100
Transcripts in the legal system

by

Helen Fraser

Fraser, Helen. 2010. Chapter 100. Transcripts in the legal system. In Expert Evidence, eds. Ian Freckelton and Hugh Selby. Sydney: Thomson Reuters.

Author information

Helen Fraser is a specialist in cognitive aspects of speaking and writing. She studied linguistics and phonetics at Macquarie University and the University of Edinburgh, and taught at several universities in the UK and Ireland before taking a lecturing position at the University of New England, NSW, which she held for 18 years. She is now an independent consultant, focusing on applications such as second language pronunciation, literacy acquisition, and forensic transcription.

Contact: helenfraser.com.au; HelenBFraser@gmail.com

Acknowledgements

This chapter has been a long time in the writing and has benefited from discussion and comments from many people, to all of whom I am grateful.

I mention particularly: Dominic Watt, Asfa Bhagdin, Philip Rose, Diana Eades, Roger Shuy, Hugh Selby, Suzanne Mendra, Stephen Keim.

Naturally any errors of fact or judgment remain my sole responsibility.

[The next text page is 100-51]

[The next text page is 100-1051]

INTRODUCTION

[100.100] Transcription plays an important role in many parts of the legal process. For example:

- transcripts provide a lasting public record of courtroom proceedings;
- transcripts provide a convenient reference to evidence gathered via formal processes, such as recorded police interviews;
- transcripts provide interpretation of evidence consisting in surreptitious recordings, such as telephone intercept or listening device product.

Such artifacts are all called "transcripts", but there is a significant difference in their status. On the one hand, the accuracy of courtroom transcripts is accepted as a cornerstone of the legal process, seldom questioned by either defence or prosecution. On the other, the correct transcription of audio evidence can be the subject of vigorous but ultimately unresolvable debate. Consider, for a famous example, David Eastman's whispered soliloquy, recorded by a listening device in his house after the 1989 shooting of Assistant Commissioner Colin Winchester in Canberra. Did it contain the words "I killed Winchester", or was it rather "I kept watching her"? There is no way to be absolutely certain.

Between these two extremes lie many points on a long continuum of accuracy and verifiability. In using transcripts of various kinds, it is clearly desirable that they be treated appropriately according to their location on that continuum. Unfortunately, this is not always the case.

While courtroom transcripts are produced and verified according to rigorous procedures developed over several centuries, justifying the high level of credence accorded them, there are few standard procedures for creating and verifying transcripts for use in evidence. Worse, there are no standard procedures for resolving disputes about transcripts once they have been brought as evidence. Consequently it not infrequently happens that inaccurate or unverifiable transcripts of barely audible recordings are treated as reliable evidence. Indeed it seems some current practices in the legal system may operate not to control but to facilitate admission of unreliable transcripts as evidence. This chapter aims to address this issue.

In order to make the necessary distinctions among transcripts of varying degrees of reliability, it is first necessary to set up a framework within which the creation and use of transcripts can be understood. That is the task of the first part of the chapter. Later sections consider factors that affect the reliability of transcripts, and discuss some of the procedures that are, should, or should not, be followed in creating and evaluating transcripts for use in court. Particular issues discussed include: the inadvisability of calling a transcript an "aide memoire"; the dangers of relying on transcripts produced by "ad hoc experts"; the importance of choosing appropriate disciplinary and subdisciplinary experts to advise on specific transcription-related issues; and, most importantly, the need for standard procedures to ensure that problems with transcripts are detected before they are tendered as evidence.

Please note: In the interests of readability by non-specialists, this chapter uses minimal referencing throughout, and supplies a classified bibliography at the end.

[The next text page is 100-2051]

WHAT IS A TRANSCRIPT?

[100.200] On the face of it, a transcript seems a straightforward thing, a simple matter of writing down what someone has said, to produce a written equivalent of the speech. One aim of this chapter is to show that this apparent simplicity masks considerable complexity: both "writing down" and "what someone says" are highly sophisticated concepts, and a transcript is never equivalent to the speech it represents.

We can start exploring this complexity by bringing to attention some familiar observations and experiences. Consider a scenario in which a conversation between two speakers, S1 and S2, is recorded at the scene by a third person, not part of the conversation but familiar with the speakers and the context. Consider now various tasks of "writing down what was said" that could be undertaken. Assume, to start with, that the conversation is a relatively formal one, conducted at normal volume in a quiet place, and the recording is of good quality. Assume further that all participants are competent speakers and writers of the language, motivated to engage with the tasks in good faith.

"Writing down"

[100.220] A first task might be for S1 and S2 to "write down what was said" from memory. This is a familiar activity: people often take notes to help them remember significant points from a conversation, as an "aide memoire". If taken a short time after the conversation, such notes can be expected to capture relevant aspects of the conversation with a certain degree of accuracy, to the extent they might have some standing in law. A record like this does have limitations, however, relating to limitations of human attention and memory. In particular, it captures only those aspects of the conversation the note taker remembers as significant at the time of writing. For this reason, we would expect some differences in the notes made by S1 and S2, though both related to the same conversation. The degree of difference could be manipulated by varying the two speakers' individual interpretation of the nature of the conversation and the purpose of the record. For example, S1 could be told the conversation formed evidence in an insurance claim, while S2 could be told it was merely an academic experiment. Recognition that individual accounts can vary in this way is the reason minutes of a meeting must always be agreed by all present, and can, in contentious meetings, become a topic of dispute.

Another task might be for S1 and S2 to write down their conversation "verbatim", ie in the exact words used at the time. This is much more difficult, and it is unlikely either speaker could do it to any level of detail even a short time after the conversation. At best the speakers could confer, to retrospectively prepare a written record which both agree to be an accurate representation of the conversation. Such a conference would likely bring to light a good deal of difference in the two speakers' individual understanding of what had been said. However, in a non-contentious case, these differences could likely be resolved through discussion, to produce a document, which, though unlikely to provide anything like an accurate record of the actual words spoken (thus not literally "verbatim"), could be accorded greater standing than an individual's aide memoire as a reliable record of the conversation.

This is the classic situation of the so-called "verbatim statement", used by police before the era of recorded interviews. Such a record is certainly useful in the absence of an audio recording,

but has a number of limitations, especially in requiring good faith on both sides. Some of these have been highlighted in the literature on "verballing", the fabrication of confessions or other evidence. However, even without considering deliberate misrepresentation, the "verbatim record", like the "aide memoire", is limited by factors of human attention and memory, and by the fact that it is necessarily a reconstruction rather than a record of what was actually said. If an important issue is later found to hinge on a memory of words having been spoken which happen not to have been written down at the time, it will not be possible to retrieve the facts of the matter. Disagreements will be one person's word against another.

It is tempting to assume such problems are resolved with the availability of an audio recording, and indeed in some cases they are. However, it is not so clear cut. Consider asking the third party to transcribe the recorded conversation. This is a task not often undertaken in the course of everyday life, and most people who attempt it for the first time are amazed to discover just how hard it is. Indeed, readers with a serious interest in the use of transcripts in court are urged to try transcribing for themselves. Even where there is no difficulty in hearing the conversation, transcribing it word by word is hard. It is not just painstaking and tedious but literally difficult to hear each individual word. Many difficult decisions have to be made, for example, how to represent false starts, "umms and ahhs", or speech errors, and how to punctuate the ungrammatical and partial sentences characteristic of conversation.

Consider now sending the recording to a professional transcriber, who we can call PT, in contrast to the amateur transcriber, AT, just considered. Since, by the assumptions of our scenario, the recording is relatively clear, transcribing it is a fairly straightforward task for PT, who is set up with appropriate equipment, has put aside an appropriate amount of time, and has been thoroughly trained in production of transcripts, including how to be consistent in decisions like those just mentioned, and when to transcribe as "unintelligible".

Imagine now that both transcripts, by AT and PT, are sent to S1 and S2 (the original speakers) for checking. PT's transcript is liable to be much easier to read, set out in a professional manner, using appropriate punctuation and other conventions, and showing an appropriate level of detail. In many cases PT's transcript will also be found more accurate. It is unlikely to be the case, however, that PT's transcript is more accurate in every respect. Even under the good recording conditions of our scenario, there are likely to be a number of cases where S1 and S2 find AT's transcript more accurate, especially if the conversation contained specialised vocabulary, unusual names, reference to specific events, or other points unfamiliar to PT.

The following excerpt from a professional transcript of a philosophical discussion shows some relevant examples, including the transcriber's unfamiliarity with words like "inductive" and "loci".

> What we're doing when we talk about natural concepts, we're talking about ways people sort stuff that turns out to be fruitful or not fruitful. I think the real contrast to the notion of natural kind is the notion of a nominal kind, the idea of a kind that's defined by whatever criteria we happen to choose. So if I have a kind that consists of all things that are either pencils or glasses of water, that's an arbitrary kind, they're not going to be any interesting explanatory laws or inducted [inductive] generalisations about those things. Whereas if I have either acids or late capitalist economies, or flightless birds, those are all going to be low sigh (?) [loci of] interesting properties that figure an explanation and induction.
>
> Transcript of discussion on radio program *The Philosophers Zone* Australian Broadcasting Corporation Radio National, 9 August 2008

Also surprising may be the number of cases where S1 and S2 themselves are unable to resolve parts found difficult by AT and PT. Though with the aid of their notes they can clearly recall the meaning and intent of their conversation, they may well not be able to reconstruct the exact form of words they used at the time, even with the recording to listen to.

Let us finish this section by considering a final stage of the scenario, in which a single "best version" transcript is agreed upon by all four participants (two speakers and two transcribers),

and sent, along with the recording, to two complete outsiders, X and Y, who need to discuss the content of the original conversation. Both are busy people, and want to extract from the conversation the information of most relevance to their upcoming meeting. X reads through the transcript, while Y listens to the recording, each making notes as a personal "aide memoire" to refer to in the meeting. Since this is a hypothetical scenario, it is not possible to state exactly what the outcome would be, but it is clear many factors could potentially affect the degree of agreement between X and Y as to the content of the original conversation. Even under the ideal conditions of this scenario, then, a transcript is not equivalent to the speech it represents.

"What someone says"

[100.240] In this section, we replay the scenario with everything held constant, except this time the recording is not clear. Even a slight change in any of the conditions can easily produce this effect.

One way to reduce the quality of the recording is to set the conversation between S1 and S2 in an environment with background noise. Little background noise is needed to severely affect the intelligibility of a recording. Even a conversation that seemed perfectly clear in context can be "hard to hear" when replayed from an audio recording. This is because of the so-called "cocktail party effect", in which features of the human visual, auditory and other perceptual systems enable participants at the scene to track a particular conversation among a hubbub of other speech much more effectively than is possible in a decontextualised recording.

Another way to make the recording less clear is to change the scenario from a formal conversation to a casual one. Even under good recording conditions, a casual conversation which is perfectly clear to participants can pose major problems of intelligibility when recorded and replayed out of context.

Alternatively, we could change the assumption that all the participants are competent speakers of the same variety of the same language. For example, let S1, S2 and AT all be speakers of a dialect unfamiliar to PT.

Consider the effect of any of these changes on the transcripts produced by AT and PT. The whole process of creating the transcript will become much harder for both of them than under the original assumptions, resulting in many more gaps, omissions and errors, only some of which will be able to be filled in by S1 and S2. Again, it is interesting to compare the gaps, omissions and errors in the two transcripts. To an even greater extent than before, these will likely differ, reflecting PT's greater skill and experience, and AT's greater knowledge of the conversation and its context.

This observation highlights the main point, to be taken up in more detail in the next section, that what makes speech "clear" is only partly the nature of the speech itself. It also depends largely on the perceiver, especially their knowledge or assumptions about what they are hearing. The role of knowledge or assumptions becomes especially evident if for some reason that knowledge is inaccurate. This is well-known in experience, and indeed by the legal system, but as we will see in what follows, can be lost sight of in certain circumstances.

Consider now removing the last assumption of our scenario – perhaps the most critical in the legal context: what if one or more of the participants was not acting in good faith, but intending to obstruct and deceive. Instead of transcribers and speakers reaching agreement, there is dispute. Which version is reliable? The only possible answer is, it depends. To elaborate on this answer, it will be useful to consider how the common observations brought to light by this hypothetical scenario are dealt with by the science of speech and writing.

[The next text page is 100-3051]

BACKGROUND

[100.300] The preceding discussion has made the point that "writing down what someone says", though it seems simple, is in fact a highly complex skill that depends on, but goes far beyond, two other highly complex skills: speech perception and writing (as well as attention, memory and other underpinning cognitive skills). Why would something that is actually very difficult be thought of as simple? This is the result of widespread misconceptions about the nature of speech and writing, which this section aims to address.

The nature of speech

[100.320] It is commonly believed that speech consists of a sequence of individual words, each separated from its neighbours by a short pause. In fact, speech is a continuous stream of sound. The process of speech perception involves listeners projecting words, learned via their socialisation into a particular language, upon this continuous stream of sound, in much the same way as seeing constellations involves projecting shapes, learned via socialisation into a particular culture, upon the continuous array of stars in the night sky. Without knowledge of the constellations, no constellations are seen, only the continuous array. Similarly, without knowledge of words, no words are heard. Listening to a foreign language gives an impression of the continuous nature of speech. It can be hard to recognise that one's own language is similarly continuous to those who do not know it, but this is in fact the case: what makes it seem to us like a sequence of discrete words is a characteristic of ourselves, not of speech.

If speech is really a continuous stream of sound, why do we not hear it as such? The point of listening to speech is not to hear the sound as such, but to access the meaning intended by the speaker. In classic linguistic terminology, speech is a signifier representing a signified. More simply put, it is sound representing meaning.

In listening to speech, therefore, we see "through" the signifier (sound) to our goal, the signified (meaning). In doing this, we pay minimal attention to the sound itself. That is why, as seen in the scenario, memory for the meaning of speech is generally much better than for the sound in which the meaning is couched. This is very much in line with other human perceptual systems, which all serve to reduce the "blooming buzzing confusion", as William James described the uninterpreted reality around us, to meaningful elements – our attention is generally on the meaningful elements, such as trees, people, furniture and so on, to the extent we barely recall the details of their physical characteristics. However, it is even more true of speech than of other perception. Speech is fleeting. As soon as a sentence is uttered, it exists only in memory. To study it objectively requires capturing it and analysing it – something most people never trouble with.

Interestingly, however, rather than acknowledging that they know nothing about speech, most people rely on assumptions and inaccuracies. The few who study linguistics, the scientific investigation of language, encounter many surprising and counterintuitive facts about speech and language. Indeed, teaching undergraduate linguistics consists as much in eliminating misconceptions as in providing reliable information. The next section looks at where some of the most pervasive of these misconceptions come from.

Writing

[100.340] Writing is often considered to be a representation of the sound of speech (the signifier). However, it is really a signifier in its own right – a different way of representing meaning, that has developed as an entirely separate medium of communication from speech. A small impression of the vast difference between writing and speech can be gained by comparing the transcript of the radio program, above, with the text of the present chapter.

Although the goal of writing is to communicate meaning, it does this by making reference to the spoken signifier. Each letter or symbol of a writing system refers to part of the signifier (not directly but in a complex manner we will not take space to consider here). The exact units referred to depend not just on the language, but also on the writing system. Most writing systems represent units approximately the size of a syllable. Alphabetic writing systems, such as the one used by English, represent (roughly speaking) individual sounds, or phonemes. People who have not learned an alphabetic writing system are unlikely to recognise individual "sounds" in the way that seems so natural to literate English speakers. They generally consider a syllable to be a unitary sound in its own right, rather than a combination of several smaller units of sound.

Learning a writing system means, of course, learning the symbols it uses to represent parts of the signifier. Much more importantly, it involves learning to conceptualise words as having parts which can be represented by symbols. What makes literacy difficult (to the point of inaccessibility for many) is not learning symbols for sounds, but learning to conceptualise words as a sequence of individual sounds to which symbols can be attached. Interestingly, however, as soon as this "phonological awareness" is achieved, learners forget how difficult it was, and consider literacy "as easy as ABC".

To help learners acquire phonological awareness, teachers tell them that words are made up of "sounds" or "phonemes" which can be represented with letters and "sounded out" into words. Of course, to become a fluent reader, it is essential to move beyond painstaking sounding-out, and learn to recognise words as whole units. As they do this, however, learners generally continue to believe they read letter by letter. This belief is completely erroneous. As with speech, our perceptual systems encourage us to see "through" the written signifier to its meaning – one reason that typographical errors are often overlooked. Research demonstrates very clearly that, in fluent reading and writing, the focus is almost entirely on meaningful words, and only minimally on individual letters. Humorous examples demonstrating this "Fascniatnig sutff" abound.

Nevertheless, the belief that words are made up of individual letters or sounds lives on, becoming so entrenched that it is difficult to hear words except as a sequence of phonemes. This creates a kind of "literacy bias" – a belief that speech is like printed text, composed of small units arranged into groups separated by short gaps, like the letters on this page. In fact, as we have seen, speech is a continuous stream of sound, and there is no clear correspondence between any parts of this stream and individual words or phonemes (a fact known in scientific phonetics as the "segmentation and invariance problem").

There is not space to elaborate on this explanation here (but see Fraser 2003, 2004). However, readers might like to reflect that if words really were a sequence of easily identifiable units of sound, many skills would be much easier to learn than they are. Not least, literacy acquisition would be far easier if it really did involve merely learning to label pre-existing sounds with letters. Similarly, if speech really were a sequence of individual sounds, like printed letters, computer speech recognition would be much easier than it is. What has made this task much more difficult than was expected when audio technology first became available is precisely the fact that words are not in fact made up from a small set of discrete phonemes.

The "literacy bias" (the erroneous belief that speech is made up of phonemes in the way a printed text is made up of letters) is mainly harmless in everyday life, and certainly does not

detract from the skills of speaking, listening, reading and writing. However, it creates a kind of "false consciousness" about speech and writing which can cause problems. As is often the case, the worst part of the problem comes not when people do not know how speech and writing work, but when they do not know that they do not know, and therefore accept their erroneous belief without seeking expert advice. Unfortunately this is a common occurrence in relation to language.

Before leaving this section it will be useful to consider another kind of "literacy bias", stemming from the extreme importance to the functioning of modern democracies of the ability to make and refer to official written records. There is a very strong sense of status granted to an agreement or decision that has been "set down in black and white" as opposed to a "merely verbal" one, considered to be, as Sam Goldwyn famously quipped, "not worth the paper it is written on". This status can easily be transferred inappropriately to any written text, to the extent that, once something has been written down, it can be credited with far more objectivity than it might deserve.

Speech perception

[100.360] We have observed that the common view that speech is a sequence of discrete phonemes, recognised one-by-one and put together into words, is a misconception of the literacy bias (even reading itself does not operate in this way). The question then arises: how does speech perception work?

The analogy was made above that speech perception is like projecting constellations onto the night sky. Actually, it is far more complex. A better analogy, presented by eminent phonetician, Peter Ladefoged, is that speech perception is a process of "hunting for words" in the continuous stream of sound. Once some "islands of meaning" have been identified, giving the general gist of the utterance, the speech perception mechanism hypothesises words between the islands to construct plausible sentences. This process is well known in experience, as shown by jokes like the one about the general's instructions "send reinforcements, we're going to advance" being passed on as "send three and fourpence, we're going to a dance".

Performing this construction process at the heart of speech perception necessarily involves combining information from the sound itself (bottom up information) with knowledge from the mind of the perceiver (top-down information). All speech perception requires both types of information, with a trading relationship between them: the better the bottom up information, the less top-down information is needed; conversely, if the bottom up information is obscured, top-down information is relied upon more. Interestingly, and again in line with other forms of perception, people generally disregard their own "top-down" contribution to speech perception, attributing their perception wholly to the stimulus.

One final, crucial, point of background needs to be given. The role of top-down information (the knowledge, beliefs and assumptions of the perceiver) means that incorrect top-down information, or mistaken knowledge, beliefs and assumptions can very easily mislead the perceptual process. Many entertaining examples are provided by mondegreens (mis-heard lyrics of songs, such as "Gladly the cross-eyed bear" being heard for "Gladly Thy cross I'd bear"). Importantly, this can happen with no diminution of confidence. Thus many people continue to believe mistaken perceptions like this for years.

Some examples of similar types of misperception have become famous in the forensic linguistics literature, for example:

> In one case […] an indistinct word in a clandestine recording of a man later accused of manufacturing the designer drug Ecstasy, was mis-heard by a police transcriber as "hallucinogenic":
>
> > but if it's as you say its hallucinogenic, it's in the Sigma catalogue.

whereas what he actually said was

but if it's as you say its German, it's in a Sigma catalogue.

In another case, a murder suspect with a strong West Indian accent was transcribed as saying in a police interview that he "got on a train" and then "shot a man to kill"; in fact what he said was the innocuous and contextually much more plausible "showed a man ticket".

Coulthard, *An Introduction to Forensic Linguistics* p 145

Indeed it is a trivial matter to experimentally manipulate a hearer's perception of speech by manipulating only their beliefs about what they are hearing (top-down information), leaving the acoustic speech wave (bottom-up information) completely intact. This can be done either at the macro level (for example a famous experiment from 1958 changed the perception of the same bottom-up information from "I tell you that our team will win the cup next year" to "I tell you that I am more hungry than you are", simply by changing a contextualising keyword from "sport" to "food"), or at the micro level (for example, one can easily change perception of a sound sequence from "pin" to "bin" by adding or removing an initial "s"). Interested readers with access to a simple computer sound editor are encouraged to try this for themselves.

Most importantly, again, this kind of manipulation typically results in no diminution in the hearer's confidence in the accuracy of their perception.

Transcription

[100.380] Now at last we can pull together all this information about the underpinning skills of speech perception and writing, and return to consider the topic of main interest, transcription.

Definition of "transcript"

[100.400] Transcription is a kind of writing, but a highly unusual kind. Most writing (as in a book or a letter) is focused on conveying meaning. It uses a signifier, but the attention of both writer and reader is on the signified. In transcribing from an audio recording, writing is not being used directly for communication of a signified. Rather, the attention is on the signifier, and the intention is to represent the signifier as accurately as possible.

This is a rather unnatural task. Indeed, only since the invention of audio recording has it been possible to do it according to current understanding and expectations. So although the word "transcribe" has been in English for many centuries, its meaning has changed radically in recent times. Originally, it meant to copy out a duplicate of an existing text. Later, it meant to take notes or minutes while a meeting was in progress and write them up or "transcribe" them at a later time. With the invention of shorthand, such notes could be taken rapidly enough for a "verbatim" transcript – traditionally the basis of court and parliamentary records.

Nowadays, the term "transcription" most often refers to the laborious process of writing down word for word exactly what someone said, based on repeated close listening to an audio recording, as discussed in the scenario above. (As an aside, it is interesting to note that the word "record" has also changed its meaning dramatically with the invention of this new technology. Originally it meant to "learn by heart" in order to be able to recite from memory, then, by extension, to take notes or a shorthand "record" as an aid to the memory, prior to full transcription.)

A transcript can be defined, then, as a written version of a spoken text – clearly distinct from a text originally created via the medium of writing. Creating and reading a transcript are very

different activities from creating or reading a written text. The differences arise precisely because the focus of a transcript is on the signifier (the sound or written marks conveying the meaning) to a much greater extent than in the case of written text, which focuses more on the signified, or meaning itself.

Levels of transcription

[100.420] Within this broad definition, it is possible to distinguish many different types of transcript. The most common is an orthographic transcript, ie one using ordinary spelling, which, because of its easy readability, is useful for many contexts involving a general readership, such as transcribing radio programs, court hearings, parliamentary proceedings, etc. However, the level of detail provided by an orthographic transcript is very low.

Many kinds of specialist transcription have been developed to record more detailed information about the signifier, especially by the various branches of linguistics. For example, a transcript intended for use in the subdiscipline of discourse analysis (the scientific study of conversation) uses augmented orthography to provide details such as duration of pauses, different kinds of "fillers" (umms and ahhs), overlapping speech, intrusion of laughter, coughs or other sounds, tone of voice, etc.

Phonemic transcription represents phonemes (individual speech sounds) systematically, avoiding the irregularities of spelling, such as use of "-ough" to represent different sounds in "through, bough, though, trough". It is often assumed that phonemic transcription, being systematic, would be much easier than spelling, but this is not at all the case. Beyond the level of simple individual words, it involves many complex judgments, and indeed there is not one system of phonemic transcription, but many. These are used in subdisciplines such as lexicography (the subdiscipline investigating dictionary making), or sociolinguistics (the subdiscipline investigating the differential use of language by different social groups) to represent pronunciation of words in different regional or social dialects of a particular language, such as the pronunciation of "tomato" in British or American English.

Phonetic transcription is significantly different from phonemic transcription, a fact obscured by common misuse of the word "phonetic" for phonemic or even modified orthographic representation. A true phonetic transcription, as used in phonetics (the subdiscipline investigating the sound of speech) gives a much more detailed representation of pronunciation, using language-independent symbols of the International Phonetic Alphabet (IPA). It is important to note, however, that even the most detailed IPA transcription leaves out a great deal of significant information. Specialised systems have had to be developed by branches of phonetics which focus specifically on rhythm, intonation and voice quality.

The point is: any transcript is necessarily an abstraction. Speech contains a limitless quantity of information of widely differing kinds. It is impossible to represent all its many aspects in one transcript, just as it is impossible to represent every detail of a room in a painting. Any one transcript, therefore, no matter how detailed and specialised, provides a limited view of speech, appropriate to the goal and context for which it is produced.

This is shown by the fact one can never reproduce speech in all its detail just by reading a transcript. One can only reproduce (and even then only to the level of one's skill) the degree of detail the transcriber abstracted from the speech and represented in the transcript. To understand the full detail of the speech, it is necessary to listen to the recording from which the transcript was made (recognising that this recording itself is an abstraction from the total situation in which the speech originally took place).

Informal recognition of this important point came at the end of the scenario, where it was seen that X and Y could come to different interpretations depending on whether they listened to the recording or read the transcript. Again, we reach the conclusion, a transcript is not equivalent to the speech it represents.

Relationship among levels of transcription

[100.440] In discussing various levels of transcription, it is important to recognise that all transcripts begin from speech perception. As we have been at pains to point out, speech perception requires not just bottom-up information from the speech signal, but top-down information from the perceiver. The next step in transcribing any recording more than a few words long, is creation of an orthographic transcript which is then reworked via repeated listening and analysis to extract the level of detail required for the particular type of transcript being created.

Since both speech perception and orthography crucially and unavoidably involve a top-down component, and any transcript crucially and unavoidably depends upon speech perception and orthography, every transcript must involve top-down knowledge from the transcriber. This means no transcript can be wholly objective.

It is of course possible to create a purely "bottom-up" analysis of speech, focusing purely on its acoustic properties, with no top-down component. This is done, for example, in various analyses performed by audio engineers and telephony experts. However, such a record is not a transcript – precisely because it takes no account of the role of the sound as a signifier. It is part of the essential nature of speech that it is not just sound, but sound that represents linguistic meaning, a signifier representing a signified. For this reason it is essential to take account of the top-down component of speech perception when analysing speech. Serious problems can be created when audio engineers with insufficient background in linguistics and phonetics are asked to analyse linguistic, as opposed to purely acoustic, aspects of speech.

It may be worth addressing here a view sometimes expressed: that greater objectivity could be achieved by use of computer transcription. It is true that speech-to-text technology has now gained extremely impressive levels of accuracy. It is essential to recognise, however, that this has been achieved not through purely bottom-up analysis, but through skillful encoding of human-like top-down knowledge into a form usable by computer processing. The result is ingenious – but not neutral or objective.

The key to achieving maximum objectivity in transcription is not to ignore or avoid the top-down component, but to recognise and control it. This point will be taken up in more detail in the next section.

[The next text page is 100-4051]

IMPLICATIONS

[100.500] The background provided in the previous section has a number of implications for the use of transcripts in the legal system.

A transcript is not an "aide memoire"

[100.520] A common response when issues about transcription are raised with legal professionals is the assertion that transcripts as such have minimal importance in legal cases, since the evidence is not the transcript but the recording itself, with the transcript acting merely as an "aide memoire".

> The tape recording (which is defined in s 281(4) *Criminal Procedure Act*) is an exhibit when it is tendered in court. The transcript of the videotape is called the *aide memoire*. It assists the court but is not the actual exhibit.
>
> (*Practitioners Guide to Criminal Law* p 172)

Based on the discussion to this point, it is clear that a transcript from an audio recording is both more and less than an aide memoire. An aide memoire is a set of notes about a meeting, event or conversation, taken as an aid to the memory of someone who has personally experienced it, as a participant or an observer – or even via a recording (as in the case where Y in the scenario above took notes from the recording of the conversation between S1 and S2, prior to attending the meeting with X). Such an aide memoire is an interpretation, focusing on the meaning of the events for the person taking the notes, in relation to the context in which they are experienced, and the purpose of remembering them. There is no necessity for such notes to accurately represent the exact words that were used in the event. Indeed, for anything above the shortest stretches of speech it is impossible for them to do so.

In forensic contexts, however, transcripts of recordings used in evidence are typically (and rightly, as will be argued shortly) produced by someone with little or no knowledge of the original event, meeting or conversation. They are then read and interpreted by people who did not experience the events, but are trying to reconstruct them as accurately as possible. Such people, by definition, cannot be remembering the events, so the transcript cannot be an aid to their memory of the events.

A transcript can appear to be a kind of aide memoire if, having listened to the recording and confirmed the transcript is a reliable record, the user then refers to the transcript as an aid to their memory of having listened to the recording. However, such use makes assumptions about the relationship between the transcript and the recording that are problematic even in cases with an agreed reliable transcript of a clear recording, In other cases, the fact that the "memory" embodied the transcript is not the user's own but that of the transcriber, can cause serious problems.

Indeed, it can happen that even seeing the transcript in advance of hearing the recording can bring into operation inaccurate top-down information, creating a kind of "false memory" that significantly biases or "contaminates" perception of the speech in the recording.

For all these reasons, it is crucial to distinguish clearly between a genuine aide memoire and a third party transcript, even an apparently "accurate" one. Calling a third party transcript an aide-memoire blurs this crucial distinction.

The accuracy of a transcript is not an objective measure

[100.540] It goes without saying that an accurate transcript is to be preferred over an inaccurate one – but what exactly does it mean to say a transcript is "accurate" or "inaccurate"? As often happens, a meaning that seems obvious in context becomes problematic when we attempt to pin down a general definition.

To say a transcript is "accurate" cannot mean that the transcript represents all the information available in the speech being transcribed. Paragraph 100.420 above argued forcefully that this is impossible. Nor can it mean that it represents the information in an entirely "objective" manner, without interpretation. Again it has been argued above (100.440) that the only way to avoid interpretation is to disregard the very aspects of speech that make it speech, as opposed to mere sound.

Similarly, to say a transcript is "inaccurate" cannot mean that it omits or alters bottom-up information, since all transcripts must do this to some extent. As discussed in more detail below, even courtroom transcripts undoubtedly do omit and alter information. Clearly we don't want a definition of "accuracy" that allows blatantly incorrect transcripts, but equally we don't want a definition so strict that even courtroom transcripts aren't included.

As observed above (100.440), transcripts require continual reworking to give the level of detail required in the context. Even considering only orthographic transcription, this level varies considerably. Bringing in the possibility of more technical transcripts adds yet more variability. The point at which a transcript is considered "complete" or "accurate enough" depends very much on its context and purpose, with a relatively arbitrary cut-off point based on diminishing returns from further reworking in relation to the purpose.

Consider the radio transcript above. Correcting the words "loci" and "inductive" would surely make it more accurate – but would it be fully accurate? It is certain that careful repeated listening to the original recording would bring to light many other cases of omission or editing. Resolving all of these would not only take a great deal more time. Importantly it would also make the transcript much harder to read. Most importantly of all, even if it were made maximally detailed in one context, it might well still fail to represent some specific aspect or detail which turns out to be crucial to another context.

This is not to endorse a generally "relativistic" view of truth. It is to emphasise that the choice of the appropriate style of transcript, as well as the many detailed judgments about how aspects of the speech should be represented in the transcript, and the decision as to when it is complete to an appropriate level of accuracy, are all judgments made by an individual in a context. It is clear, then, that judgments of accuracy or inaccuracy of a transcript must be made with recognition and respect for the context in and purpose for which the transcript is being used.

A transcript is an opinion

[100.560] The inescapable conclusion of the foregoing discussion is that a transcript is an opinion. Of course this does not render the use of transcription invalid, or make it impossible to distinguish a good transcript from a poor one: a good transcript is a valid opinion based on, and supported by, appropriate evidence. Indeed, if produced by an expert, it is an expert opinion.

It does indicate that, in judging the quality of a transcript, we are dealing not so much with a mechanistic concept of accuracy as with a concept of reliability: the extent to which it provides information relevant to its context, in a manner accurate and complete enough for its purpose.

Judging the reliability of evidence in relation to the overall context of the case is of course a matter for the court. However, it is appropriate for an expert providing advice and information about evidence to include evaluation of the reliability of the evidence and the level of certainty that should be attached to it. Thus the code of conduct of the International Association for Forensic Phonetics and Acoustics (IAFPA) includes the injunctions:

> 4. Members should make clear, both in their reports and in giving evidence in court, the limitations of forensic phonetic and acoustic analysis.
>
> 5. In reporting on cases where an opinion or conclusion is required, Members should make clear their level of certainty and give an indication of where their conclusion lies in relation to the range of judgments they are prepared to give.
>
> (IAFPA Code of Practice http://www.iafpa.net/code.htm).

What factors need to be taken into account in evaluating whether a transcript constitutes reliable evidence in a particular context?

[The next text page is 100-5051]

FACTORS AFFECTING THE RELIABILITY OF A TRANSCRIPT

[100.600] The reliability of a transcript is affected by a number of factors. Some of these have already been discussed in giving background so can be rehearsed briefly here, but it is worth setting them out as a systematic framework within which the reliability of individual transcripts in specific contexts can be assessed.

Recording factors

[100.620] The quality of the recording that provides the bottom-up information on which the transcription will be based is obviously a crucial factor. A good quality recording for purposes of transcription is not necessarily one of high fidelity, as required for music appreciation, precisely because with speech, the focus of attention is not primarily on the sound (signifier) for its own sake, but on the sound as a conveyer of meaning (signified). However two characteristics of a recording are essential: its signal to noise ratio, and its duration.

The signal to noise ratio refers to how loud the speech is in relation to any other sound that reduces its audibility, and can be affected not only by background noise, but by a buzz in the wiring, holding the microphone too far from the speaker's mouth, or using too low a "gain".

It is possible to tolerate a low signal to noise ratio as long as the noise is relatively stable. In that case, with deliberate concentration, the ear can perform a basic version of the cocktail party effect and follow the speech quite well. However, this ability is diminished if the background noise is unpredictably variable, particularly if it includes speech, from television or radio or from other speakers, at a similar level to that of the conversation being transcribed.

"Enhancement" of a recording can be valuable in some cases. However, like "transcript", the term "enhancement" covers a wide category of different techniques, few of which produce the magical effects portrayed in television cop shows. Different types of enhancement can affect perception of the speech in a wide range of ways, not all of them desirable, so it is essential to know exactly what kind of "enhancement" has been performed. The easiest type (removal of very high or very low frequency noise of stable character, and boosting of frequencies in the speech range) usually causes least interference to perception. More technically sophisticated methods of removing strongly interfering noise (such as radio or television speech, or another conversation) can detract from as much as enhance the recording if the frequencies being removed are within the same range as the speech frequencies.

The other key consideration is that the recording is of appropriate duration to enable use of top-down knowledge. This refers not just to the duration of the recording as a whole, but to the duration of each continuously audible stretch of speech or conversation, which can be affected by background or recording noise intruding at irregular intervals. The exact duration required depends on other factors, notably the various types of context discussed below.

Speech factors

[100.640] Speech varies greatly in its clarity. Some speakers have naturally "clearer" voices than others – not in terms of their accent, but in terms of the mechanics of their speech

production system, including larynx, tongue, soft palate, etc. However, all speakers range along a scale of clarity depending on various circumstances. At one end of the scale is casual conversation in a shared environment which, as discussed above, though perfectly clear to participants at the time, can be hard to decipher when heard without its original context. At the other end is prepared speech expected to be replayed out of context, best of all if speakers know they are being recorded and co-operate by speaking clearly.

Two issues are involved here. First, a casual conversation is more likely than a formal one to make elliptical reference to aspects of the physical environment, or to a context of understanding shared by participants, but not available to someone listening to the decontextualised recording. Second, since speakers in informal conversations know (subconsciously) they can rely on context to help clarify their meaning, and use a far less clear form of speech (sometimes called "hypo-articulated", as opposed to the "hyper-articulation" of exaggeratedly clear speech).

Contextual factors

[100.660] The role of context has been discussed above. Here it is worth distinguishing explicitly among three types of context: context of situation, external context, and internal (or linguistic) context. The context of situation is the room or other environment in which the conversation takes place, including all the features observable by the speakers. As just mentioned, these may be referred to implicitly rather than explicitly, causing difficulty for a transcriber. For example, if someone holds up a kettle, the words "I'll have a Milo" may seem perfectly clear to someone who sees the kettle gesture, and knows that Milo is a chocolate flavoured drink, but obscure to someone who doesn't.

The external context is the historical and social context of the conversation. Aspects of external context referred to by the speakers can be extremely important in evaluating a transcript, since there is often a verifiable trail of evidence regarding the external context.

The internal or linguistic context is the language that occurs around the particular section being transcribed. This can be extremely useful in providing knowledge relevant to the interpretation of the conversation even to those who have no access to the context of situation or external context – a fact exploited to good effect by radio plays, in which the language is carefully crafted to enable the whole conversation to be understood with no visual dimension. Importantly, linguistic context is only usable if the speech being transcribed is of sufficient duration and has sufficient "continuity", or internal cross-references, and the listener has sufficient competence both in the language and in linguistics to make proper use of it.

Listener factors

[100.680] It has been emphasised above that transcription depends upon the prior process of speech perception. This means that the transcriber is first a listener to the speech. We have seen the crucial role played by listener knowledge, beliefs and assumptions about the meaning of the speech (top-down information), especially in situations where the bottom-up information is degraded in any of the many possible ways this can happen. In general terms, someone with greater knowledge of the context (in all three senses discussed in the previous section) is better placed to interpret the speech accurately than someone who lacks that knowledge (or top-down information).

However, we have also emphasised that this advantage is a two-edged sword, since there is always the possibility that the knowledge, beliefs or assumptions about the speech may be incorrect in some way. In this case, there is a strong risk of incorrect perception, and little likelihood that this risk will be reflected in reduced confidence on the part of the listener (see para 100.360 for fuller discussion). Thus, to extend the example in the previous section,

someone who does not know that Milo is a chocolate drink but does know that Marmite is a savoury sandwich spread, might assume a slightly different context of situation, and hear the recording quite confidently as "I'll have Marmite please". Similar cases have been seen in the radio transcript and forensic linguistics examples above.

Another characteristic of the listener mentioned briefly above is competence in the language. It is important here not to let use of the term "competence" suggest any privilege for a standard or "educated" variety of the language. The transcriber may be a highly competent speaker of the standard variety, but if the speakers use a non-standard variety, it is competence in that non-standard variety which is relevant to the quality of the transcript. Indeed, for conversations involving multilingual participants, a transcriber will require competence not just in those language varieties, but in the style of "code-switching" relevant to the conversation.

Transcriber factors

[100.700] It has been emphasised above that even orthographic transcription (using ordinary spelling) is a much more difficult task than is often realised by those who have not attempted it. Untrained people are unlikely to be able to produce a reliable transcript (though they may well be able to provide information useful in the production of a reliable transcript, as in the philosophy example above). Worse, untrained transcribers are unlikely to be aware of the many issues surrounding reliable transcription, and may therefore place undue confidence in their own work.

It is interesting to note that amateur transcribers often give themselves away with transcripts that are too perfect. They may use grammatical sentences appropriate for written text but very different from the structure of real spoken discourse. Or they may show knowledge they could not have obtained from the recording. As an example from my own experience, I was asked to match the voice in an incriminating recording with a recording of a suspect reading the same words from a transcript. I was reluctant to do this, since read speech has so many characteristics different from spontaneous speech that comparison is often inconclusive, but on insistence I agreed. On hearing the recording of the read version I was struck by the fact the suspect appeared to have been able to hear accurately words in the original recording that I myself had been unable to transcribe. I was able to suggest that he possibly had some knowledge of the context in which the recording had been made.

The professional transcriber is greatly aided not just by having appropriate equipment (for example, good quality headphones, foot-operated playback, specialised software), but also by having relevant training and experience in the specialised form of listening to the signifier required for transcription, so different from the normal "listening for meaning" familiar in everyday situations. Interestingly, this specialised form of listening is sometimes known among professional transcribers as "ear squinting", by analogy with the "squinting" performed by artists to enable them to concentrate on the visual features of a scene without being distracted by top-down information.

As well, a professional transcriber knows and consistently uses editing and layout conventions to produce a transcript of appropriate readability for its context. This skill, however, does have a downside if carried inappropriately into situations where detail of the exact words spoken is more important than readability.

Even more important, however, is the transcriber's independence and good faith in rendering an unbiased account of the material. It is essential here to point out again that "bias" can be entirely unintentional, coming only from subconscious beliefs or assumptions about the nature of the material. "Good faith" in this context, therefore, must include not just absence of intention to deceive, but recognition of one's own limitations, and active intention to bring to light any potentially biasing beliefs. The role of the transcriber and the expertise required for different types of transcript is discussed further below.

Situational factors

[100.720] The final factor to be considered here is the situation in which the transcript is being prepared: the purpose of the transcript, who it is to be used by, the context of its creation and use. Several aspects have been discussed at various places above, and the point made that the transcript should include all information relevant to the purpose of the transcript, while excluding irrelevant information, which may have the effect of obscuring the important points.

Perhaps the most important aspect of the situation for ensuring reliability, however, is the degree to which the transcript can be checked, either by someone who witnessed or participated in the conversation, or against some form of external evidence.

[The next text page is 100-6051]

ASSESSING THE RELIABILITY OF TRANSCRIPTS USED IN COURT

[100.800] Considering these factors allows a clear distinction to be drawn between the various artifacts called "transcripts" described in the introduction (courtroom transcripts, ERISP transcripts and transcripts of evidence in surreptitious recordings), and gives a clear rationale for treating them differently. Consider first the production of courtroom transcripts. The various factors just discussed can be used as a framework for assessing their reliability.

Courtroom transcripts

[100.820] In principle, audio recordings are not even necessary for creating courtroom transcripts, which have been in use for longer than audio recording technology has been available. However, as was made clear by the scenario above, some kind of simultaneous record is essential, as human memory is not capable of reproducing speech verbatim after even a short interval. Traditionally, such a record has been created by a skilled stenographer who takes shorthand notes to write up (or "transcribe", in an earlier meaning of the word) while memory is fresh. These days, however, proceedings are often audio recorded instead.

Recording factors

The traditional shorthand record is created by a highly skilled professional, present in the courtroom, whose accuracy has been guaranteed through extensive training and testing. The audio recording, while perhaps not of utmost fidelity, is generally adequate, with the system having been set up specifically to allow a transcribable audio recording.

Speech factors

The participants are all enjoined to speak clearly, one at a time, in a formal style, specifically so that every word can be heard by everyone in the room, including the transcriber, as well as reproduced in the audio recording. Anyone mumbling, using gestures, or otherwise not clearly audible, is asked to repeat what they said.

Contextual factors

There is considerable continuity, creating reliable internal or linguistic context, and numerous official witnesses to every word spoken, creating verifiable external context. The context of situation is well known to all participants, and in any case the speech is of a highly decontextualised kind (for example, if an object is referred to, it is explicitly named and objectively referenced).

Listener factors

Traditionally the transcriber is present during the entire event, so able to bring considerable top-down knowledge to the transcription, but independent and disinterested, so minimising the likelihood of bias in the perception of the speech. These listener advantages are greatly reduced in the increasingly common case where transcribers are not present, as discussed in the next section.

Transcriber factors

Traditional court reporters are trained to a high level of skill not just in creating a simultaneous shorthand record but in transcribing it into typescript – part of the qualification for the job involves demonstrated 98% accuracy on a series of prepared tests.

The facts about speech perception, writing and transcription outlined above offer strong support for court and parliamentary reporters' argument that moves to cut costs by audio-recording proceedings for transcription by teams of typists working at locations remote from the court will result in a poorer quality public record. Putting aside questions of skill and professionalism, such typists, though they may benefit from being able to replay the audio recording, suffer not just from not having been present during the proceedings, but from being able to listen to the audio only in decontextualised sections.

Situational factors

The completed transcript is able to be checked by key participants in the proceedings, and for important cases such checking is required to be done within a short time of the events taking place, so that any disputes or disagreements can be resolved and corrected while memory remains fresh.

[100.840] All these factors work together to make such courtroom transcripts highly reliable – as is appropriate given their vital importance to the functioning of the legal system in our society. Indeed the procedures have been deliberately evolved to ensure maximal reliability for the purposes for which they are required. It is thus reasonable to use the transcripts with confidence, and only in rare instances is it necessary to refer back to the original recordings for confirmation or an alternative interpretation.

Nevertheless, it is important to recognise that such transcripts, reliable as they are, are neither objectively complete, nor infallible. First, a courtroom transcript, while highly useful for its purpose – that of providing an easily legible, orthographic transcript of the proceedings at a sentence level of accuracy – is not nearly detailed enough for, say, a discourse analysis of courtroom talk, or a phonetic analysis of the pronunciation of the speakers.

Secondly, reliable as such transcripts are, there are inevitably inaccuracies, whether introduced accidentally, by transcriber error, or deliberately, for the sake of rendering the speech more appropriate for a written record. This has been shown in cases where recordings of court proceedings have been subjected to various types of linguistic analysis, in which more detailed transcripts are produced and compared to the original court reporter's transcript. Again, this does not mean the original transcript was inadequate for its purpose. Far from it – overly detailed linguistic transcripts would be less useful as a general record of court proceedings. It does demonstrate that any transcript is an abstraction, involving interpretation and judgment in a context.

ERISP transcripts

[100.1000] Consider next the ERISP ("Electronic Record of Interview with a Suspected Person") – an audio or video recording of a police interview. Here too, standard procedures have been developed to ensure a reliable record of the interview is available to the court, as an agreed "official version" – avoiding the problems inherent in the former use of "verbatim" records. Again, these procedures can be assessed in terms of the framework of factors.

Recording conditions are generally adequate. Style of speech is generally formal, with the speaker on at least one side of the conversation trained and motivated to conduct an interview so as to obtain a reliable version of the "on-the-record" information being exchanged. Contextual factors are controlled, with the interviewer able to interject comments "for the

benefit of the recording". Transcription is usually done by an independent transcriber, typically with somewhat lower levels of training than for courtroom transcripts but still a dedicated professional. Recording and transcript can be checked by participants if required.

The result is a transcript perhaps somewhat less reliable than the courtroom transcript but certainly adequate to its purpose – though it is notable that legal practitioners are advised not to take this for granted:

> It is important to watch the ERISP or listen to audio tapes of records of interview. It will not only help you work out whether the transcript is accurate, but it may also indicate important aspects of the questioning and your client's manner and condition at the time of questioning which may be relevant in your case (for example, being intoxicated or not in a fit mental state).
>
> (*Practitioners Guide to Criminal Law* p 172)

Surreptitious or adventitious recordings

[100.1020] Consider now transcripts of audio evidence obtained from telephone intercepts, listening devices, answering machines, dictaphones or the like. Where can these transcripts be placed within the framework of factors? Recording conditions are frequently poor, sometimes extremely so. Style of speech is never monitored for the recording, rarely formal, and frequently very casual and/or emotional. The context of situation is not controlled, and often unknown. The linguistic context may be compromised by disjointed conversation or background noise. The details of external context may or may not be known. Most importantly, the transcript is unlikely to be checked by the speakers – or if it is their opinion may not be trusted.

Consider, as an example, the David Eastman transcript mentioned in the introduction above. The recording, made by a hidden listening device, is of very poor quality. The speech, being a whispered soliloquy, is of low volume, and produced without even a casual listener in mind. The context exists almost entirely within the mind of the speaker, and is thus inaccessible to the listener, who can at best bring assumptions rather than knowledge about its intended meaning. With such a poor recording, top-down knowledge is essential to even the most skilled of transcribers, introducing the risk of unwanted bias in perception. If that top-down knowledge includes information about the external context of the case, there is no way to check the transcript except by asking the speaker, whose opinion is necessarily unreliable in this instance.

Under conditions like these, listener/transcriber factors – the only ones that can be controlled retrospectively – become crucial. Paradoxically, however, far from the careful procedures just described for court and ERISP recordings, there are no standard procedures for ensuring reliable transcription when the recording itself is of very poor quality. In some cases, appropriate steps are taken to ensure a reliable transcript, but in others the process shows what can only be described as naivety about the nature of a transcript.

In one example, I was contacted by a defence lawyer with a request to examine a recording for evidence of tampering. This lawyer's client acknowledged the recorded conversation had taken place, and admitted being the speaker, but denied having spoken the incriminating words attributed to him in the transcript. Before recommending an audio engineer to examine the recording for tampering, I listened to the recording. Since it was of extremely poor quality, I found the relevant words to be of very low intelligibility. Poor transcription appeared to be a much more likely explanation for the disputed utterance than tampering. Upon querying this, I was alarmed to discover the transcript had been prepared by a prosecution witness and accepted as evidence in the court with no evaluation process.

While this is perhaps an extreme example, it is one of many, even in my own limited experience. Introduction of a simple but standard process for assessing the reliability of

transcripts against the factors set out in this section before they are presented in court would avoid the waste of time, and worse, that can follow from this type of occurrence.

[The next text page is 100-7051]

OBTAINING A RELIABLE TRANSCRIPT OF A POOR QUALITY RECORDING

[100.1100] We have emphasised the importance of using a professional rather than amateur transcriber. An experienced professional transcriber, given appropriate instructions and sufficient time, can often make a good transcript even of a poor quality recording. However, with such restricted bottom-up information, the transcriber's lack of top-down knowledge can be a severely limiting factor, to the extent that those involved in the case may genuinely be able to hear material that is not intelligible to the transcriber.

What is the appropriate procedure in this circumstance? It is tempting to have the recording transcribed by someone who does have knowledge of the external context. This can be a valuable part of the process. However, on its own it is a highly problematic course of action.

Dangers of using an "ad hoc expert"

[100.1120] It is common for transcripts of poor quality recordings prepared by police directly involved in a case to be admitted as expert evidence on the grounds that the transcriber's detailed study of the recordings gives him or her "ad hoc expertise" in relation to its contents.

> [2.5.270] Ad Hoc Expertise
>
> In *R v Menzies* [1982] 1 NZLR 40 at 49 Cooke, McMullin and Somers JJ and Sir Clifford Richardson concluded that a person could become a "temporary expert in the sense that by repeated listening to the tapes he had qualified himself ad hoc".
>
> In *Butera v Director of Public Prosecutions (Vic)* (1987) 164 CLR 180 [PDF]; 30 A Crim R 417 [PDF] at 195 (CLR), 427 (A Crim R) Dawson J adopted the concept of "ad hoc expertise". He held that where words in a tape recording are inaudible or unintelligible "expert evidence of its contents may be required and it has been held that an ad hoc expertise may be acquired by a witness by playing and replaying a tape so as to become more familiar with its contents than could be done by playing it only once or twice". (See too *Hopes and Lavery v HM Advocate* [1960] Crim L R 560.)
>
> In *R v Leung* (1999) 47 NSWLR 405 [PDF] at 412-413 Simpson J (with whom Spigelman CJ and Sperling J agreed) further endorsed the concept and Sperling J expressed the view that s 79 of the *Evidence Act 1995* (NSW) accommodated the notion of an "ad hoc expert". In *Li v The Queen* (2003) 139 A Crim R 281 [PDF] at 287-288 Ipp JA endorsed the approach and further applied the notion of "ad hoc expertise" in the context of an interpreter and translator who had spent many hours listening to an individual's voice.
>
> (Freckelton and Selby 2.5 *The Expertise Rule*)

This view is understandable in view of the fact that top-down knowledge is particularly necessary when bottom-up information is degraded. However several cautions are in order regarding the use of such "ad hoc expertise" – in addition to the obvious fact that police generally do not have the skills of a professional transcriber.

It is important to distinguish the superior understanding of a transcriber who has listened many times to a recording, from superior understanding conferred by knowledge of the external circumstances of the case. If it is the former that is required, a professional transcriber, independent and disinterested with regard to the content of the recording, if given sufficient time, might appropriately be termed an "ad hoc expert" in regard to the speech on that particular recording (as opposed to a general expert in linguistics and phonetics, to be discussed below).

Giving such a role to police involved in the case makes it impossible to separate top-down understanding arising from repeated listening to the material (internal or linguistic context) from top-down understanding arising from inside knowledge or prior expectations about the likely or possible content of the recording (external context).

This creates several problems. First, it removes one of the most important situational factors – the ability to check the transcript against known external evidence – often the only check available. Second, it introduces considerable risk of bias. Such bias, as remarked several times already, need not arise from any malicious intent; it is simply an inevitable part of speech perception. On the other hand, it would be naive not to recognise the possibility that an apparently superior understanding might mask, if not a deliberate intent to deceive, at least lack of intent to focus on the accuracy of the transcript more than on the need for evidence that suits the police case.

Clearly there is a role for police or other "insiders" with knowledge essential to the interpretation of hard-to-hear material. However this knowledge must be interpolated, with caution, through a genuine and independent expert in relevant branches of linguistics and phonetics, as will be discussed shortly. Allowing transcripts produced by police directly involved in a case to be defended by the police transcribers themselves in the role of "ad hoc expert" is a dangerous practice.

Value of appropriate expertise

[100.1140] Another temptation when professional transcribers have trouble with poor quality recordings is to send the recording for acoustic analysis – to find out "what the sounds really say". Acoustic analysis can indeed be valuable, particularly in relation to short "disputed utterances", such as those mentioned in para 100.360, but again there are dangers to avoid. First, it is essential to obtain the acoustic analysis from an expert qualified in linguistic phonetics, rather than audio engineering or similar subjects. Audio engineers can of course produce a scientific analysis of the sound of speech, but often have insufficient knowledge about language to interpret the results properly (100.440).

Second, it is important to recognise that even a phonetically valid acoustic analysis does not in itself determine the correct transcription. It can provide improved bottom-up information but this must be combined with top-down information from the linguistic context to arrive at a transcript of the words that were spoken. Ensuring the reliability of the top-down information requires expertise from other branches of linguistics, sometimes but not always embodied in the same person as the phonetics expertise.

Linguistics is a broad discipline with widely divergent branches. Some of these have already been mentioned – for example, phonetics, sociolinguistics, discourse analysis – but there are many others as well, notably psycholinguistics, dialectology, and literacy theory. Most linguists specialise in more than one of these but few in all of them. One subdiscipline worth special mention is forensic linguistics – the application of linguistic expertise in forensic contexts. Since forensic applications often raise significantly different issues to those encountered in the course of general research, this is increasingly developing as a subdiscipline in its own right, and it is desirable that experts consulted in relation to forensic cases have knowledge of this relatively new area.

Unfortunately, the discipline of linguistics is little known outside its own boundaries, and is often confused with general "knowledge about language". Thus it not infrequently happens that people, who may have genuine expertise in another field, claim unwarranted expertise in linguistics or phonetics. It is therefore important to choose an expert with qualifications and experience specifically in relevant subdiscipline(s) of linguistics. Any genuine linguistics expert should be able to advise on the relevant distinctions.

Appropriate experts are likely to be identifiable not just by their qualifications but also by their caution in relation to the work they undertake. For example, they are likely to show awareness of the need to avoid unintentional bias, and ensure they are not given too much information about the external context of the case in advance of their transcription. On the other hand, they are likely to engage in general discussion about the type of material and the requirements of the case so as to judge the required type of transcription and level of detail. When they produce the transcript, it is likely to come with an indication of their level of certainty (as required by the IAFPA Code of Conduct, discussed in para 100.560), and this level is unlikely to be 100 per cent. They are also likely to used conventions to differentiate levels of certainty for various parts of the transcript, and to prefer technical descriptions to overuse of terms like "unintelligible".

Expert evaluation of transcripts

[100.1160] Any transcript is an opinion, and even experts can disagree. When this happens, or when it is necessary to consider transcripts by non-experts who have relevant top-down knowledge of the external context, it is necessary to evaluate the reliability of competing transcripts in the broader context of the case as a whole.

Such evaluation involves more than simply preparing another competing transcript. It constitutes a different phase of analysis and requires expertise in a broader range of linguistic subdisciplines, including in-depth understanding of the facts about speech and writing and the framework of factors outlined above, and their relevance to the forensic context. This evaluation phase need not be overly time-consuming, and can often result in simple confirmation of the reliability of one or more parts of one or more transcripts. If this is not possible, the dispute may become a matter for the court, though, as discussed in the next section, expert advice continues to be useful in ensuring material is presented in the best possible way.

Perhaps the most important role for an evaluator, however, is where, far from confirming a transcript as correct, it is necessary to insist that a passage for which a transcript has been provided should in fact have been labeled "untranscribable".

The case of David Bain, in New Zealand, provides a useful example to illustrate this. An important piece of evidence in Bain's trial for the murder of his family was a recording of an emergency telephone call. Most of the one-minute phone call was easily comprehensible. However, one barely audible section, isolated from the rest of the speech and little more than a second in duration, was alleged by a detective, backed up by an audio engineer, to contain a confession consisting of the words "I shot the prick". At various stages of a lengthy appeal process, this recording was analysed by several independent phoneticians with different areas of specialisation, each producing an opinion as to what had been said. Finally Dr Phil Rose demonstrated convincingly that, while it was not possible to say exactly what phrase Bain had uttered, it probably had not been "I shot the prick". Bain was subsequently acquitted and released after a decade in custody.

Several lessons about transcription can be learned from this case. First, a team composed of a detective and an audio engineer forms a potentially disastrous combination for analysis of speech – an observation backed up by a number of other cases. Second, it is important to distinguish the role of providing a transcript from evaluating existing transcripts.

In the Bain case, for example, running through the factors listed above would have suggested from the outset that very little weight should be given to the original detective's transcript "I shot the prick": the recording of the disputed utterance is of poor quality, and extremely short in duration; the speech is distorted by heavy breathing; the internal or linguistic context does not enable use of top-down information; there is no external context to provide a check for the transcript; the detective transcriber is neither skilled not independent, and the audio engineer is skilled in an inappropriate discipline.

All these factors would suggest it would have been appropriate for the material to be evaluated from the outset as "untranscribable". A standard procedure insisting on evaluation of transcripts according to these factors before recordings can be produced as evidence would likely have resulted, in this case as in others, in substantial savings of court time.

Presenting transcripts in court

[100.1200] Hopefully, following a process of evaluation based on the six factors outlined above will enable many transcripts to be agreed by relevant parties before being presented as evidence. Inevitably, however, even with the best processes, there will be disputes over transcripts of potentially incriminating recordings. Further, in the absence of such a process of evaluation, there will continue to be many cases where poor quality transcripts are presented in an uncontrolled way, leading to disputes over their reliability.

Such disputes clearly have to be resolved by the court. In some cases, especially if the external context has been kept from the transcriber so the information can be judged against a known external context, this may be relatively straightforward. In others, it will be useful for the court to take advice from an expert in evaluating, as opposed to merely producing, transcripts.

It can be very confusing to a jury, or anyone, to have to listen repeatedly to a poor quality recording. More importantly, as is well understood in the legal system, it can lead to "contamination" in the sense of creating expectations that reduce their objectivity and independence. This is one reason it is desirable if at all possible for any areas of disagreement about a transcript to be localised, if not resolved, before it is tendered as evidence.

It also makes it important to control carefully the manner in which the recording and the transcript are presented – to avoid prior knowledge of the transcript inadvertently introducing biasing top-down knowledge. For example, it is useful to prepare listeners beforehand with a short introduction explaining that there is a dispute over the correct transcription of the material they are about to hear, and acknowledging that it may be difficult for them to hear – avoiding of course any reference to the detailed nature of the dispute. They should then be presented first with the audio (or video if available), preferably in good listening conditions (simply playing over speakers in the open court is not ideal) and allowed to listen to it several times to form their own judgment as to the content. Only then should they be given the alternative transcripts, presented by each side in whatever way and with whatever preamble or additional evidence is deemed necessary, preferably with an expert evaluation in support.

While a process similar to this is followed in many cases involving disputed transcripts, it is certainly not universal. Again, instituting a clear standard procedure, with advice from appropriate linguistics expertise, would surely be an advantage to the legal system.

[The next text page is 100-8051]

CONCLUSION

[100.1300] This chapter has sought to emphasise the importance of transcripts in the legal process. Having a written version of material on an audio recording is convenient, allowing easy reference to what was said, when, and by whom. More importantly, it ensures a formally agreed "official version" can be made available to all parties.

It has also sought to emphasise the need for solid understanding of the general status of a transcript as an artifact, and for this understanding to be used to distinguish clearly between transcripts with different levels of reliability. It has provided a framework to enable such distinctions to be carried out in a systematic and explicit manner.

Finally, it has sought to emphasise the existence of different kinds of expertise relevant to the use of transcripts in the legal system, specifically to distinguish between expertise in transcription, expertise in police and detective work, expertise in the audio engineering, and expertise in linguistics and its various subdisciplines.

Most importantly, it has sought to call attention to the need for expertise in linguistics as distinct from general knowledge or "common sense" about speech and language. This is in light of two related problems commonly faced by experts in phonetics and other branches of linguistics.

On the one hand, it is common for linguistics expertise to be undervalued, not just in relation to transcription but more generally. Thus evidence from linguistics experts is sometimes challenged on the grounds that it may not be sufficiently distinct from "general knowledge". This is very unfortunate in view of the many erroneous beliefs about speech and language that are part of "general knowledge". On the other hand, it is equally common for linguistics experts to be credited with far greater technical abilities than are actually available to the field.

It is hoped that this chapter will join the work of other linguistics experts (see bibliography below) in disseminating a more realistic appraisal of the role of expertise in these subjects, and in particular, of the need for expert evaluation of transcripts used as evidence in court. While the examples of poor practice presented in the chapter were resolved appropriately, it is clear that in at least some of these cases it was sheer chance that they came to the awareness of a relevant expert.

To conclude as we began: it would be advantageous if the reliability of transcripts of audio recordings used in evidence were guaranteed by processes equivalent to those that have for centuries ensured the reliability of transcripts of legal (and parliamentary) proceedings themselves – or indeed equivalent to those that are routine for most other forms of scientific evidence.

[The next text page is 100-13051]

SELECTED BIBLIOGRAPHY

Please note: This is not intended as a comprehensive bibliography, but rather to provide references for the account of transcription given above, and an accessible introduction to the field for those who want to learn more about it.

General phonetics

Ladefoged Peter, *A Course in Phonetics* (5th ed, Boston: Thomson/Wadsworth, 2005)

Ladefoged Peter, *Vowels and Consonants: An introduction to the sounds of language* (2nd ed, Oxford: Blackwell, 2005)

Laver John, *Principles of Phonetics* (Cambridge: Cambridge University Press, 1994)

Perkell Joseph and Klatt Denis (eds), *Invariance and Variability in Speech Processes* (Hillsdale New Jersey: Lawrence Erlbaum Associates, 1986)

Forensic phonetics

Note: The *International Journal of Speech Language and the Law* is a useful resource beyond the articles cited here.

Coulthard R Malcolm and Johnson Alison, *An Introduction to Forensic Linguistics: Language in Evidence* (London and New York: Routledge, 2007)

Fraser Helen, "Issues in Transcription: Factors affecting the reliability of transcripts as evidence in legal cases", (2003) 10 *International Journal of Speech Language and the Law* 203-226

Nolan Francis, *The Phonetic Bases of Speaker Recognition* (Cambridge: Cambridge University Press, 1983)

Rodman Robert, "Linguistics and the law: how knowledge of, or ignorance of, elementary linguistics may affect the dispensing of justice" (2002) 9 *International Journal of Speech Language and the Law* 92-101

Rose Philip, *Forensic Speaker Identification* (London: Taylor and Francis, 2002)

Shuy Roger W, *Language Crimes: The Use and Abuse of Language Evidence in the Courtroom* (Oxford: Blackwell, 1993)

Transcription

Baron Dennis, *A Better Pencil: Readers, writers and the digital revolution* (Oxford University Press, 2009)

Branley Duncan, "Making and Managing Audio Recordings" in Seale Clive (ed) *Researching Society and Culture* (2nd ed, London: Sage, 2004), 207-228

Bucholtz Mary, "The politics of transcription" (2000) 32 *Journal of Pragmatics* 1439-1465

Dixon David, "Videotaping Police Interrogation" (2008) *University of New South Wales Faculty of Law Research Series* 28

Eades Diana, "Verbatim courtroom transcripts and discourse analysis" in Hannes Kniffke (ed) *Recent Developments in Forensic Linguistics* (Frankfurt: Peter Lang, 1996), 241-254

Green Judith, Franquiz Maria and Dixon Carol "The myth of the objective transcript: Transcribing as a situated act" (1997) 31 *TESOL Quarterly* 172-176

Haworth Kate *An analysis of police interview discourse and its roles in the judicial process* (PhD University of Nottingham, 2009)

Kerswil Paul and Wright Susan, "The validity of phonetic transcription: Limitations of a sociolinguistic research tool" (1990) 2 *Language Variation and Change* 255-275

Lapadat Judith C, "Problematizing transcription: purpose, paradigm and quality" (2000) 3 *International Journal of Social Research Methodology* 203-219

Walker Anne Graffam, "Language at work in the law" in Levi Judith and Walker Anne Graffam (eds), *Language in the Judicial Process* (London/NY: Plenum Press, 1990), 203-244

Wray Alison, Trott Kate and Bloomer Aileen, *Transcribing speech orthographically*, in Wray Alison, Trott Kate and Bloomer Aileen (eds), *Projects in Linguistics* (Arnold, 1998), 201-212

Language and linguistics

Mesthrie Rajend, Swann Joan, Deumert Ana and Leap William L, *Introducing Sociolinguistics* (Edinburgh: University of Edinburgh Press, 2009)

Taylor John R, *Cognitive Grammar* (Oxford: Oxford University Press, 2002)

Speech perception

Erard Michael, *Um … Slips, stumbles and verbal blunders, and what they mean* (New York: Random House, 2007)

Field John, *Psycholinguistics: The Key Concepts* (London: Routledge, 2004)

Fraser Helen, *Teaching Pronunciation: A guide for teachers of English as a second language* (CD-ROM, updated, Canberra: Commonwealth of Australia, Department of Education Training and Youth Affairs, 2004)

Warren Richard, *Auditory perception: a new analysis and synthesis* (2nd ed, Cambridge, UK; New York, NY, USA: Cambridge University Press, 1999)

Writing and literacy

Byrne Brian, *The Foundation of Literacy: The child's acquisition of the alphabetic principle* (Hove [England]: Psychology Press, 1998)

Gillon Gail, *Phonological Awareness: From research to practice* (New York: Guilford Press, 2007)

Olson David, *The World on Paper: The conceptual and cognitive implications of writing and reading* (Cambridge: CUP, 1994)

Port Robert, "The graphical basis of phones and phonemes" in *Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production* Murray J Munro and Ocke-Schwen Bohn (eds), (Amsterdam: John Benjamins, 2006)

Postman Neil, *Amusing ourselves to death* (New York: Penguin, 1985)

Tannen Deborah (ed), *Spoken and Written Language: Exploring Orality and Literacy* (New Jersey: Ablex, 1982)

General perception, cognition and memory

Chalmers Alan, *What is this Thing called Science?* (3rd ed, Open University Press, 1999)

Frith Chris, *Making up the Mind: How the brain creates the mental world* (Oxford: Blackwell, 2008)

MacDonald Matthew, *Your Brain: The Missing Manual* (O'Reilly, 2008)

Miller Judi B, deWinstanley Patricia D, and Carey P, "Memory for Conversation" (1996) 4 *Memory* 615-632

Research Board of the British Psychological Society, *Guidelines on Memory and the Law: Recommendations for the scientific study of human memory* (Leicester: The British Psychological Society, 2008)

Legal references

Fernandez Lester, *Practitioners Guide to Criminal Law* (3rd ed, Sydney: NSW Young Lawyers Criminal Law Committee, 2004) http://www.criminal.younglawyers.com.au

Jarrett Janet, "Murder of Assistant Commissioner Winchester: The tragic end of an era", *The Canberra Times* (1995, reprinted in Australian Federal Police Platypus Magazine Oct 1999) http://www.afp.gov.au/about/publications/platypus_magazine/platypus_magazine_previous_editions/1999/october_1999/murder.html)

Freckelton Ian and Selby Hugh (eds) *Expert Evidence* (Sydney: Thomson Reuters subscription service)

The Wikipedia article "David Bain" brings together a useful collection of information giving a detailed history of the case.

[End of Volume 5]