‘Enhancing’ forensic audio: false beliefs and their effect in criminal trials

Helen Fraser

To cite this article: Helen Fraser (2018): ‘Enhancing’ forensic audio: false beliefs and their effect in criminal trials, Australian Journal of Forensic Sciences, DOI: 10.1080/00450618.2018.1491115

To link to this article: https://doi.org/10.1080/00450618.2018.1491115

Published online: 06 Jul 2018.

Submit your article to this journal

Article views: 54

View Crossmark data
‘Enhancing’ forensic audio: false beliefs and their effect in criminal trials

Helen Fraser

Discipline of Linguistics, University of New England, Armidale, Australia

ABSTRACT

Indistinct covert recordings admitted as evidence in criminal trials are routinely ‘enhanced’ to assist a jury in making out their contents. But just what is ‘enhancing’, and how effective is it? This paper uses two short experiments to demonstrate that a subjective impression that ‘enhancing’ has made the audio ‘clearer’ does not necessarily indicate there has been an objective improvement in intelligibility. It then outlines, in a non-technical manner, the capabilities and limitations of various ‘enhancing’ techniques, and discusses implications in relation to current legal practices around the admission and use of ‘enhanced’ audio in Australian criminal trials. Finally, it recommends that ‘enhanced’ versions of forensic recordings should only be admitted on the basis of objective evidence of the extent to which they have genuinely improved the intelligibility of the specific audio being used, noting that such evidence is easy to obtain and provide.

KEYWORDS

Enhancing; transcription; linguistics; audio

1. Can ‘enhancing’ reveal JonBenét’s brutal killer?

The recent twentieth anniversary of the brutal murder of a six-year-old beauty queen in Colorado, USA, was marked by the release of a television movie, The Case of: JonBenét Ramsey (now available on YouTube). The film purported to shed new light on the still-unsolved crime, by way of re-analysis of several pieces of evidence (for background on both the murder and the movie, Wikipedia makes a good starting point).

One key item examined in the movie is a six-second snippet of audio heard after the end of the 911 call in which JonBenét’s mother, Patsy Ramsey, reported her daughter missing. The original audio is unintelligible, to the point that it is not clear it is even human speech. However, the movie makers claim that ‘enhancing’ this audio has revealed words which, along with other evidence presented in the movie, point to the identity of JonBenét’s killer. (‘Original’ and ‘enhanced’ versions of this audio snippet, along with all the other short audio samples mentioned in this article, can be found at forensictranscription.com.au/audio. Where possible, it is recommended that readers listen to all the audio before proceeding.)

There are many points worthy of discussion regarding the claims made by this movie, and especially the use it makes of the 911 call. Here we focus on the dangerous false beliefs it perpetuates about the potential for ‘enhancing’ to make unintelligible audio intelligible, and the effect these false beliefs have on admission and use of forensic audio in Australian criminal trials.
2. What is ‘enhancing’?

The core meaning of the word ‘enhance’ is ‘to improve the quality of (something that is already good)’. It is accurately used in this sense by audio engineers who improve the aesthetic quality of studio-recorded music, movie soundtracks, etc., typically for the entertainment industry. Less accurately, but frequently, it is used interchangeably with ‘restore’, to refer to the process of improving the quality of audio that has become degraded over time, typically in re-mastering old vinyl records. In both of these cases, skilled audio engineers, using high-end technology, can create impressive results.

In the forensic context, the meaning of the word ‘enhance’ is extended to include processes applied to indistinct covert recordings, i.e. conversations recorded secretly and used as evidence in criminal cases\(^2\). This semantic extension is unfortunate, since the forensic context is different to that of the entertainment industry in several relevant ways. First, the audio is often of significantly worse quality than even badly degraded studio recordings, to the point it can be literally unintelligible in the absence of prior knowledge about its content. Second, the ‘ground truth’ regarding the content of forensic audio is not known, making objective evaluation of the effect of ‘enhancing’ difficult. In fact, the content of the audio is often the very issue the ‘enhancing’ seeks to resolve, creating the potential for dangerous circularity of reasoning. These points will be taken up in more detail after discussion of two illustrative examples.

3. A brief Australian case study

3.1. The case

The JonBenêt movie makes a dramatic introductory example, but forensic audio ‘enhanced’ in similar ways is routinely admitted by Australian criminal courts to ‘assist’ the jury in understanding the content of indistinct forensic audio. As discussed further below, it is rare for its admission to be challenged. However, in one recent trial, I provided expert evidence that ‘enhancing’ techniques applied to the audio had not objectively improved its intelligibility, and that using the ‘enhanced’ version had the potential to mislead the jury. The judge excluded my report on the grounds that speech perception is a matter of common knowledge, not requiring expert opinion. Instead, he listened personally to both original and ‘enhanced’ versions, and ruled that the ‘enhanced’ version was indeed clearer, so could be made available for the jury to consider whether it might assist their hearing (this may seem surprising to non-lawyers, but, as discussed further below, it is standard practice).

3.2. An informal experiment

To test the validity of the judge’s opinion (after the trial), I conducted an informal experiment using the online survey tool, Qualtrics. Participants were 12 graduate students in law. The experiment used original and ‘enhanced’ versions of four 1-minute segments of the (lengthy) covert recording, chosen for the crucial role played in the trial by words the prosecution claimed were made more intelligible by the ‘enhancing’.
Participants were given a very brief introduction to the concept of ‘enhancing’, and to the distinction between listenability (how pleasant audio is to listen to) versus intelligibility (how well listeners can understand the words). They were then told their task was to rate the listenability and intelligibility of pairs of audio samples that had been ‘enhanced’ with two different methods, in order to help scientists determine which method was more effective. In fact, participants were presented, not with two ‘enhanced’ versions, but with the original and the ‘enhanced’ version.

Part 1 presented the four pairs of samples in random order. Participants were asked to listen to each pair once only, and to rate each version for listenability and intelligibility (operationalized in terms of how many times they thought they would need to hear the audio in order to make out the words). They then moved on to Part 2, which used only one of the four pairs. Participants were randomly assigned to hear either the original or the ‘enhanced’ version of this pair. They were asked to listen as many times as they wished, and write down what they heard in the audio. Finally, they were invited, optionally, to move on to Part 3. This offered some general context about the case, without suggesting specific words, and again asked participants to listen as often as they wished, and write down what they heard.

3.3. Results

In Part 1, for all samples, the original version was subjectively rated by participants as considerably more ‘listenable’ and somewhat more ‘intelligible’ than the ‘enhanced’ version (i.e. the exact opposite of the intention of the ‘enhancing’ – and the exact opposite to how the judge evaluated them).

Recall that by rating audio as intelligible, participants – who at this stage had listened only once – were indicating their belief that, if they were to listen repeatedly, they would be able to make out what was said. However, when tested, in Part 2, for what they could actually hear upon repeated listening, none of the participants, listening without context to either the original or the ‘enhanced’ version of one section, could write down any words at all.

Few participants went on to the optional Part 3 (reflecting the difficulty of the task, a point taken up below). With the addition of contextual information, most could hear some words. However, there was no evident effect of the ‘enhancing’ on what was heard, and none of the participants heard anything remotely like the words in the police transcript whose intelligibility the prosecution claimed their processing had ‘enhanced’.

3.4. Discussion

It is not possible to draw strong conclusions on the basis of this small and informal experiment. However, the results do surely call into question the judge’s confident assertion that the ‘enhanced’ version was ‘clearer’, and thus capable of assisting the jury in hearing any words it contained. Of course, there is no suggestion here that the judge was intentionally misdirecting the court. He undoubtedly did experience the ‘enhanced’ version as clearer. The point (to be returned to) is that subjective evaluation is not a reliable basis for deciding the effectiveness of ‘enhancing’ forensic audio.
4. The ‘fish’ experiment

4.1. Aim

This experiment took advantage of an opportunity, provided by a BBC science show about forensic audio, to compare the subjective and objective effects of ‘enhancing’ indistinct audio in a more formal way. The radio host presented the original and an ‘enhanced’ version of a short, indistinct audio sample in a before- and-after format. The clear assumption was that the audience would recognize the ‘enhanced’ version as objectively more intelligible than the original, and thus understand how techniques like these could be useful in forensic contexts. Of course, all this was done in good faith, and there is no suggestion that there was any intentional manipulation of the audience. It was notable, however, that no objective test was done to determine whether the ‘enhanced’ version actually was more intelligible than the original. The current experiment aimed to determine this.

4.2. Materials

In the absence of a real forensic example, the radio show had used a 4-second snippet of poor-quality audio from a non-forensic field recording, first in its original form, then in an ‘enhanced’ version created by the presenter. I later contacted the presenter to obtain permission to use the snippet, and to ascertain the ‘ground truth’ regarding the words contained in the original.

4.3. Method

The experiment was deployed via the website forensictranscription.com.au, using a web-based quiz application. It had two parts. In Part 1, participants were asked to listen once only to original and ‘enhanced’ versions of the audio snippet (without being told which was which), and indicate which version seemed to them, on first impression, to be ‘clearer’. In Part 2, each participant was presented with one of the versions, randomly selected, and asked to listen as many times as they liked, then write any words they heard into a response box. See Table 1.

Table 1. Design of the ‘fish’ experiment.

<table>
<thead>
<tr>
<th>Part 1</th>
<th>Part 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stimulus</strong></td>
<td>Original version of snippet* AND enhanced version of snippet, presented in random order, without identification</td>
</tr>
<tr>
<td><strong>Instruction</strong></td>
<td>Play each version once only</td>
</tr>
<tr>
<td><strong>Question</strong></td>
<td>Which version sounds clearer?</td>
</tr>
</tbody>
</table>

*Snippet consists of two short utterances: Utterance 1, less than 1 second; Utterance 2, approximately 3 seconds (see Section 4.5.2).
4.4. Participants

In total, the experiment was done by 60 people, recruited via several email lists, some of which had a forensic theme. Half the participants transcribed the original version and half the ‘enhanced’ version. No demographic information was collected.

4.5. Results

4.5.1. Part 1

After listening once only, 37 of the 60 participants (62%) felt the ‘enhanced’ version sounded ‘clearer’, while 23 (38%) thought the original was ‘clearer’. Note that chance in this context is 50%.

4.5.2. Part 2

Overall, no participants interpreted either the original or the enhanced version remotely correctly. Whichever sample may seem ‘clearer’, neither of them is actually ‘clear’ enough for listeners to determine its content in the absence of contextual knowledge (or assumptions) about what was said.

Beyond this, it is worth looking at the content of the transcripts participants entered. For purposes of analysis, the snippet can be divided into two utterances (see Table 1), which we look at in turn (all participants of course heard both utterances as part of the snippet).

4.5.3. Utterance 1 transcripts

Utterance 1 is just a few syllables in duration. Of the 60 participants, 22 (37%) transcribed the word ‘fish’, which was in fact spoken. This was by far the most accurately transcribed word in the entire recording. However, the surrounding words were transcribed variously, with only one correct response:

- Molly’s fish
- Polly fish
- Hobbies fish
- Bobbie’s fish
- I’ll use fish
- What are these fish?
- All these fish
- This is fish.

Interestingly, 15 of the 22 (68%) who correctly heard ‘fish’ were listening to the ‘enhanced’ version (50% of that group of 30), and one of these was the only participant to transcribe the first utterance correctly. By contrast, only seven of the 22 (32%) who correctly heard ‘fish’ were listening to the original version (23% of that group of 30), with none hearing the full utterance correctly.

Before concluding that the ‘enhancing’ had actually made the audio clearer, it is worth considering the rest of the results.
The next most commonly heard phrase in Utterance 1 was ‘police station’, which was not accurate (possibly influenced by the ‘forensic’ context in which some participants heard about the experiment). This was erroneously transcribed by six of the 60 (10%) participants, of whom five were listening to the ‘enhanced’ version (17% of that group), and only one was listening to the original (3% of that group).

Ten participants heard a range of other phrases, exemplified by the following:

- this is special
- what is this?
- the copies switched
- switched on
- it’s fiction
- the platitude
- English
- restriction.

In total, 22 of the 60 participants (37%) found the first utterance uninterpretable. Interestingly, 16 of these 22 (73%) were listening to the original version (53% of that group of 30). By contrast, only six of the 30 (20%) in the group that listened to the ‘enhanced’ version were unable to transcribe the first utterance. Again that might seem at first to indicate the ‘enhancing’ had made an improvement – until you recall that only one of the 24 (4%) who offered a transcript of the ‘enhanced’ version heard the utterance correctly. All the others, though no doubt believing themselves to be listening carefully and hearing objectively, were effectively guessing, sometimes wildly.

Arguably, in a forensic context, ‘uninterpretable’ is a more reliable transcript than a wild guess confidently heard.

4.5.4. Utterance 2 transcripts
Utterance 2 was slightly longer than Utterance 1. No participants transcribed it in a way remotely like the ground truth of what was actually said. No consistent differences between transcripts based on original or enhanced versions were evident, so results are presented together.

Overall, 52 of the 60 (87%) heard phrases including one or two uses of the word ‘time’. In fact, however, the word ‘time’ was not spoken in the original utterance. The most common interpretation was ‘I need some time’ or ‘I’ll need some time’, heard by 31 of the 60 (52%) participants overall, including 20 (67%) of those listening to the original and 11 (37%) of those listening to the ‘enhanced’ version.

Others heard other phrases involving ‘time’ or words such as ‘tone’, ‘tonight’ or ‘tight’ embedded in a wide range of different phrases (none remotely correct). Some representative transcripts (of the whole snippet) give an impression of the variety:

- I’ll use fish, so I’ll need some time, so tonight.
- Bobby’s fish. I’ll need some time (thyme) so for the thousandth time
- this is special, something and tights
- fish something Time’s quality enhancing tone
• Police station. I’ll need some time, so call the cops on time.
• Patty’s said sh-. Patty’s in time, like what did house in time.
• the copies switched. I’ll need some time so we’ve got accounting time.
• It’s fiction but I need some time so ? some time
• This is fish. I’ll need some time so I’ve got a ‘planting’ time
• Restriction. At least in time/tone so we’ve got an enhancement tone
• I need some time so I can go out hunting tonight.
• I need some time cause I’ve got a concert tonight.

4.6. Discussion

Again, caution is required in drawing detailed conclusions from this very simple experiment. However, it seems fair to say it adds weight to the proposition that subjective evaluation of the effects of ‘enhancing’ on forensic audio is not reliable. The radio presenter (who, like the judge in the case study experiment, had prior information about the content of the audio) evidently assumed that ‘enhancing’ had made it clearer in both senses (more listenable and more intelligible). In fact, however, only 62% of experiment participants rated the ‘enhanced’ version as subjectively ‘clearer’ than the original, and even for those, testing showed the ‘enhancing’ had made no objective improvement to the intelligibility of the audio.

5. What are the capabilities of forensic ‘enhancing’?

Both these experiments suggest the effectiveness of the ‘enhancing’ techniques used is considerably less than claimed. So what are the possibilities for improving the quality of indistinct forensic recordings? Are these examples representative?

As in the entertainment industry, if a forensic recording is intelligible but unpleasant to listen to (due to abrupt variations in amplitude, the presence of interfering noise, etc.), there are good possibilities for improving the listenability of the audio. However, this is rarely done with forensic audio, where the aesthetics of the recording are not the main concern. The court’s aim is to determine what is being said and who is saying it – a far more problematic endeavour than deciding whether it ‘sounds pleasanter’.

Understanding the problem is made easier by starting from analogy with visual images. Thanks to the prevalence of digital cameras, most people have become familiar with the wonderful results that can be obtained by ‘enhancing’ photographs – as well as with the potential of ‘photoshopping’ to completely alter an image. These results, however, are only possible when the original image is of reasonable quality. If the original is so blurry that the subject matter is barely recognizable, ‘enhancing’ has little effect. As the saying goes: ‘garbage in, garbage out’.

The situation with forensic audio is similar – though less well known, since fewer people have sufficient personal experience undertaking aesthetic audio enhancement to have developed a sense for the boundaries of what is possible. For many, the nearest equivalent experience is television dialogue that cannot be heard clearly without subtitles. The audio that is sent for forensic ‘enhancement’ is typically of far poorer quality than a subtitled broadcast, however, as seen in the examples above.
The possibilities for improving the intelligibility of audio of this quality are very limited. Under very specific circumstances, short sections can be improved via detailed hand processing by an expert in phonetics—though great care must be taken to protect the expert from unconscious bias that might cause a ‘photoshop’-type alteration of the original.

This kind of processing by a genuine expert in phonetic science is rare in casework, however. Most ‘enhancing’ is done by audio engineers. These vary greatly in expertise. Some are school leavers using free software downloaded from the internet. Others are well qualified in audio-engineering (a very different discipline from phonetics) who use specialist equipment such as Cedar Studio (https://www.cedar-audio.com). This kind of equipment enables the engineer progressively to refine the frequency distributions of the audio until the speech sounds ‘clearer’. As with aesthetic enhancing, however, this is acknowledged to be a subjective process, greatly reliant on the personal judgement of the individual engineer.

While the technicality of ‘enhancing’ makes it seem like a science, it is really an art. Two engineers working on the same input are unlikely to produce identical outputs. Further, as we see from the experiments above, making audio seem ‘clearer’, in the sense of having less interfering noise, does not necessarily mean it will be ‘clearer’ in the sense of being more intelligible. There is also the possibility that the engineer might alter the audio to make relevant words sound clearer, with less attention to the effect on other words. This might perhaps have been the reason that more participants listening to the ‘enhanced’ version of the ‘fish’ recording heard the word ‘fish’. A crucial point, however, is that the radio presenter knew for a fact that the word ‘fish’ was part of the ground truth of what had been said. In a forensic context, the content of the audio is, by definition, not factually known (although, depending on how they have been briefed, an engineer might well have conscious or unconscious opinions about it).

Despite being the subject of well-funded research over several decades, to date there is no standard, widely applicable process that can reliably improve the intelligibility of indistinct forensic audio. Maher provides an authoritative but accessible overview of the engineering background. For more detail see ‘Forensic audio analysis’ in the Interpol International Forensic Science Managers Symposium. For an informal impression, consider the enormous unmet demand for such a process: if one had been developed, it would be very widely used, making the inventor justly famous.

All this of course raises the question: if forensic ‘enhancing’ is so ineffective, why is it used at all?

6. Why is forensic ‘enhancing’ used at all if it is so rarely effective?

Listeners to indistinct audio often have the sense of being nearly able to hear the words. It seems the speech is merely covered by a mask, and that someone with the right tools should be easily able to remove the mask to reveal the words. Indeed, this can sometimes be done – but only if the original audio was recorded with the words and the noise on separate tracks. This is extremely rare with genuine forensic audio – although rather common in demonstrations, where a mask is applied to a clear recording, the resulting ‘unintelligible’ audio is played to an unsuspecting audience, then the mask is removed in a ‘hey presto’ revelation of the power of ‘enhancing’.
In the normal case, the noise and the words, though originating from separate sources, are convolved by the recording into one acoustic signal. Our ears, by focusing on the meaningful words and ignoring the noise, can do an extraordinary job of de-convolving the signal. To date, there is no mechanical method that can routinely and reliably achieve better results than (trained and unbiased) human ears. And, of course, beyond a certain signal-to-noise ratio, even the best human ears falter and fail.

The reason is that listeners’ sense that words in a poor-quality speech signal are ‘masked’ is inaccurate. The reality, in most cases, is that the information needed to hear the words does not exist in the signal. Rather than removing noise, improving the quality requires adding acoustic information, which of course cannot be done reliably without ‘ground truth’ knowledge of what was said.

Unfortunately, all this, though well established in science, remains little known by the wider community. As the number of indistinct forensic recordings increases year by year, audio engineers are urged to ‘do what you can’, and clients can be surprisingly uncritical of the results – even when responsible engineers explain that their processing has improved only listenability, not intelligibility. This credulity of course opens the door for unscrupulous practitioners, but that is not the worst problem.

An even more unfortunate consequence is that well-intentioned researchers focus on demonstrating they are inching towards the reliable ‘enhancing’ techniques the public so badly wants, rather than on explaining the current situation clearly. Thus, it is rare to find simple experiments, like those above, aiming to disabuse the legal community of unrealistic expectations, and help create a fresh start, by seeking a more practical way to solve the problems posed by indistinct forensic recordings.

Perhaps the nearest equivalent is a BBC White Paper responding to complaints that up to 60% of viewers have trouble understanding the dialogue in television dramas. Having explained the reasons for this, and canvassed various options for processing the audio to make it easier to understand, the White Paper concludes

> The literature available indicates that there is little if any chance of audio processing improving intelligibility of speech in noise, and a real danger of degrading it. Whilst audio processing can be used to create cosmetic improvements in a speech signal it cannot be used to improve the ability of an audience to follow the words.

> The best recommendation is for viewers to use subtitles.

If this conclusion holds for broadcast-quality audio with known content, surely there must be some flaw in the concept of using ‘enhancing’ to improve the intelligibility of extremely indistinct audio whose content is to be used as forensic evidence.

### 7. Why are listeners so uncritical of the effects of ‘enhancing’?

If the possibilities for genuinely improving indistinct forensic audio are so limited, why are listeners – even a sober judge – so easily misled into believing it has been successful?

Again, the visual analogy is useful in explaining this. A poor quality image is not necessarily uninterpretable. It has the potential to be interpreted as something it is not – especially if those viewing it have inaccurate assumptions about what the image represents. When the ‘enhancer’ shares these assumptions, even if not fully consciously, it is
easy to ‘photoshop’ the image to emphasize features that, unintentionally or otherwise, strengthen an inaccurate interpretation.

All this is well known in relation to images (see, for example, the useful and accessible explanation provided by the UK Forensic Science Regulator\textsuperscript{7}). In Australia, thanks to important work in the psychology of facial recognition, use of indistinct photographs as evidence is becoming better regulated, though it is still far from ideal\textsuperscript{8}.

The situation is exactly the same with audio, except that, again, despite being well established in phonetic science, it is less well known in the wider community. The defining characteristic of ‘indistinct’ audio is not just that it is hard to hear any words at all, but that listeners are liable to hear different words under different conditions\textsuperscript{9}. Most importantly, listeners’ confidence in their perception correlates poorly with their objective accuracy. This makes them highly susceptible to a simple trick, which can be perpetrated intentionally or unintentionally.

The effect of ‘enhancing’ audio is rarely demonstrated by straightforward comparison against the original, as in the experiments above. Far more commonly, the original is first played alone, then the ‘enhanced’ version is played with a transcript, or some other suggestion as to what the listener should expect to hear. This is what was done, for example, in the JonBenët movie, which shows the engineer twiddling knobs to make speech waves dance on the screen, while the investigators announce the words they claim are revealed by the ‘enhancing’ (failing to mention that these words were in fact suggested to them from another source\textsuperscript{1}). Then the ‘enhanced’ version is played, not alone, which would give a direct comparison, but with the ‘transcript’ provided as subtitles for viewers to follow while listening to the audio.

From a scientific point of view, this clearly confounds two variables: the enhancing and the transcript. However, few viewers recognize this, and, to judge by the impact of the movie, many are persuaded that the ‘enhancing’ has indeed revealed the words suggested by the transcript. This reflects a strong but false belief of common knowledge that speech perception is a matter of ‘picking up’ sounds that are ‘there to be heard’, then putting them together to make words, phrases and sentences. In this view, ‘enhancing’ the quality of the sounds must inevitably ‘enhance’ the quality of the perception. However, this view is wrong.

Phonetic science has long known\textsuperscript{10} that perception of speech, especially but not only of indistinct speech, works in exactly the reverse manner. Listeners start with an expectation of what words and sentences they might hear, then use ‘cues’ in the acoustic signal to confirm or disconfirm their hypothesis. This is the reason that priming listeners with a transcript can cause the transcript to be confidently accepted even when it is manifestly wrong\textsuperscript{11}.

In this scientific view, ‘enhancing’ speech is expected to have a small effect relative to the effect of the transcript, which is indeed the case\textsuperscript{12}. Unfortunately the relevant scientific research is virtually unknown in the wider community. Most people experience a transcript as merely ‘assisting’ their perception, not realizing the extent to which the transcript actually creates their perception. Further, most people trust what they then ‘hear with their own ears’, not realizing there is nothing in listeners’ experience that is capable of letting them distinguish reliably
between accurate and inaccurate perception. More importantly, there is nothing in public education to alert listeners to the low correlation between their subjective confidence and their objective accuracy.

The effect of all this is that listeners, assisted in the perception of indistinct audio by both ‘enhancing’ and a transcript, are liable to wrongly attribute the assistance they receive to the ‘enhancing’, not noticing the effect of the transcript – indeed likely to deny it has had any influence upon them.

8. How is ‘enhanced’ audio used in Australian courts?

With this background, we turn now to consider the criteria for admitting ‘enhanced’ versions of indistinct forensic recordings in Australian courts.

The impression given by the case study above is indicative of standard practice. In a key ruling used as a precedent in many subsequent trials, the judge admitted ‘enhanced’ audio on grounds that what the engineer had done was ‘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 straightforwardly wrong. A magnifying glass makes an image appear bigger. This can certainly help a viewer see more detail in a high quality image, but it is not capable of improving a poor quality image. For audio, the equivalent of magnification is making the signal louder. This of course can be useful when the only factor making the audio hard to hear is that it is too soft. However, this is rarely the case with forensic audio – and when it is, boosting the amplitude, or using headphones, can be done without requiring an engineer.

Where engineers come in is when the audio is indistinct. Making any effective change to such audio, whether an improvement or not, necessarily requires altering the structure of the sound. As discussed above, this can be done in a wide range of ways, with widely different results. Yet following Giovannone and other rulings, judges admit a proffered ‘enhancement’ as ‘the’ enhanced version, frequently with little or no scrutiny of either the techniques or engineer, as shown by this example from a real case (not the one referred to above):

CROWN PROSECUTOR: [...] Is this one of the discs that was enhanced?

DETECTIVE: Yes, it was.

CP: By an out-company; is that right?

D: Yes.

CP: That is called [NAME OF MUSIC COMPANY]?

D: Yes.

HIS HONOUR: What does that involve, can I ask?

DETECTIVE: My understanding is we gave them a copy of the master CD and they placed it in the computer program and they put a number of filters through the audio recording in an attempt to enhance the sound.
This was the end of the explanation – evidently the court was satisfied with this response. Even when engineers are cross-examined about the ‘enhancing’ (as in the case used for the experiment above), they typically recount the technical processes they applied, rarely if ever being asked to demonstrate the objective effects of the processes on the intelligibility of the audio.

This apparent laxity can perhaps be understood by recalling that evaluating the effect of ‘enhancing’ processes is considered a matter of common knowledge, and thus left for the jury to do. The ‘enhancement’ is provided along with, not instead of, the original. The jury is invited to listen to both (optionally; in their own time), and use the ‘enhanced’ version only if they find it assists their perception. In this context, it is easy to see why a judge might prefer to err on the side of admitting an ineffective ‘enhancement’ rather than risk excluding an effective one. However there are two important flaws in this reasoning.

First, the need to play and compare multiple versions of indistinct audio places a substantial burden on the jury. To see this, recall that in the case study experiment above, few participants continued to Stage 3, since comparing several 1-minute samples had already taken too much of their time and energy

Second, as we have seen, people genuinely (though erroneously) believe in the power of ‘enhancing’ to improve intelligibility. In the context of a trial, the ‘enhanced’ version is always heard in conjunction with a transcript. Listeners are already liable to hear indistinct audio in line with a transcript even if the transcript is manifestly inaccurate. Belief that they are listening to an ‘enhanced’ version of the audio is liable to increase this effect, with the result that what gets ‘enhanced’ is not the intelligibility of the audio but the credibility of the (potentially unreliable) transcript. The risk of using ‘enhanced’ forensic audio in court is that, in evaluating the transcript against the audio, insufficient recognition is given to the fact that the transcript has itself already contributed to evaluation of the audio. This creates dangerous circular reasoning.

Rather than relying on subjective evaluation of the effects of ‘enhancing’ by judge and jury, the burden of proof should be on the party seeking admission of the ‘enhancement’ to demonstrate objectively that it has made a genuine and reliable improvement to intelligibility. This demonstration must include more than a general description of the technical processes applied to the audio and their intended effects. It is essential that it include demonstration of the actual effects on the audio being evaluated in the trial. Such demonstration is easy to provide via a simple experiment, like those described above.

9. Conclusion

This paper has reported two small experiments whose findings should create unease and disquiet in the forensic science community, about the frequent but poorly regulated use of ‘enhanced’ audio in our criminal justice process. This is one of several issues that have moved Australian linguists to deliver a Call to Action to the judiciary, seeking review and reform of the handling of covert recordings used as evidence in criminal trials. While it will no doubt take some time for this Call to Action to yield results, in the meantime there is scope for lawyers and scientists to demand simple, objective tests of the efficacy of ‘enhancements’ admitted in trials.
Perhaps the ‘fish’ experiment above offers a useful model that can be applied to any ‘enhancing’ as a basic ‘first pass’ test of its effectiveness.

**Acknowledgments**

Thanks for comments on earlier versions of this paper are due to Alex Bowen, Peter Gray, Roger Shuy and two anonymous reviewers. Thanks to participants in both experiments, and to relevant parties for permission to use the stimulus materials.

**Disclosure statement**

No potential conflict of interest was reported by the author.

**References**

14. 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.