish a great deal of theoretical knowledge about the nature of speech and how speech perception works – notably the massive variability of speech at all levels of description, and the role that textual and contextual priming play in assisting listeners to understand indistinct audio, as discussed above.

However, developing this theoretical knowledge does not necessarily give researchers the skill of actually deciphering indistinct audio with unknown content and context. Indeed, while researchers are rarely tested for this skill, anecdotal evidence suggests that phonetics experts are not always better than non-phoneticians at forensic transcription.

This situation is reminiscent of the one uncovered by Markham (1999) in relation to determining speakers' regional dialects. At that time, it was assumed that phonetics experts who were highly skilled in describing and analysing regional dialects would naturally also be highly skilled in determining the dialect of speakers whose regional origin they did not know. However, it turned out that, when deprived of contextual information, the experts made a surprising number of errors. This does not at all negate the scientific knowledge gained through phonetic analysis of known dialects. It simply differentiates that knowledge from the skill of determining a regional dialect in the absence of external information, or, even more difficult, in the face of misleading information (for further background on the role of contextual information in determination of speakers' regional and social origin, see Fraser, 2009).

A similar situation exists with transcription. The only way to be absolutely sure that a transcript of indistinct audio is correct is to evaluate it against 'ground truth' (indisputable knowledge of its content) – which of course is rarely possible in real forensic cases. Further, the lack of testing means experts may be unaware that their opinions are not always as accurate as they think they are. There may be a tendency to assume that experts can rely on acoustic evidence to determine the content. However, there is no strong evidence to support this assumption. In fact, there are good reasons to argue that acoustic evidence alone cannot reliably reveal the content of indistinct audio.

6. *Acoustic evidence alone is not enough*

6.1 Bottom-up vs top-down

One of the best established findings of phonetic science, as discussed at some length above, is that speech perception is not a 'bottom-up' process. For perception to occur, information from the speech wave must be combined with 'top-down' information from other sources, notably from the listener's understanding of the internal and external context. Nevertheless, there is a strong tendency for experts to assume that acoustic-phonetic analysis can resolve errors in police transcripts.

Of course, acoustic analysis is an extremely useful skill. However, as discussed, it typically involves starting from known words, and then observing their acoustic characteristics. Going in the reverse direction, from acoustic characteristics to words, is a very different matter – and the abilities even of experts are known to be limited.

This is seen in the spectrogram-reading competitions sometimes run by speech science associations. Even with short, clear phrases in high-quality recordings, members need a strong hint about the topic to be able to guess the phrase – and

even with the hint, many are likely to guess wrongly (that is what makes the competition fun). A similar phenomenon was seen in the audio demonstration with excised words discussed earlier – and it is important to recognise that having the spectrogram does not help experts identify the words.

Of course, in both cases – once the content is known – phoneticians can readily explain why the words have the acoustic characteristics they do, and why those characteristics make them so hard to recognise purely from bottom-up acoustic information. But the fact remains that they had not been able to use those acoustic characteristics to understand the words, or to identify any of their phonemes before they knew the content.

This is certainly not to belittle in any way the expertise of phoneticians. It is simply to acknowledge the well-known fact that, contrary to the misconceptions of 'educated common knowledge', this expertise does not allow us to 'read' the content of indistinct audio from an acoustic representation in an objective, context-free manner. Yet that is what the law asks experts to do – and what some experts claim to be able to do – with forensic audio.

6.2 Disputed utterances

In fact, the law rarely asks experts to do forensic transcription in an open-ended way. A far more common request is for evaluation of a 'disputed utterance' – a transcript suggested by one side and opposed by the other (usually but not always a police transcript put forward by the prosecution and opposed by the defence). In such cases, acoustic analysis can be useful in treating the suggested transcript as a hypothesis for evaluation. However, it has two important limitations.

First, for all the reasons discussed above, acoustic evidence is unlikely to reveal, unambiguously, the true content of the disputed utterance. After all, if the audio is auditorily indistinct, the acoustic evidence is likely also to be indistinct.

Acoustic analysis is generally more effective in ruling out an inaccurate suggestion, as in the 'shot the prick' and 'pact' examples discussed above. However, with indistinct audio, even ruling out cannot always be done with 100% certainty.

Importantly, the fact that a hypothesis cannot be ruled out does not necessarily mean that it is right. The limited acoustic information in indistinct audio may well mean the content is simply not able to be resolved with demonstrable reliability sufficient for the high stakes situation of a criminal trial.

Generally, all an expert can do is support the disputed suggestion to a greater or lesser degree. Responsible experts are usually cautious in expressing the degree of such support, pointing out the kinds of limitations discussed above. The problem is that experts have little control over how their evidence will be represented in the trial process. For one thing, they can provide their evidence only in response to questions from barristers who have (understandably enough) limited knowledge of phonetic science. More importantly, after the expert leaves, the barristers may sum up the phonetic evidence in ways that suit their case – but that might well reduce

the expert's caution and nuance. This already bad situation may be exacerbated by a second consideration.

6.3 Managing priming

The process of evaluating a disputed utterance inevitably exposes the expert to its textual priming. Of course, experts are less likely to be misled by a completely erroneous transcript (like those suggested in the 'pact' and 'prick' cases). However, inaccuracies are not always so egregious as these examples – and there is no reason to believe that experts are immune to having their hearing influenced by an erroneous but plausible transcript.

This is why it is important to recall that the transcript being evaluated arises from the contextual information available to one side of the dispute (usually but not always police) – before that information has been tested by the trial process. Further, since experts are typically briefed directly by the side hiring them, they are likely to receive subtle (or not so subtle) hints about contextual information that supports that side's view of the content (for examples, see Fraser, 2021).

Under these conditions it is difficult or impossible for an expert to be certain that their hearing has not been influenced by the suggestion they are evaluating. As seen in the excised-words and spectrogram-reading examples above, acoustic analysis depends greatly on the analyst's hearing in context. This is true even with good quality recordings, and far more so with indistinct audio.

For these and other reasons it is essential to manage the expert's exposure to suggestions about the content and context. Since we know that priming is essential to perception of indistinct audio, and that its effect cannot be controlled by an effort of will, it must be managed via the evaluation process.

7. *The Australian approach*

It seems clear that taking an 'evidence-based' approach to forensic transcription requires more than just providing acoustic-phonetic evidence to support or refute a hypothesis about the content of indistinct audio. It is necessary also to control the process by which hypotheses are generated in the first place. This is not easy when an individual expert interacts directly with a client (whether prosecution or defence).

At the Research Hub for Language in Forensic Evidence, we are developing a process which first accredits transcribers in reliable transcription of indistinct audio, and then ensures that they follow a process whereby relevant, reliable contextual information can be provided in a managed and accountable way.

We are also researching the best way to report on forensic transcription in court. The first step is, of course, for the expert to form a reliable opinion about the content of the audio. However, the ultimate aim is not for the expert to reach the right conclusion about the audio, but for the jury to reach the right conclusion, so that the audio evidence can be combined appropriately with all the other evidence they are evaluating in order to reach their verdict. This makes it important to consider a

wider range of factors that might potentially cause the jury to misinterpret the audio than the transcript itself. In particular, it is essential for transcripts to be presented in ways that do not inadvertently feed common misconceptions about language and speech known to affect the legal process (Fraser, Loakes, 2020).

8. *Conclusion*

This paper has reviewed the use of indistinct forensic audio in criminal trials, showing that it is powerful evidence with potential to be powerfully misleading. Current legal procedures, in Australia and elsewhere, are incapable of fully protecting the court from the influence of potentially misleading transcripts. Phonetic science has an important part to play in solving this problem – first by participating in a law reform process, and second by developing evidence-based methods for creating reliable transcripts of indistinct audio to be used as evidence in court.

The latter goal requires acknowledgment that acoustic evidence alone is not enough to confirm the content of indistinct audio. It is necessary to take a broader view of 'evidence-based', which recognises and understands the legal context in which forensic audio is used, and especially of how priming is managed behind the scenes of the actual trial (see Fraser, 2020; 2021).

This requires a new approach from phonetic science, in which we not only explain our knowledge to the law but seek to understand important differences between scientific and legal concepts of evidence. For example, phonetic science typically uses acoustic-phonetic analysis to provide evidence that confirms or disconfirms a theoretical explanation. The issue is not to determine the content of the audio, but to decide how best to represent the content in order to use it as data relevant to the research. This is very different from the issue in forensic transcription, which seeks to assist a third party in determining the content, and then using the content to help reach a verdict. Recognition of this and other differences opens the path to new and interesting developments in phonetic science, with potential not just to improve the provision of forensic evidence, but to contribute to development of our field as a whole.

Acknowledgements

Bibliography

BURRIDGE, K. (2017). The dark side of mondegreens: How a simple mishearing can lead to wrongful conviction. *The Conversation*. Retrieved from [http://theconversation.com/the-dark-side-of-mondegreens-how-a-simple-mishearing-can-lead-to-wrongful-conviction-78466].

FRASER, H. (2003). Issues in transcription: Factors affecting the reliability of transcripts as evidence in legal cases. *International Journal of Speech Language and the Law*, 10(2), 203-226. [doi: 10.1558/sll.2003.10.2.203].

FRASER, H. (2009). The role of 'educated native speakers' in providing language analysis for the determination of the origin of asylum seekers. *International Journal of Speech Language and the Law*, 16(1), 113-138. [doi: 10.1558/ijsll.v16i1.113].

FRASER, H. (2018). Forensic transcription: How confident false beliefs about language and speech threaten the right to a fair trial in Australia. *Australian Journal of Linguistics*, 38(4), 586-606. [doi: 10.1080/07268602.2018.1510760].

FRASER, H. (2019). Don't believe your ears: "Enhancing" forensic audio can mislead juries in criminal trials. *The Conversation*. Retrieved from [https://theconversation.com/dont-believe-your-ears-enhancing-forensic-audio-can-mislead-juries-in-criminal-trials-113844].

FRASER, H. (2020a). Enhancing forensic audio: What works, what doesn't, and why. *Griffith Journal of Law and Human Dignity*, 8(1), 85-102.

FRASER, H. (2020b). Forensic transcription: The case for transcription as a dedicated area of linguistic science. In COULTHARD, M., MAY, A. & SOUSA-SILVA, R. (Eds.), *The Routledge Handbook of Forensic Linguistics*. London: Routledge, 416-431.

FRASER, H. (2020c). Introducing the Research Hub for Language in Forensic Evidence. *Judicial Officers' Bulletin*, 32(11), 117-118.

FRASER, H. (2021). The development of legal procedures for using a transcript to assist the jury in understanding indistinct covert recordings used as evidence in Australian criminal trials: A history in three key cases. In *Language and Law/Linguagem e Direito*, 8(1), 118-134.

FRASER, H., KINOSHITA, Y. (2021). Injustice arising from the unnoticed power of priming: How lawyers and even judges can be misled by unreliable transcripts of indistinct forensic audio. *Criminal Law Journal*, 45(3), 142-152.

FRASER, H., LOAKES, D. (2020). Acoustic injustice: The experience of listening to indistinct covert recordings presented as evidence in court. *Law Text Culture*, 24, 405-429.

FRASER, H., STEVENSON, B. (2014). The power and persistence of contextual priming: More risks in using police transcripts to aid jurors' perception of poor quality covert recordings. *The International Journal of Evidence and Proof*, 18(3), 205-229. [doi: 10.1350/ijep.2014.18.3.453].

FRASER, H., STEVENSON, B. & MARKS, T. (2011). Interpretation of a crisis call: Persistence of a primed perception of a disputed utterance. *International Journal of Speech Language and the Law*, 18(2), 261-292. [doi: 10.1558/ijsll.v18i2.261].

FRENCH, P., FRASER, H. (2018). Why "ad hoc experts" should not provide transcripts of indistinct forensic audio, and a proposal for a better approach. *Criminal Law Journal*, 42, 298-302.

LANGE, N.D., THOMAS, R.P., DANA, J. & DAWES, R.M. (2011). Contextual biases in the interpretation of auditory evidence. *Law and Human Behavior*, 35(3), 178-187. [doi: 10.1007/s10979-010-9226-4].

MARKHAM, D. (1999). Listeners and disguised voices: The imitation and perception of dialectal accent. *International Journal of Speech Language and the Law*, 6(2), 290-299. [doi: 10.1558/sll.1999.6.2.289].