

Transcription of indistinct forensic recordings: Problems and solutions from the perspective of phonetic science

Helen Fraser

Forensic Phonetics Australia

Abstract. *Covert recordings (speech captured on audio or video without the knowledge of the speakers) can provide powerful forensic evidence. Unfortunately, however, since it is difficult to control their recording conditions, covert recordings are often indistinct, to the extent the speech is unintelligible to listeners who do not already have knowledge or expectations about their content. This raises the question of how to present the speech evidence in court so that the trier of fact (jury, magistrate, judge, etc.) can make use of the information contained in the recording. In many jurisdictions, the answer is to have a transcript made by those who know the context (typically police working on the case), and provided to the trier of fact as an aid to perception of the speech. The present article outlines several problems with this approach, then suggests some solutions that allow maximal value of the intelligence contained in covert recordings, while reducing the risk of injustice through biased perception of indistinct audio. A key part of the suggested solution is to make a clear distinction between investigative and evidentiary uses of indistinct covert recordings, and to ensure that transcripts of evidentiary recordings be produced by professional transcribers independent of the case.*

Keywords: *Forensic phonetics, forensic transcription, covert recordings, digital forensics, forensic surveillance.*

Resumo. *Gravações clandestinas (fala capturada em áudio ou vídeo sem o conhecimento dos locutores) podem fornecer poderosa evidência forense. Infelizmente, no entanto, uma vez que é difícil controlar as condições de captura, as gravações por interceptação são muitas vezes indefinidas, na medida em que a fala pode ser ininteligível para os ouvintes que ainda não têm conhecimento ou expectativas sobre o seu conteúdo. Isso levanta a questão de como apresentar essa evidência no tribunal para que o juiz de fato (júri, magistrado, juiz, etc.) possa fazer uso das informações contidas na gravação. Em muitas jurisdições, a solução é ter uma transcrição feita por alguém que conheça o contexto (alguém da polícia que trabalhe no caso), e fornecida ao juiz de fato como auxílio à compreensão da fala. O presente artigo descreve vários problemas com essa abordagem, e depois sugere algumas soluções que permitam aproveitar ao máximo o conteúdo dessas*

gravações clandestinas, reduzindo o risco de injustiça por percepção enviesada de áudio com problemas para compreensão. Uma parte fundamental da solução é fazer uma clara distinção entre as utilizações dessas gravações clandestinas pouco claras, para investigação ou para prova, e para garantir que as transcrições de gravações de prova sejam produzidas por transcritores ou profissionais não envolvidos com o caso.

Palavras-chave: *Fonética forense, transcrições fonéticas, gravações por interceptação.*

Introduction

Technological advances are making it ever easier to collect covert recordings (conversations recorded electronically without the knowledge of the speakers). Legally obtained¹ covert recordings can potentially yield powerful evidence in criminal trials, allowing the court to hear speakers making admissions or giving information they would not have been willing to provide in person, or in an overt recording (one made with all speakers' knowledge, for example in a police interview).

A major limitation of covert recordings², however, is that it can be hard to control their recording conditions, with the result that the audio is often of very poor quality, to the extent it is difficult to hear what is said. This is the problem that is the topic of the present paper.

In certain circumstances, it is possible to overcome this limitation by providing a transcript to assist listeners, notably the trier of fact (the judge, magistrate or jury deciding the verdict of the trial), to hear what words are spoken. Of course, the reliability of a transcript used for this purpose is crucial. Otherwise there is a danger it might 'assist' the listeners to hear words other than those spoken in the conversation originally captured by the covert recording, and thus to reach an inappropriate evaluation of the evidence.

Unfortunately, ensuring reliable transcription, especially of poor quality audio, can be highly problematic. This is well known across multiple branches of linguistic science, where transcripts are frequently used for research purposes (Bucholtz, 2007; Heselwood, 2013), and specifically in forensic phonetics, which has a small but well-established branch dealing specifically with transcription of indistinct audio (French and Stevens, 2013; Coulthard and Johnson, 2007; Shuy, 1993).

However, these issues in transcription seem to be less well known in legal circles, where practices for the handling of audio evidence have developed with little reference to the relevant science. For example, it is common, internationally, for indistinct covert recordings played in court to be accompanied by transcripts produced by police investigating the crime. This may seem sensible, on the grounds that investigators' knowledge of the context of the recording helps them to make out words that are unclear to others – and indeed it is important to recognize that contextual knowledge can aid perception. However, as discussed below, use of transcripts by those involved in the case raises serious problems, which can affect the fairness of trials.

The present paper starts by setting out some general background about transcription. Though each element of this background is well known in the sciences of language and speech, they are rarely brought together to be viewed as a whole in the legal context. This background is then used to highlight several problems associated with use of

police transcripts of indistinct covert recordings. The paper finishes by suggesting, as a starting point for collaborative research between phonetic science and the law, some directions in which solutions to the problems might be sought, while still retaining the advantages of access to police contextual knowledge.

Definitions and distinctions

It will be useful to start by clarifying some concepts and terminology regarding speech recordings and their transcription. This will enable important distinctions to be maintained throughout the discussion, and allow forensic transcription (i.e. transcription of indistinct audio evidence) to be set in a broader framework. One aim of this section is to outline some features that make transcription in general a more difficult task than is sometimes recognised. Another is to demonstrate several ways in which the handling of forensic transcription differs from standard practice in general transcription.

What is a transcript?

The term 'transcript' was first used in the middle ages, before the development of the printing press, to denote a copy of a hand-lettered text, 'transcribed' or 'written across' from an original (Oxford Dictionary). Later, the word was used for a 'fair copy', written up from notes made during a meeting or event. This usage continues to the present day in relation to work such as that of court or parliamentary stenographers, who take shorthand notes and transcribe them into a written text, which becomes (after appropriate checking) the official version of the proceedings.

Alongside this has developed a new usage, made possible by the introduction and rapid spread of audio-recording technology. Nowadays, 'transcribe' most commonly means 'to write out (i.e. to represent in written form) the speech captured in an audio recording', and a whole new industry of audio-transcription has emerged, servicing legal, medical, research and general markets (Mills, 2010).

An unfortunate consequence of this semantic evolution is that connotations that had, reasonably, accrued to earlier uses of the word have been inappropriately transferred to the modern usage. In particular, it is often assumed by those who have never tried it that transcribing from an audio recording is a simple matter of 'writing down what you hear', a little like schoolroom dictation, making the product essentially a copy of the audio, albeit in a different medium.

Despite its ubiquity, however, this idea is inaccurate. As discussed below, it is well known in the linguistic sciences that a transcript is never anything like a copy of the audio. Nor is transcription ever simple — not even when the recorded speech seems easy to hear, and certainly not when it is indistinct. Ensuring reliability of a transcript, therefore, requires careful attention to a range of factors, some of which are canvassed briefly in the following sections.

Speech recordings: Purpose

Speech by nature is fleeting, disappearing the moment it is uttered. With audio technology it can be captured, allowing it to be heard again, in a different context or by different listeners. This can be done for a range of different reasons. One familiar purpose is to create a record of an official event, for example, a police interview or court proceedings.

An important aspect of overt recordings like these is that everyone knows their speech will be transcribed. For this reason, the speech is monitored for clarity, by the speakers themselves, or by others admonishing them to 'speak up for the tape'.

Speech recordings are also used by language scientists, to capture speech for various kinds of analysis and research. This naturally requires high-quality recordings. In the early days, this restricted research to short, clear utterances, sometimes called 'laboratory speech'. Later it became possible to capture high quality recordings of spontaneous conversational speech, though this remains notoriously difficult (Wray and Boomer, 2013).

Spontaneous conversational speech also proved extremely difficult to transcribe, even in a good quality recording. Though the overall meaning may be quite clear to the listener, detailed identification of every word is surprisingly difficult, and transcription is extremely time-consuming. However, the efforts of researchers across several decades have provided many insights into the vast differences between monitored and spontaneous speech (Shockey, 2003; Cauldwell, 2013).

These insights have had profound implications for our understanding of how speech is processed in the mind – which unfortunately remain little known outside academic circles (Cutler, 2012; Fraser, 2003). Most importantly, they indicate that speech is not perceived 'bottom up', i.e. purely from information in the acoustic signal. To a far greater extent than is evident from everyday experience, perception relies on the hearer's knowledge of the context, both internal context (available from the speech itself) and external context (from the background situation within which the recording was obtained).

This is seen, for a topical example, in the effects of recording courtroom proceedings and outsourcing the audio to casual transcribers. The 'bottom up' view would predict that having the audio should increase accuracy, as the transcribers can listen again to the actual words. In fact the opposite is true. Transcribers who have witnessed the proceedings produce better transcripts, even when relying only on shorthand notes, without reference to the audio (Wilson, 2013).

All the examples mentioned so far have been overt recordings, where the purpose is to retain a record for reference or analysis by those who heard, or could have heard, the original speech in its context. It is often possible, or even required, that those present be consulted before finalisation of the transcript. This is quite different from the purpose of a covert recording, which is to provide evidence of unmonitored conversation that would not have occurred in the presence of those listening to the recording. Checking the transcript with those present is problematic.

Within the overall category of covert recordings, an important distinction is whether the purpose is investigative or evidentiary/evidential (Haworth, 2010; French and Harrison, 2006).

Investigative uses are those related to the investigation of a matter, before it goes to trial, when detectives or other investigators attempt to uncover the facts surrounding an alleged crime. For example, if a covert recording reveals suspects making plans to meet at a particular address, this may prompt police to raid the premises at the relevant time. If the raid is successful, its results become evidence to be used in the trial. The recording itself may never be played again.

Evidentiary uses of covert recordings are those where the audio is played in court as evidence of the crime itself. Importantly, it is not just the fact of particular words having been uttered that is relevant. The manner in which the words are uttered is essential to the trier of fact in evaluating the intentions of the speakers, and the significance of the

audio evidence in relation to all the other evidence being weighed to reach a verdict. This 'manner of speech' can never be represented objectively in a transcript, no matter how detailed, but must be interpreted by the human ear. This is the reason most courts insist that it is the audio, not the transcript, that is the evidence, with the transcript provided only as an aid to listeners' perception of indistinct audio (more will be said about this below).

Speech recordings: quality

Another factor to be considered in ensuring reliable transcription is the technical quality of the recording itself. In principle, the technical quality of a recording is a separate issue from the clarity of the speech being recorded. However, in practice these interact substantially. For example, careful, pre-planned speech may be understood even when recorded with cheap equipment, while casual conversation can be surprisingly hard to understand even when the best equipment is used (as discussed above). For present purposes, then, it is useful to define quality in terms of clarity of speech, and to distinguish three levels.

Clear recordings, are those where most of the speech can be understood readily, in one sitting, by anyone with competence in the language, even if they know little about the context of the recording. This does not require studio quality, such as that appropriate in listening for pleasure. Many overt recordings are of only fair quality, and may have significant background noise. All that is required for them to fit into the *clear* category is that the speech in general is easily intelligible. Thus, while there may be issues of analysis, or of interpretation of speakers' intentions, overall there is little room for disagreement about what words are actually spoken.

Poor or *unclear* recordings are those where the speech itself is hard to understand, and listening is significantly unpleasant. On first hearing, listeners, especially those lacking background knowledge of the context, may pick up only an impression of the conversation, and not its detail. To hear it properly requires headphones, software to enable isolation of particular sections of audio, and repeated, patient listening, with note-taking. However, with these requirements in place, it is possible to hear most of what is said, to the extent that several independent transcribers are liable to reach overall agreement on most of the content – though of course there may still be differences of opinion regarding the analysis or interpretation of the recorded conversation.

Some poor quality recordings can be further classified as *indistinct*. These are recordings so poor that casual listeners are liable to understand little or nothing of the content. Even with specialised equipment, repeated listening, and contextual knowledge, it is still difficult to hear the conversation, and multiple transcribers are liable to produce significantly different versions of the content.

For most purposes, indistinct recordings are discarded, and for many, even poor quality audio is not used. Another unusual feature of covert recordings used as evidence in trials, then, is that they may be, and often are, used even if substantial portions are indistinct.

Transcripts: Purpose

In general, the most common purpose for a transcript is convenience. Written text is far easier to navigate, index, annotate and refer to than audio, which must be listened to in real time. Of course, this convenience depends on the transcript being reliable.

With court proceedings, for example, or even the minutes of meetings, elaborate checking processes have been developed over centuries to ensure the written record is not compromised by any kind of fraud or personal bias on the part of the transcriber.

Convenience also depends on the written text containing the relevant information in accessible form. The criteria for this vary considerably according to specific aspects of the intended purpose. For general use, an orthographic transcript (i.e. one using ordinary spelling) in a standardised format is usually appropriate, as it can be easily read and referenced by a wide range of users. It is worth noting, however, that this accessibility comes at the cost of considerable loss of detail. Thus even so-called ‘verbatim’ transcripts do not literally represent each and every word spoken (Eades, 1996).

Transcripts intended for research such as discourse analysis, for example, require far more information to be represented than is possible with normal orthographic conventions, and make extensive use of special symbols and conventions relevant to the specific topic being studied (Sidnell and Stivers, 2012; Edwards, 2008). The advantage is, again, that once the transcript has been accepted as reliable, analysis can be conducted on the written symbols, reducing (but not eliminating) the need to refer repeatedly to the actual audio. Of course, this level of detail comes at the cost not just of increased time required for preparation of the transcript, but also of reduced legibility to a general audience.

Most importantly, it is very widely recognised that even the most detailed transcription involves selection and interpretation, whether according to explicit or implicit criteria, of the features to be represented (Jefferson, 2004; Green *et al.*, 1997). This goes not only for discourse-level transcripts but also for those used in phonetic research (Heselow, 2013; Cox, 2012). Speech is an extremely complex signal (Kreiman and Sidtis, 2011; Laver, 1994), and it is literally impossible to include every aspect in a transcript (Kerswill and Wright, 2008). This is why it is misleading to treat any transcript, no matter how detailed, as a copy, or even an objective representation, of the audio it represents.

Turning now to indistinct covert recordings used as evidence in criminal trials, we find a totally different purpose for the transcript. It does not simply allow convenient reference to words that have previously been heard clearly by multiple people. It is emphatically not intended as any kind of substitute, however limited, for the audio itself. Rather, it is intended to assist the listener in making out words that they would find difficult or impossible to hear without the transcript, leaving them free to concentrate on interpreting the speakers’ intentions without undue influence from the transcriber’s opinions (see Fraser, 2015 for further discussion).

Transcribers: skill level

Recognition that there is much more to producing a transcript than simply ‘writing down what you hear’, makes clear that a major factor in the reliability of a transcript, and its usefulness for its intended purpose, is the skill level of the transcriber. For the present context, we can distinguish several categories.

Untrained transcribers are those who undertake transcription with little or no training or systematic, reflective experience. (Of course this refers to training and experience specifically in transcription; they may have high levels of expertise in other professional skills.) Untrained transcribers may work very hard on a transcript, listening many times to the audio, and updating frequently to reflect changes in their hearing. However, due

to their lack of experience with the variability of speech perception, they may be inclined simply to accept their most recent hearing as accurate. They may also have little understanding of the effect of layout on transcript usability.

Professional transcribers have considerable experience in the kind of analytic listening needed to choose among multiple potential perceptions, as well as explicit training in representing speech and laying out transcripts in the manners suitable for particular purposes. They may even have been tested to ensure the accuracy and usability of their output.

Experts in transcription are those with a theoretical understanding of the nature of transcription, and its many complexities — some of which are summarised in this paper.

Experts in forensic transcription are those with high-level qualifications in relevant branches of phonetics, enabling them to evaluate the acoustic evidence supporting (or not) a particular transcript. It is important to acknowledge that such experts may be employed by police-based forensic analysis units. These officers have little direct involvement in investigation of cases, creating a very different situation from the one discussed below, where indistinct audio is transcribed by police working on the case, who lack any expertise in transcription.

Expertise in forensic transcription is often sought when there is a ‘disputed utterance’ (French, 1990). Without it, resolving the dispute becomes a mere ‘battle of the ears’. It is worth emphasising though, that forensic transcription is a broader topic than resolution of disputed utterances. Many covert recordings have indistinct portions lasting hours or even days. Unfortunately, however, as discussed below, these seem less likely to be sent for expert analysis than shorter segments with explicit alternative transcriptions.

It is notable that in virtually all contexts in which transcripts are used, it is a standard requirement that they will be prepared by professional transcribers, and that the less clear the audio the more skill the transcriber needs. Even with covert audio, transcripts of clear or poor recordings are normally produced by transcribers with at least some training and experience. Again indistinct covert recordings are an exception, being often transcribed by police with no training whatsoever in transcription.

Transcribers: relationship to material

Another factor that is usually given considerable attention in ensuring the reliability of transcripts is the relationship of the transcriber to the material being transcribed. We have seen already that, while general background knowledge of the context is useful or essential in preparing a transcript that is reliable and useful for its purpose, steps must be taken to ensure effects of personal bias are avoided.

However there is another kind of bias which is more difficult to avoid. Cognitive bias (Kahneman, 2011) is the tendency for people to perceive something they expect, assume or want to be present, even if it is not objectively there. Importantly, cognitive bias is unconscious. It can exist even in those who feel themselves to be neutral, and it cannot be controlled by a mere effort of will (Thompson, 2011). This is well known as the reason that medical and other sciences insist on ‘double-blind’ analysis of experimental results. A very similar, though less widely publicised, effect exists in speech perception, where it is often called ‘priming’. It is discussed further below.

In view of these observations, we can classify transcribers according to their relationship to the audio being transcribed. An *independent* or *impartial* transcriber has

little or no information about the specific context of the recording, beyond general background knowledge essential for producing the desired type of output, and no prior opinion about, or particular interest in, the ultimate use that will be made of the transcript.

By contrast, an *involved* transcriber does have such an opinion or interest. In linguistic research, considerable effort is made to ensure transcripts are prepared by independent, impartial transcribers. Even where this is difficult for practical reasons, academic standards require at least a proportion to be transcribed in a way that ‘blinds’ the transcriber to the purpose, and the level of agreement between this and other portions to be reported.

Here again, transcription of indistinct audio diverges from standard practice, with transcripts by involved transcribers, notably police, often used.

How does forensic transcription relate to transcription in general?

With this brief background, we are in a position to locate forensic transcription within a broader framework of understanding regarding what a transcript is, and the standard processes that must be followed to ensure it is reliable.

Most importantly, the discussion makes clear that transcription of lengthy indistinct covert recordings is very different from most other forms of transcription. First, the quality of such audio is far worse than that used for most other purposes. Second, the intended purpose of the transcript is very different, going beyond mere convenience of reference, to influence on perception. Third, despite this, there is heavy reliance on transcripts by untrained, involved transcribers. Finally, it is worth noting that the negative consequences of errors in a forensic transcript are considerably higher than those in transcripts used for most other purposes. When the comparison is set out explicitly like this, it becomes unsurprising that it results in a range of problems.

Problems in current practice regarding transcription of indistinct forensic recordings

The majority of covert recordings are obtained on behalf of police, so it is natural that the audio would go first to them to determine if any of the material is pertinent to their investigation. To save time, they may obtain a summary transcript (notes on the content of the audio) prepared by an independent transcription agency, to help guide them to relevant parts. Police then note any information of investigative value, which they act on appropriately. At this investigative stage, to the extent police transcribers’ contextual knowledge of the case helps them to hear indistinct audio accurately, the intelligence provided by the recording will be valuable, while inaccuracy in their transcription is likely to result at worst in some waste of time or effort.

It is only later, when the investigation is complete, and a case is being prepared for trial, that it is necessary to decide which parts of the covert recording are relevant as evidence in their own right, and require detailed transcription for use in court. Clear and poor recordings are sent to independent professional transcribers. It is indistinct parts, too hard for the professionals, that may be transcribed by detectives from the case.

The audio, along with the transcript, is then disclosed to the defendant or representatives for checking, admission of the evidence to the trial is sought, and the audio is ultimately played to the trier of fact with the transcript provided as an aid to perception of indistinct sections. As indicated earlier, this process has numerous problems.

Inaccuracy of content

In terms of the classifications above, police transcribers are typically *untrained* and *involved*, and it is these characteristics that create the problems discussed here, not the fact that the transcribers are police. Thus the present remarks apply equally to transcripts by any untrained, involved transcriber, regardless of whether their transcripts are used by prosecution or defence.

The most notable problem with transcripts by untrained, involved transcribers is that they are liable to be inaccurate. The advantage such transcribers have over a professional, independent transcriber — and the reason their transcripts are used — is their knowledge about the specific background and context of the recording. And it is important to acknowledge, as discussed above, that such knowledge can sometimes enable them to hear words that are unintelligible to others. However, due to the effects of cognitive bias, this contextual knowledge is a double-edged sword, creating more problems than it solves.

First, involved transcribers tend to focus on sections they consider to have most relevance to their investigation. This may lead them to pay less attention to parts they consider unimportant, with heavy use of fillers such as ‘indistinct’ or ‘indecipherable’ or simply ‘[...]’. Of course, from a less involved, or differently involved, perspective, these sections may contain crucial information.

Second, they may mis-hear (i.e. hear words or phrases even when these are contradicted by the acoustics) or over-interpret the audio (i.e. hear words or phrases even when they are not well supported by the acoustics). Unfortunately, to untrained listeners, the experience is the same whether valid contextual knowledge is helping them to hear accurately, or assumptions or expectations are creating perceptual error. In both cases, they feel confident they are simply ‘hearing what is there to be heard’, and accept their perception uncritically.

Since this phenomenon is rather little known outside phonetic science, it can seem hard to accept when first encountered. However, it is a very well established feature of human speech perception, with strong experimental and experiential support going back at least to the 1950s (Miller, 1951; Cutler, 2010). A particularly clear example is given by one of the experiments that first brought this interesting feature of human speech perception to light. Bruce (1958), investigating the effect of listeners’ mental ‘set’, or context-based anticipations, created a number of sentences with the following form:

Sentence 1: I tell you that our team will win the cup next year

Sentence 2: You said it would rain but the sun has come out now

Participants heard these sentences ‘masked’ with a hissing noise, which he calibrated so as to make the sentences around 25% intelligible. Next, he presented the same sentences in the same level of masking noise — but this time he preceded each with a keyword that gave it a context. For example, the keyword for Sentence 1 was SPORT, for Sentence 2, WEATHER, and so on.

As predicted, sentences preceded by their keyword were more intelligible. However, an unexpected discovery was what happened when he played the masked sentences with the wrong keyword. This did not hinder perception, as had been predicted. Rather it created a different perception. For example, playing Sentence 1 with the keyword

FOOD (instead of SPORT) led participants to hear a range of sentences quite different from the one that had actually been spoken, such as:

Sentence 1 (FOOD): I tell you that I feel more hungry than you are

Playing the same Sentence 1 with the keyword TRAVEL led participants to hear yet another range of sentences, again different from the one that had actually been spoken, such as:

Sentence 1 (TRAVEL) I tell you that I too will leave next year

Crucially, these erroneous perceptions were heard with no diminution of confidence. Listeners felt they were simply hearing what was there to be heard, in just the same way they did when their perception was accurate – an early indication that listeners' personal confidence in their perception of indistinct audio is a poor guide to the accuracy of a transcript.

Though these findings were surprising at the time, and remain less widely known by the general public than similar effects in other areas of forensic evidence (Ridley *et al.*, 2013), the role of this kind of contextual priming in speech perception is now well understood in the phonetic sciences, as discussed above. Therefore, the argument that police transcribers might be similarly affected carries no suggestion of any personal or moral bias. However it is essential to recognize the role of cognitive bias, in audio as in other forms of evidence. It would be more surprising to find that police transcripts were not affected by cognitive bias than that they are.

Inadequacy of checking procedures

A common legal response, upon mention of these problems with police transcripts, is to point out that the transcripts are not just accepted uncritically. Most jurisdictions have processes of checking that must be undergone before the transcript is admitted. Unfortunately, however, these processes are frequently inadequate. They commonly involve the defendant and/or legal representatives checking the transcript against the audio. This may be acceptable, if not ideal, when the audio is fairly clear, and the transcript is professionally laid out. Such checking may lead to identification of one or more 'disputed utterances' which can be sent for expert analysis.

However, well known findings of phonetic science, discussed above and taken up again shortly, suggest that with indistinct audio it is not effective to evaluate transcripts in this way. Inexpert listeners may hear (accurately or not) a few of the phrases they read in the transcript, and assume the rest of it must be reliable – and of course, at this stage they do not know which utterances are going to be picked out as relevant in the trial, and thus require special attention.

A further feature of police transcripts can make this kind of checking even less likely to be effective than it might otherwise be, namely: their poor layout. We have emphasised above that an important part of training in transcription is learning how to format a transcript in a manner appropriate to its purpose. Police have no such training, and their transcripts are typically quite unsuited to the purpose of careful checking against the audio by another listener. Here, as an example, is an exact replica of a section of a police transcript from a real case (please note this represents less than a minute from within a 136-page transcript of a covert recording featuring several hours of barely audible conversation).

Male voice (F) huh?'

Male voice (P) '(indec — possibly fix the counter) (indec).' Male voice (F?) 'fix the what?'

Voices (indec).

Motor attempting to start.

Male voice (R?) 'no petrol.'

Motor attempting to start.

Voices (indec) swearing (fuck).

Male voice "What, What you want me to do?" (F?)

"No it's all good" (P)

Female voice (idec)

The question of exactly how forensic transcripts should be laid out for maximal usability by listeners in court is a complex one (the subject of research in progress by the present author), but it is clear layouts like this one are far from ideal (regardless of accuracy — which in this case was low). The mixing-up of speaker attributions, words heard, unconfident suggestions, reference to background noises, etc., makes it extremely hard to read this against rapidly-passing indistinct voices with a great deal of loud background noise.

Unrealistic expectations of trier of fact

Ultimately, the content of the audio must be evaluated by the trier of fact, as one piece of evidence to be weighed in with all the other evidence in a case, with the aim of reaching a verdict as to the guilt of the person undergoing trial. As discussed earlier, it is the audio that is the evidence. The transcript is intended only as an aid to the perception of listeners who may find an indistinct recording difficult to hear.

Again, while this use 'only as an aid' may sound reasonable based on everyday understanding of speech perception, phonetic science vigorously opposes it. Two recent experimental studies have demonstrated the reasons for this opposition, using indistinct audio and police transcripts from real cases, and closely simulating the experience of juries in interpreting a poor quality recording with the aid of an inaccurate police transcript.

Fraser *et al.* (2011) used a disputed utterance from the famous the case of David Bain (Innes, 2011), convicted in 1995 of murdering his entire family, then acquitted on all charges in 2009, after 13 years in prison. Subjects, divided into two experimental groups, listened repeatedly to the same audio, while evidence about the case, including the inaccurate police transcript, was gradually revealed to them across seven 'evidence points'. At each evidence point, they were asked what they heard in the audio, along with various other questions.

At first, virtually no one in either group heard anything remotely like the police transcript. However, at evidence point 4, when a transcript was explicitly suggested, over 30% of Group A, who were given the inaccurate police transcript, confidently heard the exact words of the transcript, with many others displaying perception clearly influenced by the transcript.

Subsequent evidence points attempted to convince participants that the police transcript was inaccurate, but brought about little change in perception. The final evidence

point explained the purpose of the experiment, with information the materials had been deliberately chosen to show how easily a demonstrably inaccurate transcript could mislead listeners' perception of indistinct audio. However, even after being told that experts on both sides of the case had agreed the police transcript was inaccurate, 17% of Group A claimed to hear its exact words, with, again, many more influenced by it in a variety of ways.

Most interesting of all was the response of Group B at this last evidence point. Group B had been primed with a different transcript, and had never heard anything like the police version. However, mention of the police transcript at the final evidence point prompted 12% of Group B to hear the audio exactly in line with the transcript, and many others to be influenced by it.

These results suggest that it may be unrealistic to expect a jury to use a transcript 'only as an aid', and to reach their own independent conclusion as to what is said in an indistinct recording. A useful and important follow-up study (Bonifaz, 2014) demonstrated that explicitly informing participants at the outset that they might potentially be primed by the suggested transcript did not significantly reduce their tendency to be affected by the prime.

The second experiment (Fraser and Stevenson, 2014) carried this work forward by looking at how knowledge, or assumptions, about the context can affect perception of words in an indistinct recording. Again, the experiment used audio from a real case, this time a short excerpt from a 38-minute recording of extremely poor quality – along with the police transcript used in the trial and later demonstrated to be inaccurate. It was conducted in two parts.

The first part demonstrated that, in the absence of contextual knowledge, the police transcript was actually quite implausible. Its priming effect was less than usual when first presented to participants, and even the few who initially accepted it, quickly abandoned it when presented with a more plausible alternative. This raises the question of how the transcript could ever have been accepted in court. An answer was suggested by the second part of the experiment.

The second part began by giving participants contextual knowledge about the trial, similar to the background available to lawyers in the case, and, later, to the jury. These participants were far more likely than those in the first part to hear the audio in line with the inaccurate police transcript, and far less likely to be swayed by the alternative, more plausible, transcript. As with the 2011 experiment, many subtle effects were found on the perception even of participants who rejected the police transcript.

Again, these results were interpreted as evidence it is unrealistic to instruct a trier of fact that they should use the transcript only as an aid. They also went further to suggest that leaving evaluation of forensic transcripts to defendants and their legal representatives gives insufficient protection from inaccuracy – and to demonstrate other problems with the law regarding police transcripts.

Finally, in both experiments, before-and-after questions revealed a strong effect of seeing the inaccurate transcript on participants' ultimate opinions about the guilt of the defendant, even in those who rejected its exact words. Participants believed their understanding of the case was influenced by the audio, not realising how much their perception of the audio was influenced by their understanding of the case.

Towards some solutions

It is hoped that this brief discussion has demonstrated some dangers of using transcripts by untrained, involved transcribers as an aid to perception of indistinct covert recordings presented as evidence in criminal trials. In summary, it is highly likely that such a transcript will be inaccurate in its representation of what is said in the recording, highly unlikely that errors will be picked up through the currently standard checking processes, and highly likely that the perception of the trier of fact will be unconsciously influenced by the transcript, with consequences for their evaluation of the significance of the audio in relation to the overall verdict.

The effect is that transcribers provide a ‘view’ of the audio evidence that may unwittingly influence those who believe themselves to be reaching an independent interpretation – much in the way that eye witnesses can be unwittingly influenced by others’ descriptions of what they have seen (Loftus and Palmer, 1974).

This section turns to consideration of how to resolve these problems. To start, it may be worth noting one potential solution that is unlikely to be effective: letting the trier of fact hear indistinct covert recordings with no transcript. While this clearly reduces the direct influence of the transcript on a trier of fact’s perception, it certainly does not ensure they will hear the audio accurately, due to the strong effect of the context itself on perception – demonstrated by Fraser and Stevenson (2014). Listeners need a reliable transcript to hear indistinct audio properly.

The question that remains is: how to ensure the transcript provided is reliable. In particular, are there practical ways to limit the disadvantages of using police transcripts for this purpose, while still retaining their potential to provide intelligence essential to the solving of crimes?

In fact, it may be relatively easy to achieve this desirable outcome through close collaboration between phonetic science and the legal system, keeping in mind the distinctions outlined above. This section outlines some suggestions that emerge from the earlier discussion, and may provide a starting point for such collaborative research.

Distinguish clearly between investigative and evidentiary uses of covert recordings

Perhaps the most important recommendation is to recognise the major difference in transcripts used for investigative purposes, as opposed to those used for evidentiary purposes. As long as indistinct covert recordings are used only for investigative purposes, little harm and much advantage can accrue from reliance on interpretations of involved transcribers, even if they are untrained in transcription.

The dangers described earlier arise when police transcripts are used for evidentiary purposes. Here, then, it can be recommended that the barriers to moving indistinct covert audio with police transcripts from investigative use to trial evidence should be far higher than is currently common. It may be worth emphasising that this is true even if perception of some words in the transcript may have been shown to be accurate by their value during investigative stages of the case. While this may lend credence to the transcription of those particular words, it is no guarantee of overall reliability of the transcript.

Decrease reliance on indistinct audio as evidence

The next recommendation is an overall reduction in use of indistinct audio as evidence in trials. In some cases, especially with shorter utterances, detailed analysis can allow experts in forensic transcription to provide a reliable transcript. However, in many other cases, even expert analysis has inconclusive results. This indicates the audio is simply untranscribable. It has a status similar to that of a smudged fingerprint, or inconclusive DNA results. The appropriate response, with audio as with other kinds of forensic evidence, is to exclude the evidence from the trial – not to admit it, with a police interpretation, ‘for the jury to decide’.

While exclusion of indistinct audio may be frustrating at first, it might have the advantage of greater attention being given to ensuring that covert recordings are obtained in such a way as to ensure the speech is clear.

For evidentiary uses, insist that covert recordings be (re-)transcribed by an independent, professional transcriber

Where indistinct covert recordings are admitted as evidence, it is essential that they should be accompanied by a reliable transcript. A first step is to ensure the transcript was produced by a transcriber who is independent of the case, with no stake in its outcome, and minimal contextual knowledge, and no influence from seeing a police version. However, the value of gradually providing specific contextual knowledge to the transcriber through a process of ‘sequential unmasking’ (Thompson, 2011) is a topic of current research by the present author.

Any disputes regarding words heard in the transcript should be resolved before the audio is admitted as evidence, through evaluation of the audio by a genuine expert in forensic transcription. Naturally this expert should also be independent of the case, in line with requirements increasingly being enforced in other branches of forensic science (Edmond and San Roque, 2012).

Present indistinct audio to the trier of fact in a way that enables reliable evaluation

Indistinct audio evidence should be prepared so as to make realistic demands of listeners. That means, for example, restricting the overall amount of audio (in one trial known to the author, the court had to listen to covert recordings for more than six days straight); dividing it into short sections containing coherent parts of conversations; providing headphones and software allowing replay at will; and allowing time for it to be listened to carefully – i.e. many times longer than the duration of the recording (cf. French and Stevens, 2013).

It also, of course, means presenting the audio with a reliable transcript, laid out in a way that assists listeners to follow the speech and find their way around the recording, enabling them to pay attention to intonation, tone of voice and other aspects affecting interpretation of the speakers’ intentions. Finally, reading or quoting from the transcript by barristers should be discouraged (Haworth, 2010), and emphasis placed on the need for interpretation of the evidence to be based on listening to the audio.

Conclusion

This paper has demonstrated a paradoxical situation in the legal process, whereby audio of worse than usual clarity is subjected to transcription practices of less than usual rigour.

It is notable, for example, that relatively clear, overt recordings, such as police interviews, are transcribed with more accountability than indistinct covert recordings. A range of resulting problems has been discussed, and directions for solutions suggested on the basis of well established research in the linguistic sciences.

A first step in solving these problems is increased publicity for general issues such as those raised in this paper. Unfortunately, it may not be sufficient for experts in forensic transcription to wait to be asked for assistance with specific disputes in individual recordings. In many cases, arguably those with the worst audio and the most unreliable transcripts, the legal process relies on its own checking procedures, not realising their inadequacy.

Ultimately, there is a need for collaborative research between the phonetic sciences and the law in developing evidence-based practices to ensure indistinct covert recordings used as evidence in criminal cases are accompanied by reliable and usable transcripts. It is hoped the present paper may spark interest in this kind of initiative.

Acknowledgments

This is a modified and expanded version of a paper presented at IAFPA, Zurich, 1-3 September 2014. It has benefitted from useful discussion with colleagues at that meeting, and later at *Theories, Practices and Instruments of Forensic Linguistics*, Rome, 1 December 2014, as well as from comments by Jose Antonio Mompeán González, and two reviewers. Of course, the views expressed are those of the author and all responsibility for errors or omissions rests with her.

Notes

¹This paper is based mainly on experience with the Australian context, where covert surveillance is governed by a variety of legal instruments, including the *Commonwealth Surveillance Devices Act 2007* and related State-based legislation (see Australian Law Reform Commission 2008). It is hoped that parts of the discussion may also be relevant in other jurisdictions.

²Covert recordings can be divided into two broad classes: telephone intercepts and environmental recordings (made with a microphone placed in the environment of the speaker). This paper deals mainly with environmental recordings, though some of the remarks may be relevant to telephone intercept material, where this is of poor quality.

References

- Australian Law Reform Commission, (2008). *For Your Information: Australian privacy law and practice*. Commonwealth of Australia: ALRC Report 108.
- Bonifaz, S. (2014). The effect of priming awareness when listening to disputed utterances. Master's thesis, University of York.
- Bruce, D. (1958). The effect of listeners' anticipations on the intelligibility of heard speech. *Language and Speech*, 1, 79–97.
- Bucholtz, M. (2007). Variation in transcription. *Discourse Studies*, 9(6), 784–808.
- Cauldwell, R. (2013). *Phonology for Listening*. Birmingham: Speech in Action.
- Coulthard, R. and Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. London and New York: Routledge.
- Cox, F. (2012). *Australian English Pronunciation and Transcription*. Cambridge: Cambridge University Press.
- Cutler, A. (2010). Abstraction-based efficiency in the lexicon. *Laboratory Phonology*, 1(2), 301–318.

- Cutler, A. (2012). *Native Listening*. Cambridge, MA: MIT Press.
- Eades, D. (1996). Verbatim courtroom transcripts and discourse analysis. In H. Kniffke, Ed., *Recent Developments in Forensic Linguistics*. Frankfurt: Peter Lang.
- Edmond, G. and San Roque, M. (2012). The cool crucible: Forensic science and the frailty of the criminal trial. *Current Issues in Criminal Justice*, 24(1), 51–64.
- Edwards, J. (2008). The transcription of discourse. In D. Schiffrin, D. Tannen and H. E. Hamilton, Eds., *The Handbook of Discourse Analysis*. Oxford: Blackwell Publishing Ltd.
- 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.
- Fraser, H. (2015). Transcription of indistinct covert recordings used as evidence in criminal trials. In H. Selby and I. Freckelton, Eds., *Expert Evidence*. Thomson Reuters.
- Fraser, H. and 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. *International Journal of Evidence and Proof*, 18, 205–229.
- Fraser, H., Stevenson, B. and 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.
- French, P. (1990). Analytic procedures for the determination of disputed utterances. In H. Kniffke, Ed., *Texte zu Theorie und Praxis Forensischer Linguistik*. Tübingen: Niemeyer.
- French, P. and Harrison, P. (2006). Investigative and evidential applications of forensic speech science. In A. Heaton-Armstrong and E. Shepherd, Eds., *Witness Testimony: Psychological, investigative and evidential perspectives*. Oxford: Oxford University Press.
- French, P. and Stevens, L. (2013). Forensic speech science. In R. Knight and M. Jones, Eds., *Bloomsbury Companion to Phonetics*. London: Continuum.
- Green, J., Franquiz, M. and Dixon, C. (1997). The myth of the objective transcript: Transcribing as a situated act. *TESOL Quarterly*, 31(1), 172–176.
- Haworth, K. (2010). Police interviews as evidence. In M. Coulthard and A. Johnson, Eds., *Routledge Handbook of Forensic Linguistics*. Abingdon and New York: Routledge.
- Heselwood, B. (2013). *Phonetic Transcription in Theory and Practice*. Edinburgh: Edinburgh University Press.
- Innes, B. (2011). R v David Bain: A unique case in New Zealand legal and linguistic history. *International Journal of Speech, Language and the Law*, 18(1), 145–155.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. Lerner, Ed., *Conversation Analysis: Studies from the first generation*. Amsterdam: John Benjamins.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar Straus Giroux.
- Kerswill, P. and Wright, S. (2008). The validity of phonetic transcription: Limitations of a sociolinguistic research tool. *Language Variation and Change*, 2(3), 255–275.
- Kreiman, J. and Sidtis, D. (2011). *Foundations of Voice Studies: An interdisciplinary approach to voice production and perception*. Oxford: Wiley-Blackwell.
- Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Loftus, E. and Palmer, J. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 585–589.
- Miller, G. (1951). *Language and Communication*. New York: McGraw-Hill.

Fraser, H. - Transcription of indistinct forensic recordings
Language and Law / Linguagem e Direito, Vol. 1(2), 2014, p. 5-21

- Mills, L. (2010). *Jump-Start Your Work at Home General Transcription Career: The Fast and Easy Way to Get Started!* Amazon.
- Ridley, A., Gabbert, F. and La Rooy, D. (2013). *Suggestibility in Legal Contexts*. Oxford: Wiley-Blackwell.
- Shockey, L. (2003). *Sound Patterns of Spoken English*. Oxford: Blackwell.
- Shuy, R. (1993). *Language Crimes: The use and abuse of language evidence in the courtroom*. Oxford: Blackwell.
- Sidnell, J. and Stivers, T. (2012). *The Handbook of Conversation Analysis*. John Wiley & Sons.
- Thompson, W. (2011). What role should investigative facts play in the evaluation of scientific evidence? *Australian Journal of Forensic Sciences*, 43(2–3), 123–134.
- Wilson, R. (2013). Warning in Patel case highlights court transcript weakness. *Sunshine Coast Daily*, Available from: <http://www.sunshinecoastdaily.com.au/news/warning-patel-case-highlights-court-transcript-wea/2050544/>.
- Wray, A. and Boomer, A. (2013). *Projects in Linguistics and Language Studies*. London: Routledge, 3 ed.