
Forensic Transcription

Helen Fraser

helenfraser.com.au

Text of a presentation at
Expert Evidence Conference
National Judicial College of Australia and ANU College of Law
12-13 Feb 2011

this text and powerpoint slides also available at
<http://helenfraser.com.au/forensic/#section-2>

Summary

The legal system has effective systems in place to ensure the reliability of transcripts of court proceedings, police interviews, etc. Unfortunately, the systems available to ensure the reliability of transcripts of covert recordings used in evidence (forensic transcripts) are less effective. As a result, it happens not infrequently that poor quality transcripts come to be accepted as reliable evidence.

This paper aims to assist in improving this situation by providing background information about six factors affecting the reliability of transcripts, and suggesting several easy-to-implement recommendations for minor reform in the legal system. Principal among these are: the need to curtail the increasingly common practice of allowing police to validate their own transcripts in the role of 'ad hoc expert'; and the need to routinely evaluate all forensic transcripts against a simple check-list at the voir dire stage.

1. Some preliminary definitions

Phonetics is the scientific study of speech. **Cognitive phonetics** is the scientific study of the complex interplay of mental, physical and social processes that underpin the skills of acquiring, producing, perceiving and representing speech. Though the two are closely related and interdependent, the focus of cognitive phonetics is more on the skills than on speech as such.

To use plainer language, cognitive phonetics is about **speaking and listening, spelling and pronunciation – all components of transcription**, the topic of today's talk, a brief summary of more lengthy treatments in Fraser (2003 and 2010).

Transcription is the representation of speech with written symbols, to create a stable record of what was said. Transcripts are commonly used as official court or

parliamentary archives, and are also the basis of most kinds of linguistic analysis, since it is hard to analyse speech directly, even if electronically recorded.

Forensic transcription is the representation of speech captured in a **covert or surreptitious recording** (for example via a telephone intercept, listening device or bodywire) for use as evidence in a legal investigation.

2. Overall theme

The legal system has extensive systems to ensure that transcripts of court proceedings, police interviews, etc. are reliable. Of course these systems are not infallible – many in the legal professions can cite examples, entertaining or otherwise, of inaccurate transcription – but they are generally adequate to their purpose of providing a written record of spoken transactions.

The systems available to ensure the reliability of forensic transcripts, however, seem to be far less effective. Thus even in my own limited experience I have encountered an unacceptably large number of instances of poor quality transcripts being treated as reliable evidence, at considerable cost in time, expense, and, ultimately, the potential for miscarriage of justice. Indeed some aspects of the legal system seem to actually promote poor practice. These include: reliance on so-called **ad-hoc experts** to verify transcripts, use of the term **aide-memoire** for transcripts, and insufficiently rigorous questioning of the provenance of transcripts at the **voir-dire** stage of trials.

One reason may be that a good deal of what seems in our society like ‘common sense’ about speech and writing is significantly inaccurate. It is hoped that providing a better understanding of cognitive phonetics – not its technicalities, but its **conceptual framework** – might enable common sense to be used more appropriately in judging the reliability of transcripts.

3. What is transcription?

To ‘transcribe’ originally meant to ‘write across’, in the sense of **copying a written text** by hand from one sheet to another (in the days before printing). The current

meaning has evolved via several extensions from this original meaning, which for convenience we can refer to as **transcription-1**.

The first extension of meaning (**transcription-2**) occurred when the word was used to refer to the process of creating a written record of a meeting, interview, etc. by 'writing down' what was said. Though this may seem like a similar process of 'copying', however, there is a major difference between 'copying across' a written text and 'copying down' spoken language. **Where writing stays still to be referred to during copying, and checked afterwards, speech is fleeting.** This means that making a written record of speech relies heavily on memory. For speech more than a few minutes in duration it is essential to take notes, as an 'aide memoire'.

This makes transcription-2 very different from transcription-1. Its reliance on memory means it necessarily involves **subjective interpretation**¹. For example, even with the best of good faith, there can be significant differences in the memory of various participants about what was said at a meeting. In recognition of this well-known fact, it has become common practice to ensure that transcripts of meetings (now called 'minutes') are produced by an impartial transcriber, and confirmed by all participants.

The subjective interpretation involved in transcription-2 opens the **opportunity for bad faith**, as was clearly seen some decades ago with widely publicised police 'verbals' – inaccurate or misleading transcripts of interviews with suspects. In response, it is now standard practice to make electronic recording of police interviews. Such recordings can be sent for professional transcription to produce a reliable written record suitable for use in court.

Indeed **'copying down' of speech from an electronic recording** (whether of a police interview, radio program or other audio) has become the most common meaning of the word *transcribe*. However, this usage involves a significant extension to the meaning of the word, worthy of another superscript: **transcription-3**. Most importantly, a transcript-3 is typically produced by a professional transcriber who was not present at the meeting. So, though it may be more convenient to refer to in court

¹ It is sometimes wrongly thought that short-hand provides an objective record of speech, but even short-hand is really only notes used as an aide-memoire – albeit more detailed than notes taken in normal writing. This is seen by the fact it is difficult for one stenographer to copy up another's notes. Indeed even one's own short-hand must be written up soon after the event, before memory fades.

than the recording itself, and thus certainly an aid, it is not an aide memoire in the same sense as the minutes of a meeting are an aid to the memory of participants.

4. Forensic transcription

The next extension of meaning arose through another use of audio recordings increasingly deemed essential to the legal system. **Covert recordings** provide a valuable source of evidence in many cases. The problem is, due to the manner of their recording, their quality can be rather poor – sometimes throughout, sometimes in one crucial section, giving rise to a ‘disputed utterance’, in which there is disagreement about the exact words spoken.

The natural response is to **send the poor quality recording to an expert**. The result is a **forensic transcript**. Unfortunately a forensic transcript can look very similar to a professionally prepared transcript of a police interview, making it tempting to treat it as equally convenient and reliable – or even more so, since it is now backed by an ‘expert’. But is it equally reliable, and can it be used in similarly convenient ways? Sometimes, but sometimes not. As stated in the introduction, **there have been numerous incidences of forensic transcripts being treated as reliable evidence when in fact they were inaccurate and misleading**.

In moving to discussion of the forensic transcript, or transcript of a covert recording which forms evidence in a trial, **we have now entered into a new meaning for the word *transcript*, which we can designate as transcript-4**.

Though the same word is used, there are many differences between a transcript-3 (e.g. electronic record of police interview of a person of interest) and a transcript-4 (e.g. conversation recorded by a hidden listening device in the home of a person of interest). The quality of recording is just the first of **six factors** affecting the reliability of the transcript that often differ substantially between transcript-3 and transcript-4:

1. recording factors
2. speech factors
3. contextual factors
4. listener factors

5. transcriber factors
6. situational factors.

In the remainder of this talk I will look at each of the 6 factors, starting with 2, *speech factors* and returning at the end to discuss 1, *recording factors* in more detail. In surveying the six factors, I will be intertwining two parallel arguments.

1. Transcript-3 (e.g. transcript of police interview) and transcript-4 (e.g. transcript of covert recording), though they are referred to with the same word, are actually two quite different kinds of entities. They typically differ not just in recording quality but in all six factors mentioned above (and discussed in more detail below).

2. Both transcript-3 and transcript-4 have far more in common with transcript-2 (e.g. minutes of a meeting written up from personal notes) than is usually realised. In other words, 'having the audio' can be an advantage, but does not make the transcript 'objective'.

I believe the key to improving the reliability of forensic transcripts (i.e. transcript-4) lies in understanding how the six factors work together to create transcripts of varying degrees of reliability. Once this is well understood, ensuring good practice to produce reliable transcripts can be a matter of common sense. Unfortunately, however, at the moment, though aspects of each individual factor are well known, they are not always put together in a way that makes their implications clear. The main reason for this is our society's poor knowledge of phonetics, which allows widespread misconceptions about the nature of speech to go unchallenged.

5. Widespread misconceptions about the nature of speech

Many people assume speech is like an auditory version of a printed text – a sequence of discrete, invariant sounds (phonemes) grouped into words of standard form, each separated by a short pause. It is this belief which allowed the meaning of transcript-1 (copying across a written text) to be extended to cover transcript-2 (copying down the speech at a meeting).

In fact, nothing could be further from the truth. Speech is really a continuous flow of sound. For present purposes, it is useful to compare it not to print but to handwriting

– as long as it is noted that speech is like handwriting lacking spaces or punctuation, and that even the clearest of speech is like extremely messy handwriting².

As with handwriting, speech is highly variable from person to person and context to context. Thus one set of factors affecting the reliability of a transcript is the clarity of the speech itself – the speech factors.

6. Speech factors

In a ‘normal’ transcript-3 (such as a transcript of a police interview), the speech is relatively clear. The speakers are aware they are being recorded, and are monitored by the interviewer, who may ask others to ‘speak up for the tape’. Even so, the recording can be very hard to transcribe. Anyone interested in this topic who has not tried to transcribe a recorded interview is urged to give it a go (an informal radio interview makes an acceptable substitute if no police interview is available), as it is extremely instructive to realise how unclear even carefully monitored speech can be.

This highlights the many differences between spoken and written language, a topic knowledgeably and entertainingly discussed by Michael Erard in his book *Um: Slips, stumbles and verbal blunders, and what they mean*. Erard emphasises that, though it is often assumed the features of conversational speech are the result of ‘sloppiness’, most of them actually aid comprehension. Speech and writing are not different representations of one message, but two genuinely different modes of communication, as is easily observed by noting how hard it is to read transcribed speech, or listen to a written text read aloud. Thus even for a police interview, making a transcript usefully legible takes a good deal of grammatical ‘tidying up’, which is part of the skill of the transcriber.

With a forensic transcript, the speech is by definition not monitored for the tape (participants being unaware they are being recorded). Typically it is spontaneous conversational speech, very different from the monitored speech of transcription-3.

² It may be worth mentioning for those interested that the handwriting analogy, while more useful in the present context than the print analogy, is still significantly imperfect, in that handwriting can be validly considered to be a distortion of a discrete, invariant printed form. Though some early theories of speech perception and production suggested that speech ‘output’ is a distorted version of an underlying discrete string of invariant phonemes, most current theories find problems with this view, and prefer other ways of conceptualising speech than as an unintended ‘distortion’ of an ‘ideal’ form.

Spontaneous conversational speech is highly context bound. To follow the handwriting analogy, it is like scribbled personal notes (lacking spaces or punctuation). In addition, there are frequently several intertwined voices, whether from other participants in the conversation interrupting or overlapping the speaker of interest, or from nearby conversations, live or on television or radio.

For all these reasons, covert recordings are generally extremely difficult to transcribe, even when the technical quality is good, and even when the transcriber was present and found the conversation clear at the time. This raises the interesting question of ‘How do we ever manage to understand each other?’, and leads into the next set of factors that affect the reliability of a transcript.

7. Context factors

in listening to speech – as indeed in reading handwriting, or even print – we rely very heavily on context. It is well known that comprehension depends only partly on **‘bottom-up’ information**, i.e. information from the signal, and also requires **‘top-down’ information**, i.e. information from the context. These two kinds of information interact in a **trading relationship** – the clearer the bottom up information, the less context is needed; the more context available, the less attention is paid to bottom up information – but every act of speech perception necessarily involves both.

Top-down contextual information can be usefully divided into two kinds. **External context** is the listener’s knowledge of what is going on in the situation in which the speech is heard. **Internal context** is the listener’s knowledge of the language itself. In conversation, participants constantly predict, from internal and external context, what others will say, and use bottom up information to confirm or refute the predictions.

This points to the importance of **duration and continuity** of a recording as factors affecting the reliability of a transcript, and again these can differ greatly between ‘normal’ and ‘forensic’ transcripts. A police interview, for example, typically lasts for at least 10 or 20 minutes, and shows considerable continuity (i.e. the conversation is roughly on a single continuous topic, with an orderly sequence of turns). A covert recording on the other hand can involve short bursts of conversation (especially if

poor recording values mean the speech is frequently drowned out) and the topic can vary dramatically from one section to another.

8. Listener factors

The need to use context and top-down processing in speech perception shows the essential role of the listener in making speech seem 'clear'. The ultimate demonstration of this is the fact that speech in a foreign language sounds like an indecipherably rapid stream of sound to us, but perfectly clear to someone who speaks the language.

Listeners' knowledge usually assists perception. This is seen in the well-known case where fluent readers' knowledge of the internal context allows them to easily understand a text even if the letters are significantly garbled (see powerpoint p.18). However, **top-down knowledge can also mislead**. This is seen in the equally well known case where a reader who knows or guesses what a text is supposed to say is liable to miss significantly egregious typographical errors.

Knowledge of external context can be even more misleading, as shown in a classic experiment. In 1958, Donald Bruce played a number of sentences to participants through masking white noise. The sentences were of the 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

He then played the same sentences through the same white noise, this time preceded by one of several 'hints', such as 'sport' or 'weather' (appropriate for the sentences above) or 'food', 'travel', and others appropriate for other sentences. He was expecting to find that the hints made their appropriate sentences easier to understand, but was surprised to find they also affected the perception of inappropriate sentences. For example, listeners heard Sentence 1 above quite differently depending on which hint they received (shown in brackets).

Sentence 1 (food)

I tell you that I feel more hungry than you are

Sentence 1 (travel)

I tell you that I too will leave next year.

Crucially, the sentences heard with a misleading hint were perceived with just as much confidence as those heard with the appropriate hint.

Since then, this phenomenon has been investigated from a wide range of perspectives, and confirmed again and again as one of the most robust findings of psycholinguistics.

As in many areas of phonetics and linguistics, it is well known to researchers that confidence of perception correlates poorly with accuracy of perception – to an even greater degree with speech than with visual perception.

9. Transcriber factors

So far we have discussed the skill of listening to speech. Transcribing speech requires a whole additional set of skills. It is worth recalling that transcription is an ‘unnatural act’. Over the span of millennia, speaking, listening, reading and writing have developed naturally for use in meaningful communication. Transcription is a very recent development and uses a very different set of skills, requiring extreme patience and perseverance. Most people who sit down to transcribe a recording are not only surprised to discover how hard it is, but actually produce a significantly inaccurate transcription. Professional transcribers undergo rigorous training and induction, and need to pass stringent tests of accuracy, and conventions of presentation, in order to be accredited.

However, even the best transcript by a well qualified and experienced professional courtroom transcriber cannot be said to be ‘objectively accurate’. Any transcript involves abstraction from the vast quantity of acoustic information available in speech, and transcriber judgment plays a significant role in deciding what is appropriate to include and omit in the particular situation the transcript is intended for. This leads into the sixth and final set of factors, those concerning the situation in which the transcript is created and used.

10. Situational factors

The purpose for which a transcript is to be used dictates the level and type of detail that is required. The 'verbatim' transcript typically used for police interviews is just one kind among many. Significantly, even the best of these 'verbatim' transcripts are not considered sufficiently accurate by linguists looking to study courtroom interactions. Such linguists generally require a far more detailed representation of the discourse, which may, depending on the nature of the research being carried out, include marking the duration of pauses, the exact phonetic character of particular utterances, and details of every slip of the tongue with its revisions or recasts.

Does this mean that court room transcribers should 'lift their game' and produce transcripts suitable for discourse analysis? Of course not. Linguists' transcripts are useful in their own situation but would be extremely frustrating to use in a court. They trade legibility for accuracy in representing details that are quite irrelevant for the purposes for which court room transcripts are used. The point is that **the situation in which a transcript is made affects judgement of its accuracy and reliability.**

Perhaps the most crucial situational factor affecting the reliability of transcripts in the legal context is the degree to which it can be checked. With transcripts of court or police proceedings, for example, the speaker(s) being transcribed can be given the recording and invited to raise objections regarding the way their words have been represented. With forensic transcripts, there tend to be reasons why such critiques cannot be relied upon.

In such cases, it is essential to have some other way of verifying the transcript. One useful method is to check the transcript against **external evidence** about the case that could not be known to the transcriber except from the recording. For example, a transcriber's ability to accurately represent names, places, events and other facts of a case lends credibility to the transcript. Thus the extent to which the transcriber has other means of knowing those facts (besides hearing of them in the recording) can be a significant factor in ascertaining the reliability of a transcript.

11. Recording factors

We return now to the first factor affecting the reliability of a transcript, the quality of the recording. Audio quality is frequently given all the blame for difficulty of transcription, with recordings often sent straight off for 'enhancement'. However, as we have seen, recording quality may not be the only problem. For example, police interviews often have rather poor recording quality, yet the speech can usually be transcribed satisfactorily because the other factors are favourable. On the other hand, a forensic recording can be very hard to transcribe even if the recording quality is technically quite good. However, since recording conditions of listening devices and other covert techniques is rather difficult to control, it does frequently happen that forensic recordings are of poor quality, adding greatly to the difficulties caused by unfavourable conditions in the other factors. In these cases, it is important to be able to distinguish among various types of 'poor quality' and to know their effects and how best to compensate for them. In particular, it is important to understand which types of degradation yield to which types of 'enhancement', and which do not.

12. What does it all mean?

The upshot of the preceding discussion is that deciding *what is said* is as much an **interpretation** as deciding *what it means in a particular context*. The difference is that we have far greater conscious awareness of and control over 'higher-level' interpretations of contextual or pragmatic meaning.

In this, speech perception is entirely parallel to many other aspects of perception and cognition. Chris Firth's book *Making up the Mind: How the brain creates our mental world* can be recommended as an easy-to-read but reliable account of this phenomenon. One of his arguments is that we tend to ascribe perception, especially auditory, wholly to the objective stimulus, ignoring our own subjective contribution.

The upshot is that in any transcription, what is transcribed is a judgment. The implication is that whether a transcript is 'accurate' or 'reliable' is not a fact but an opinion, and needs to be handled as such by the court. It is failure to recognise this that underlies most examples of poor practice in handling forensic transcripts.

Ad hoc expert

It has become increasingly common for police to be allowed to not only create but also evaluate and verify their own transcripts of forensic recordings in court, especially where these are hard to hear – on the grounds that their extensive familiarity with the voices gives them ‘expertise’ in understanding what they are saying. However, though police may have a good ability to understand voices with which they have become familiar through extensive listening, this ability is not expertise in the usual sense. Indeed, **the term ‘ad hoc expert’ is an oxymoron**, as real expertise requires in depth study well beyond one specific ‘ad hoc’ situation.

Further, from all that has been said above, **the familiarity with the voices that is such an advantage in understanding the speech, can equally be seen as a bias (albeit unconscious), with the potential to undermine objectivity**. Police confidence in their own transcripts, unfortunately, is no guarantee of accuracy. Again this is not to impugn police integrity in any way. It is simply a fact about speech perception well known to any genuine expert on the topic.

Police and other ‘interested parties’ can be of great assistance in creating forensic transcripts but the transcripts they produce must be evaluated for the court by a genuine expert.

Audio engineers

It has been common in recent years for hard-to-hear recordings to be sent to an audio engineer for expert transcription. Audio engineers may be genuine experts on ‘audio’, or sound in general. However, while speech is undoubtedly ‘sound’, it is a very particular kind of sound, whose understanding requires very specific training in phonetics and linguistics.

Though audio engineers can produce impressive technical descriptions of the sound of speech, they typically know little more than the average person about the specifics of phonetics, and even less about cognitive phonetics. For this reason, audio engineers should not be used as ‘experts’ in relation to speech and language evidence.

Phonetics and cognitive phonetics

Phonetics has several different branches, and it is important to choose an expert with the appropriate background for the task at hand. Acoustic phonetics can provide very useful evidence in cases of 'disputed utterance' by significantly refining the available 'bottom up' information. As we have seen, however, 'bottom up' information is only part of the picture. This can be easily demonstrated by following the handwriting analogy. A fragment of handwriting (similar to the fragment of speech in a 'disputed utterance' case) can be subjected to extensive analysis without necessarily revealing its true interpretation – simply because the latter requires the context that is missing from the fragment (see accompanying powerpoint, p.25).

It is essential to distinguish between creating a transcript and evaluating a transcript. For the latter task, an expert in cognitive phonetics, with a solid understanding not just of acoustics but of all the six factors discussed above, is necessary.

The most important function of an expert in cognitive phonetics may be that of declaring a recording or part thereof to be 'untranscribable', and unsuitable for use in a courtroom situation.

Aide memoire

A common response to my raising of these issues has been reassurance that in the legal system a transcript is merely an 'aide memoire', with the actual recording being the official evidence. From all that has been said so far, however, it is clear that a transcript as usually used in the legal system is not an aide memoire in the usual sense of that term. Such transcripts are typically made by someone other than those using them in the court (judge, jury, barristers, etc.) so reflect someone else's memory, and, crucially, interpretation, of what is said in the recording. Though this may be understood in principle by court-room personnel, it is well known that the words used to describe an entity reflect and affect subconscious understanding of the nature of that entity. It is best to use appropriate terminology wherever possible, and 'aide-memoire' is really not an appropriate term for a transcript, especially a forensic transcript.

13. Conclusion and recommendation

It is important to be very clear what I am not recommending here. **There is no suggestion that every forensic transcript must be evaluated in detail by an expert in cognitive phonetics.** In by far the majority of cases, evaluation is a straightforward matter that can easily be handled by court-room personnel.

The essential thing is to catch those cases that are not straightforward, and ensure that those are handled with advice from a genuine expert in cognitive phonetics.

Achieving this could be as simple as ensuring that all forensic transcripts are subjected at voir dire stage to a simple checklist based on the six factors discussed briefly above and in greater detail in Fraser (2010).

The best way to create, disseminate and monitor use of such a checklist would surely be for representatives from the legal, judicial, police and academic communities to collaborate in a working party.

14. References

- Erard, M. 2007. *Um: Slips, stumbles and verbal blunders, and what they mean*. New York: Random House.
- 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. 2010. Transcripts in the legal system. In I. Freckelton & H. Selby (Eds.), *Expert Evidence (Chapter 100)*. Sydney: Thomson Reuters.
- Frith, C. 2008. *Making up the Mind: How the brain creates the mental world*. Oxford: Blackwell.