HOW 'ENHANCED' FORENSIC AUDIO IS EVALUATED IN CRIMINAL TRIALS: WHAT IF ALL THAT REALLY GETS ENHANCED IS THE CREDIBILITY OF A MISLEADING TRANSCRIPT?

Helen Fraser

University of New England, NSW, Australia
hfraser@une.edu.au

ABSTRACT

Covert recordings can provide powerful evidence in criminal trials. Since the audio is often of poor quality, many jurisdictions allow an 'enhanced' version to be admitted, along with a transcript, to assist the trier of fact in understanding the content of forensic audio. But how is 'enhancing' evaluated? In Australian courts, it is simply a matter for the jury to decide whether the processed audio 'sounds clearer' than the original. This paper presents two perception experiments showing, first, that 'enhancing' can make audio sound 'clearer' in the sense of 'less noisy' without making it objectively more intelligible; and, second, that 'clearer' audio makes listeners more likely to accept an unreliable transcript. This is a problem in view of common practices that result in admission of unreliable police transcripts as 'assistance' to juries.

Discussion urges researchers to consider the legal context in which their work will be interpreted.

Keywords: forensic phonetics, audio enhancing, speech perception

1. INTRODUCTION

Covert recordings (typically obtained by 'bugging' houses or vehicles) are used during investigation of most major crimes. Increasingly, they are also used as evidence in court, when the crime comes to trial.

Unfortunately, due to the manner of their capture, many covert recordings are indistinct to the point of being unintelligible. In such cases, the jury, or other trier of fact, needs assistance to understand the evidence. Legal procedures for providing this assistance vary from one jurisdiction to another. The present paper focuses on Australia, but many of the comments are relevant more widely.

The most common form of assistance is a transcript, typically prepared by police [1]. However, here we consider another frequently used aid to perception: an 'enhanced' version of the audio.

'Enhancing' is a vague term covering a wide variety of processes which are poorly understood by the community in general, and by the legal community in particular [2]. For example, one influential Australian judgment ruled that enhancing is 'the aural equivalent of the use of a magnifying glass to enhance an individual's capacity to perceive the relevant record' (*R v Giovannone* [2002] NSWCCA 323 at para. 58). This is plainly incorrect [2]. However, in combination with other factors, its use as a precedent means criteria for admission of 'enhanced' audio in court are very lax [2]. There are known cases (one discussed below) of ineffective and potentially misleading 'enhancements' being admitted as reliable assistance.

There is thus a need, especially in view of general moves towards greater rigour in forensic science [3], [4], for better control over the validation of audio enhancements used as evidence in court. Indeed some valuable moves have recently been made in this direction [5].

The current paper discusses two sets of considerations that may be worth keeping in mind during development of criteria for validating audio enhancement. The first set stem from findings about the role of priming in speech perception. The second set stem from observations about how covert recordings are actually used in trials.

2. CONSIDERATION 1: THE POWER OF PRIMING

For speech in general, and especially for indistinct speech, the signal underdetermines the message it transmits [6]. Speech perception requires information from the signal to be combined with information from other sources. Priming is the powerful but generally unnoticed cognitive process via which listeners use contextual expectations to 'get ready' to perceive speech. These expectations come from knowledge (or assumptions) about the situation, and/or from suggestion of specific words that might be heard. In a forensic context, the latter typically takes the form of a transcript.

In everyday situations, priming is generally helpful, since listeners' expectations are usually in harmony with 'ground truth' (the actual content of the speech). Rare cases of error are readily corrected when the right interpretation is revealed.

It is notable, however, that acknowledging the role of priming problematises the everyday concept

that a recording is objectively either 'clear' or 'unclear'. With indistinct audio, the same speech can seem perfectly clear to listeners primed with appropriate expectations, but highly unclear to those who lack background knowledge of the content. More importantly, priming with misleading expectations can override acoustic cues, creating confident but erroneous perception [7].

This makes it very easy, in forensic contexts, where 'ground truth' regarding the content of the recording is by definition not available as a corrective, for listeners to be unwittingly misled by an unreliable transcript or by misleading contextual expectations – and very difficult for them to recover from perceptual errors. Cases are known of actual and potential injustice arising from use in court of unreliable transcripts of indistinct audio [8].

While priming itself has been given considerable attention in forensic phonetics in recent years [8], the question of how priming interacts with enhancing has so far been little studied. This paper makes a start by reporting two simple experiments that bear on that issue.

Both use short snippets of indistinct audio for which both the 'original', and a version 'enhanced' in a manner similar to those admitted in Australian courts, are available in the public domain. Unlike most real cases, however, 'ground truth' regarding their content is known. This enables us to use participants' responses to evaluate objectively whether the 'enhancing' has been effective in revealing what was said. More importantly – and the main intention of these studies – it enables us to test listeners' ability to evaluate the effect of enhancing.

2.1. The 'fish' experiment

The 'fish' experiment [2] played a short snippet of (non-forensic) audio in original and enhanced versions to 60 participants.

In Part 1, participants played each version once only (without being told which was which, or given any information about the audio), and were simply asked to state which one they found 'clearer'. Responses indicated that 62% found the enhanced version clearer, while 38% found the original clearer.

In Part 2, participants were divided into two groups of 30, randomly assigned to listen to either the original or the enhanced version (without being told which it was). This time they were asked to listen as many times as they wished and transcribe what they thought was said (again, no information was provided about the recording or its context).

The most consistent response was that 37% correctly heard one key word ('fish'). Of these, 32% were listening to the original, while 68% were

listening to the enhanced version. This initially looks like support for the effectiveness of the enhancing. However further results show this impression to be misleading.

First, there was little difference in the content of responses from groups listening to 'enhanced' and 'original' versions. Both gave highly variable transcripts, none with any resemblance (beyond the one word 'fish') to what was really said. This suggests the enhancing had had no objective effect in improving the intelligibility of the audio.

Second, the group listening to the 'enhanced' version was more likely (80% vs 47%) to offer *some* interpretation (albeit inaccurate), while the group listening to the original was more likely (53% vs 20%) to indicate the audio was unintelligible – which is arguably the 'right answer' in this situation.

Results were interpreted as showing that it is possible for 'enhancing' to make audio seem 'clearer' (presumably via reduction of noise) without objectively improving its intelligibility. This potentially compromises the reliability of audio evidence. The next experiment shows how.

2.2. The 'JB' experiment

The JB experiment [9] aimed to further investigate the interaction between enhancing and priming. Here the recording was a short snippet of unintelligible audio related to a real murder. Although the evidence has not been tested in court, the murder and the audio have been widely discussed in the media, due to claims that 'modern enhancing techniques' reveal four phrases spoken by family members of the deceased, indicating they were involved in the murder.

However, while the phrases (evidently originating from a police transcript [10]) have been widely accepted as valid, it seems the audio is not speech at all, but the sound of typing [9], [11].
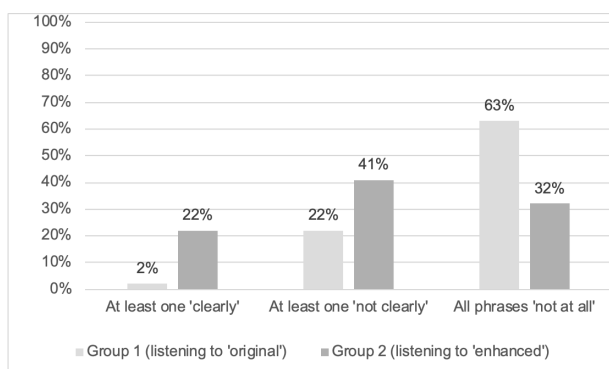
The experiment used 78 participants in two groups. In Step 1, the groups were randomly assigned to listen to either the original or the enhanced version (without being told there were versions, or given any other information about the audio). They were invited to listen as often as they wished, and transcribe what they heard.

Only a quarter of all participants heard any kind of speech at all, and no one in either group heard anything remotely like the four alleged phrases. This was interpreted as demonstration that the 'enhancing' had had no objective effect in 'revealing' the incriminating content. Interestingly, however, as with the 'fish' experiment, those listening to the enhanced audio were substantially more likely (24% vs 5%) to hear *some* words, though

no two participants heard the same words, and few transcribed coherent phrases.

Step 2 primed participants in both groups by explicitly suggesting the movie's four phrases as a possible interpretation of the audio. Overall, 44% now claimed to hear at least one of the phrases, even if not clearly. Importantly, however, those listening to the enhanced version were substantially more likely than those listening to the original (63% vs 24%) to accept the (misleading) phrases, and to hear at least one of them 'clearly'. See Figure 1.

Figure 1 After priming (Step 2), how many participants in each group claimed to hear at least one of the alleged phrases 'clearly', 'not clearly', or 'not at all'. Note: Before priming (Step 1) no participants in either group heard anything remotely like any of the phrases.

These results were interpreted as further demonstration (adding to the previous experiments) that enhancing can make audio seem 'clearer' without objectively improving its intelligibility.

However, they also allowed an additional important observation: making audio seem 'clearer' can, far from allowing listeners to hear more reliably, exacerbate the already powerfully misleading effect of priming with an unreliable transcript.

2.3. Does it matter?

It might be suggested that these results merely indicate the ineffectiveness of the particular enhancing techniques used in these examples. Indeed, it could well be true that better results are possible – there is certainly a good deal of research aiming to develop more effective forensic enhancing techniques [12].

The current question, however, is not whether it might be possible to do better enhancement of any particular audio, but whether, in general, ordinary listeners can reliably evaluate whether any particular enhancement they are presented with has objectively improved the intelligibility of the audio. Results of the experiments just discussed indicate it is not.

Again, it might be suggested that more qualified listeners, perhaps experienced phoneticians, could do a better job than the listeners in the experiments. That also might be true.

However, what really matters is not whether a particular expert might hypothetically evaluate enhanced audio reliably. What really matters is how enhanced audio is actually evaluated by the courts in which it is used as evidence. That is the question to which we now turn.

3. CONSIDERATION 2: HOW 'ENHANCED' AUDIO IS USED IN TRIALS

As with other aspects of covert recordings, rules for admission and use of enhanced audio were developed without consultation of the linguistic sciences [1]. Consequently, as mentioned above, they embody poor understanding of what is involved. 'Enhancing' is commissioned from a wide range of 'experts', who seldom have any background in linguistic or cognitive phonetics, and often have low qualifications even in audio engineering.

Admission is rarely contested, and when it is, the final criterion is whether the judge, listening personally, considers that the jury might potentially find the 'enhanced' version 'clearer' than the original. This is because evaluation of the effectiveness of an enhancement is considered ultimately to be a matter for the jury, who are invited to use it if and only if they find it more helpful than the original in discerning the content of the audio.

Of course, judge, jury and everyone else involved in the trial are well primed, both by extensive background knowledge or assumptions about the context, and by a transcript (whose evaluation is also, ultimately, a matter for the jury [1]). Based on findings like those discussed above, it seems possible they might find a less noisy 'enhanced' version 'clearer' even if it makes no objective improvement to the intelligibility of the audio.

This possibility is confirmed by an informal experiment [2] which used audio from a real trial in two versions: the original, and an 'enhanced' version that was admitted, despite expert evidence that its apparent clarity was potentially misleading, on the grounds that the judge heard it as 'clearer'.

Experiment participants given no information about the audio were unable to hear any words at all in either version. For both versions, when prompted with contextual information, some participants heard words, but they all heard different words, none remotely like the police transcript.

This confirms that the 'enhancing' had not made the audio objectively more intelligible – though it might well have made it seem 'clearer' in the sense

of making the (unreliable) police transcript more credible to the jury than it would have if they had listened only to the (extremely indistinct) original.

4. THE PROBLEM

It seems clear there is a problem with the handling of forensic audio used as evidence in criminal trials, at least in Australia. Before recommending a solution, it is worth thinking through the nature of the problem, and the relevance of the considerations outlined above, in more detail.

First, whatever solution might be offered, it is essential that it should take account of how the law actually works. After all, the best enhancement in the world could be useless or worse if used in court under conditions that tend to mislead listeners. Indeed, solving problems with forensic enhancing will likely need discussion as to how legal practice can be changed so as to better align with scientific principles.

Importantly, however, before getting to that point, it is essential for those recommending solutions to know enough about how the law works to ensure their recommendations are feasible, and will not risk causing as many problems as they solve.

Second, whatever solution might be offered, it is essential that it should take account of how speech perception actually works. After all, the purpose of enhancing audio is to make it easier for listeners to perceive what is said in the recording, and speech perception is relevant at every step in the process.

Consider, for example, a possible recommendation that enhancing should always be done by qualified experts, protected from biasing knowledge or assumptions about the case. This would indeed be a step forward from current practice. However, understanding of speech perception is still relevant, for several reasons.

First, even the most responsible professionals evaluate their own enhancing by listening. Contrary to popular belief, there is no automatic technique that can take unintelligible audio as input and produce intelligible speech as output [13]. Audio enhancing is acknowledged to be an inherently subjective process: analysts typically proceed by carefully explore the boundary between helpful processing and detrimental over-processing.

Most importantly, even to the extent that enhancing can be said to make audio sound 'clearer', in a real forensic context the question must always remain as to whether that clarity reliably reflects the ground truth of what was actually said at the time the recording was made. The fact that experiments, using audio for which ground truth is known, show that listeners can be easily led by apparent 'clarity'

to confident but erroneous perception, must surely urge caution regarding reliance on what 'sounds clear' in forensic contexts – where ground truth is necessarily unknown.

It is interesting to note that the UK Forensic Regulator [14] has urged similar caution in relation to image enhancement – though not (yet) in relation to audio enhancement [11].

Finally, it is worth repeating that, no matter how expert enhancers themselves may evaluate their own work, what really matters is how the court evaluates it – and they will always be listening under very different conditions from those of the expert. In particular, they will inevitably be heavily influenced, at least by the very contextual knowledge from which the experts may have carefully protected themselves, and often also by a potentially misleading police transcript.

For all these reasons and more, it is essential to ensure that strategies for validating audio enhancement are developed in an interdisciplinary manner that takes full account of cognitive, as well as technical, aspects of phonetic science.

5. THE SOLUTION

Enhancing is one of several issues (along with transcription, translation and speaker attribution) that have led Australian linguists to create a Call to Action, requesting the judiciary to set in train a process of review and reform regarding procedures for admission and use of covert recordings as evidence in criminal trials [1].

The aim is for authorities to recognise the need to instigate a collaborative, evidence-based, interdisciplinary project involving law, law enforcement and multiple branches of linguistic science, in order to create a process capable of ensuring covert recordings are used reliably and fairly in criminal trials.

While bringing this about will inevitably be a long-term project, there is a lot that can be done by speech science experts in the meantime.

Perhaps most important is for genuine experts to take every opportunity, not just to display the capabilities of their techniques, but to disabuse law and law enforcement personnel of the common misconception that making indistinct audio 'sound clearer' is necessarily helpful in revealing what was actually said in a forensic recording.

6. REFERENCES

[1] Fraser, H. 2018. Thirty years is long enough: It's time to create a process that ensures covert recordings used as evidence in court are interpreted reliably and fairly, *Journal of Judicial Administration*. 27(3), 95-104.

[2] Fraser, H. 2018. 'Enhancing' forensic audio: False beliefs and their effect in criminal trials, *Australian Journal of Forensic Sciences*.
DOI: 10.1080/00450618.2018.1491115

[3] National Research Council Committee on Identifying the Needs of the Forensic Sciences Community. 2009. *Strengthening forensic science in the United States: A path forward.* Washington, DC: National Academies Press.

[4] Forensic Science Regulator. 2017. *Codes of practice and conduct for forensic science providers and practitioners in the criminal justice system.* London: UK Forensic Science Regulator.

[5] Bartle, A., Eardley, M., Hargreaves, C., Harrison, P., Kaplan, A., Rhodes, R., Smith, C. 2018. An approach to audio enhancement validation, IAFPA Conference, Huddersfield.

[6] Magnuson, J., Mirman, D., Myers, G. 2013. Spoken word recognition, in Reisbert, D. (ed) *Oxford Handbook of Cognitive Psychology*. 412-441.

[7] Bruce, D. 1958. The effect of listeners' anticipations on the intelligibility of heard speech. *Language and Speech*, 1. 79-97.

[8] Burridge, K. 2017. The dark side of mondegreens: How a simple mishearing can lead to wrongful conviction, *The Conversation*.

[9] Fraser, H. 2019. Enhancing forensic audio: What if all that really gets enhanced is the credibility of a misleading transcript? *Australian Journal of Forensic Sciences*. DOI: 10.1080/00450618.2018.1561948

[10] McDonell-Parry, A. 2016. Three big ways 'The Case of: JonBenet Ramsey' got it wrong, *Rolling Stone*, 20 September.

[11] Fraser, H. 2019. Don't believe your ears: 'Enhancing' forensic audio can mislead juries in criminal trials, *The Conversation*.

[12] Nic Daeid, N. 2010. *International Forensic Science Symposium Interpol: Review Papers*. Centre for Forensic Science, University of Strathclyde.

[13] Maher, R. 2015. Lending an ear in the courtroom: Forensic Acoustics, *Acoustics Today*, 11(3), 22-29.

[14] Forensic Science Regulator. 2016. *Forensic image comparison and interpretation evidence: Guidance for prosecutors and investigators*. London: UK Forensic Science Regulator.