Chapter 8

The identification of the individual through speech

Dominic Watt

1. Introduction

Research on language and identity, including studies carried out by contributors to this book, reveals that the language choices we make are a central element of our conception of ourselves not just as members of social groups but as self-contained individuals distinct from all others. This chapter explores some of the evidence for and against this view of our own linguistic uniqueness, by looking at ways in which an individual can be identified by others through his or her speech patterns.

For reasons of space it will not be possible to give an account of other means by which linguistic identification of a person might be attempted, such as through handwriting analysis and the sophisticated stylometric techniques developed by literary scholars and forensic document analysts (see further Chaski 2005; Grant 2008). Since the aim is to identify collocations of features that are unlikely to be shared by more than a few people, the methods used to try to attribute a written text to a particular author resemble those employed in analysis of speech recordings, and in some criminal investigations both have been carried out in parallel (for example, Windsor-Lewis 1994; Ellis 1994). Our focus will henceforth be exclusively on speech, however.

We will firstly consider what has been called ‘lay’ or ‘naïve’ speaker identification: that is, impressionistic identification of individuals by listeners lacking specialised linguistic or phonetic training. It is a task we perform on a day-to-day

basis, and it seems plausible to suppose that the cognitive mechanisms that permit the recognition of known voices are unconsciously activated whenever we are exposed to a voice we have not previously heard.

A second type is the 'technical' speaker identification used for forensic applications. Here, a speech analyst is instructed by the police or a lawyer to scrutinise a voice recording using a set of formalised procedures, usually as part of an investigation of a crime in which one or more speech recordings have been adduced as evidence. Experts have a range of analytical methods at their disposal. Among these are the International Phonetic Alphabet (IPA), which through careful listening permits detailed transcription of spoken utterances, and acoustic phonetic analysis, which is made fast and efficient by dedicated and ever more powerful speech analysis software. This combined auditory/acoustic approach has proven successful in profiling speakers in investigations as yet lacking a suspect, and in cases in which comparisons are made of an incriminating speech sample with reference recordings of a known individual.

Thirdly, there now exist automatic speaker recognition (ASR) and speaker verification systems reliable enough to be used for gatekeeping applications, for example to verify the identity of callers to telephone banking operations by matching utterances to voice samples held in a database. ASR software is also used in surveillance and criminal investigations by police forces and intelligence agencies, who may wish to attempt to match an incriminating recording with a speech sample from a known individual.

These approaches to speaker identification depend upon the notion that it is legitimate to associate a particular voice with a specific person. Common sense might dismiss this as a truism but, as we shall see, we should exercise caution when making

assumptions about the uniqueness of individual voices, and the existence of unique 'idiolects'. We must also be wary of overestimating our skills at identifying voices as ones we think we have heard before. In forensic contexts an uncritical reliance on these assumptions has been, and in many jurisdictions still can be, literally a matter of life and death.

Nevertheless, we should not dismiss our capabilities as listeners too casually, as in many ways they are impressively sophisticated, whether or not the listener has had prior formal training. We turn first to look at aspects of the identification of individual speakers through their speech by 'lay' listeners.

2. Informal speaker identification

Almost everyone reading this book will have had the experience of answering the telephone and recognising the voice of the caller, perhaps after only one or two syllables. We can perform this task fairly well even in the absence of any non-linguistic information about the talker, and in spite of the degradation of the acoustic signal imposed by the limited bandwidth (c. 300 - 3,500Hz) of the telephone line. It is not difficult to elicit anecdotal examples of cases in which people have recognised voices they have not heard for substantial periods of time: the voice in question might be that of the presenter of a television programme forgotten since childhood, say, or that of a schoolmate with whom contact had been lost for decades. Even if we cannot immediately name the person we think is talking, we may feel absolutely certain that the voice is known to us.

This suggests that we store detailed information about the voices of individuals we encounter throughout our lives (Meudell et al. 1980; Hollien and Schwartz 2001), just as we store information about aspects of people's appearance,

such as details of faces, hairstyles, and clothing (for example, Mäntylä 1997; Burgess and Weaver 2003; Yarmey 2004). Not surprisingly, the amount of attention paid to the heard speech on the part of the listener appears to affect the accuracy of subsequent speaker recognition, as does whether the listener actively participated in a conversation or was a passive eavesdropper, conditions which predict better and worse performance respectively (Hammersley and Read 1985). The amount of exposure a listener has had to a voice is obviously crucial too; it constitutes, of course, much of the difference between novel and familiar voices. Not surprisingly, novel voices that are heard for longer are more reliably identified afterwards (Pollack et al. 1954). Numerous experimental studies, beginning in the 1930s with McGehee's research on the effects of delay on voice identification accuracy (McGehee 1937; Yarmey et al. 2008), also demonstrate a fairly rapid decline after initial exposure in our ability to pick a previously-heard but unfamiliar voice out of a line-up of voices selected for their similarity to the target voice.

However, the fact that we can internalise a representation of a voice such that a novel talker relatively quickly becomes familiar suggests that our capacities here are quite highly developed. There is a ring of plausibility to Hollien's hypothesis that the ability to associate voices with individual in-group members and potential rivals or enemies evolved as a survival strategy in early *Homo sapiens* (Hollien 2002: 17). Research on other species indicating that birds, cetaceans and even amphibians can identify conspecific individuals through their vocalisations (for example, Clark et al. 2006) shows that humans are not unique in this regard. The number of individual voices we may retain memory traces of is still unknown, however. It is probably unsafe to assume that because our memory capacity for faces seems to run into the thousands (Dudai 1997; Quiroga et al. 2005) our memory for voices is necessarily

equally well developed, but the latter seems likely to extend to at least three figures. Given that some estimates of the typical size of an individual's social network fall in the 100-300 range (for example, McCarty et al. 2001; Hill and Dunbar 2003), and that it is not improbable that that individual would stand a fair chance of correctly identifying a substantial proportion of network members by their speech, we can start to hypothesise in a principled way what a working minimum might be. In view of what is expected of audiences of TV and more particularly radio shows themed around impersonation of celebrities and politicians (Eriksson, this volume), we need not restrict ourselves to considering just individuals known personally to the listener. Further research will help to resolve this question.

2.1 Sources of individual variation

So what sorts of features make voices different from each other? Some variation is attributable to differences in anatomy - the dimensions of the oropharynx, dentition, palatal arch curvature, vocal fold thickness, and so on. Others are related to vocal tract function. An example is the degree of vocal fold adduction during phonation: incomplete closure yields a breathy or whispery voice quality, for instance (Laver 1980). Habitual failure to lower the velum so as to admit airflow into the nasal cavity, giving the voice a denasalised 'adenoidal' quality, is another. The anatomical structure of an individual's vocal tract is essentially fixed in early adulthood, although foreign objects (orthodontic braces, piercings to the tongue or uvula, and so on) may sometimes be fixed into the vocal tract, and with ageing certain significant morphological changes take place (Beck 1999). Ossification of the laryngeal cartilages and tooth loss are examples. Some vocal characteristics may be pathological in origin - the harsh irregular phonation nicknamed 'smoker's voice'

may be symptomatic of permanent damage to the vocal folds caused by tobacco smoke, and lisps, stammers and other dysfluencies may stem from problems with motor control of the speech organs (Miller, this volume). The involuntariness of certain articulatory habits and settings is, however, not always easy to judge. Some may be entirely idiosyncratic and not subject to the speaker's control, while others may be part of the mosaic of phonetic features that makes up a regional or social accent (for example, Stuart-Smith 1999; Coadou and Rougab 2007). It may be the case that the phonetic distinctiveness of a person's speech derives from a mixture of features not normally found in combination in a single individual, perhaps as a result of extended residence in different areas (see further Nolan 1983, 1993; and for other relevant discussion, Remez forthcoming).

Given the degrees of freedom involved in speech production, and therefore the huge number of possible combinations of segmental and prosodic features available to speakers, it is not at all surprising that speech patterns can (at least potentially) vary down to the idiolectal level. And this is before we consider non-phonological (grammatical and lexical) resources speakers exploit for communicative and identity-marking purposes. Studies of identical twins by Nolan and Oh (1996) and Johnson and Azara (2000) indicate that despite these individuals having vocal tracts as nearly alike as two vocal tracts can conceivably be, as well as having had closely comparable parental input and social and educational backgrounds in the majority of cases, they still exhibit differences in speech production. Although they might result from such anatomical differences as do exist, these discrepancies encourage the conclusion that even within the bounds imposed by anatomy, physiology, dialect background, and so forth, individuals can still exercise a degree of choice over how they speak.

2.2 Limitations in informal speaker recognition

For all our often impressive skills in correctly identifying speakers by their voices, these abilities are anything but infallible. Many readers will no doubt have been in the embarrassing situation of phoning a friend, family member or colleague and being mistaken in thinking that the intended person has answered, when in fact the answerer was someone else. We know rather little about how untrained listeners gauge the similarity of two voices - whatever analysis takes place must typically be fairly automatic and well below the level of conscious awareness - but errors of this sort presumably arise when there is a sufficiently close match between the vocal characteristics of the person talking and the listener's stored representation of the voice of the call's intended target. This would usually necessitate a degree of consistency in the acoustic cues to the talker's perceived sex (probably based principally on voice pitch), and in terms of his or her broad accent and voice quality characteristics, relative to those of the target individual. We can at times even get the sex of the speaker wrong, which is less surprising than it sounds given that the ranges of average fundamental frequency (the physical correlate of pitch) for men's and women's voices overlap to a considerable degree (Künzel 1989).

In other cases we may fail to recognise the voice of someone who is well known to us. This might be expected where the signal quality is degraded by extraneous noise (if a call is made from a moving vehicle, say), or because of distortion brought about for technical reasons (for example, through loss of signal strength). Also, an individual's voice characteristics can vary markedly in line with factors such as health, fatigue, intoxication, or emotional state (Nolan 2005). We may fail to recognise a familiar voice if the speech is shouted, as demonstrated experimentally by Blatchford and Foulkes (2006), and voices with which we have

previously been familiarised can be hard to identify if purposefully disguised, such as by whispering (Hollien et al. 1982; Masthoff 1996; Künzel 2000).

Accurate attribution of a voice sample to a known individual may even be difficult in near-optimal conditions. Peter Ladefoged admits that when presented with a series of good-quality recordings of a mixture of talkers of varying levels of familiarity he failed to recognise the voice of his own mother saying *hello* and a longer sentence; only when she had finished reading a 30-second passage did he suggest that the talker was 'possibly' his mother (Ladefoged and Ladefoged 1980: 49). McClelland (2008) reports comparably poor performances in a study she carried out among members of her own family. Similarly, Foulkes and Barron (2000) found that among a tightly-knit network of ten young British men reliable attribution of eight- to ten-second speech samples to the appropriate peer-group members was surprisingly variable. Misattributions of the speech of outgroup 'foils' to network members also occurred, and in one case a participant failed even to recognise his own voice. The last of these findings is probably a consequence of the fact that perception of one's own voice is mediated by sound transmission through the bones of the skull as well as through the air, so that we do not hear ourselves as others hear us.

An earwitness's age appears to relate to the reliability of his or her identifications. Listeners between the ages of 16 and 40 were found by Clifford et al. (1981) to perform better in speaker identification tasks than older (40+) listeners, while Mann et al. (1979) reported that only those children in their sample over 10 years of age could identify speakers at adult-like levels of accuracy. These findings have clear implications in forensically-relevant scenarios involving child witnesses, whose testimony must in any case be treated with particular caution (Parker et al. 2006).

The correlation between identification accuracy and listeners' confidence ratings - that is, how sure they feel about their judgements of whether a voice has been heard previously - has repeatedly been found to be alarmingly weak (for example, Philippon et al. 2007). Indeed, when witnesses are instructed to describe a voice verbally before being asked to identify the target voice in a voice parade, the accuracy of their judgements is impaired, despite their confidence ratings remaining unaffected (Perfect et al. 2002). This decline in performance is attributed to what is known as the 'verbal overshadowing' effect (see also Cook and Wilding 2001; Vanags et al. 2005). In light of these findings there is merit in considering carefully whether, when taking statements from earwitnesses to a crime, police officers should avoid asking for a description of the perpetrator's voice because the witness's memory of the voice may be compromised as a result. The risk would then be that the earwitness might fail to identify the wrongdoer when exposed to his or her voice in a voice parade, or worse, to 'recognise' an innocent foil speaker. If, on the other hand, police have not yet identified a suspect and the earwitness's description of the perpetrator's voice could lead to an arrest being made, there is little alternative but to elicit a verbal description. The methods used by police forces in the UK and elsewhere to obtain earwitness descriptions of voices in general appear rather *ad hoc* (with some exceptions, such as the detailed interview protocol developed for use in the Netherlands), and research on how best to gather relevant information from witnesses while minimising the influence of overshadowing is urgently needed. Comprehensive summaries of existing literature on earwitness reliability may be found in Bull and Clifford (1984), Broeders and Rietveld (1995) and Kerstholt et al. (2004).

2.3 Somatic impairment and speaker identification

It seems clear from both informal observations and experimental evidence that individuals vary widely in their ability to identify people solely by their voices. In rare cases, this ability is severely impaired or altogether absent. This condition, known as phonagnosia, is normally acquired through damage to the right cerebral hemisphere resulting from stroke or other injury (Van Lancker et al. 1988). However, the first reported case of developmental phonagnosia came to light only in 2008 (Garrido et al. 2009). KH, an otherwise normal 60 year old woman, has extreme difficulty recognising voices, including her daughter's. Surprisingly, KH had had a successful management consultancy career even though she had avoided answering the telephone unless the caller had specified a time in advance. Garrido and her colleagues assessed KH's skills in recognising the voices of celebrities, identifying emotional information in speech samples, general speech perception, and processing of non-speech sounds. They conclude that because KH exhibited no sensory or cognitive impairments except in her ability to assign names to the celebrity voice samples she heard, those areas of the brain which handle memory for individual voices must be neurologically distinct from those responsible for more general speech processing tasks.

A particularly well-developed faculty for recognising individual voices has been anecdotally claimed for blind listeners. The assumption, it appears, is that the lack of one sense is compensated for by another, which then becomes unusually acute. Research on the topic has failed to demonstrate that visually impaired listeners have any advantage over normally-sighted individuals, however. Although Bull et al. (1983) found that blind subjects outperformed sighted ones in a series of voice identification tests, more recent research refutes their results. Eladd et al. (1998)

simulated a robbery witnessed by three groups of listeners: voice identification experts, totally blind people, and untrained control listeners with normal vision. The listeners then tried to identify which voice among a line-up of foil voices was the one they had heard during the robbery. Correct identification was most accurate among the voice experts, and the blind listeners performed no better than did the untrained sighted listeners. Contradictory results in this area have apparently not deterred Belgium's federal police service from recruiting a unit of blind officers because they are thought to be more skilled than sighted analysts at discriminating voices and determining place of origin by accent in recordings of criminal activity or of the speech of suspects (Macaskill 2008).

It should be noted, however, that trained listeners of the sort enlisted by Eladd's group do not necessarily perform very much better than untrained ones in speaker identification tasks. Shirt (1984) compared the performance of phonetically naïve subjects with that of 20 volunteer phoneticians in a set of tests in which both groups listened to the same materials. The phoneticians' average accuracy scores were in many cases not markedly higher than those of the untrained subjects, although the former group's individual scores tended to be more consistent with one another. It could be concluded from her results that extensive training in phonetics does not automatically make one a better judge of voice similarity. We should remember, however, that Shirt's study lacked forensic realism - her voice samples were very short, for instance, and she did not distinguish between types of error, some of which were made for valid phonetic reasons.

Experts of course do not have to rely exclusively on their ears, and the instrumental aids to analysis that are available to contemporary forensic phoneticians are more developed than they were when Shirt conducted her study. The results of a

collaborative exercise reported by Cambier-Langeveld (2007) make encouraging reading, in that while the experts who participated in the mock speaker comparison case made use of a wide assortment of methods (fully automatic, semi-automatic, and auditory-acoustic) and had varied linguistic backgrounds and levels of casework experience, the number of correct judgments greatly exceeded the number of incorrect ones.

There is now also greater control over how forensic speech science is practised, at least in Europe and North America, than was the case until quite recently. In part this has come about through the foundation in 1991 of the International Association for Forensic Phonetics and Acoustics (IAFPA), a principal aim of which is to develop and enforce standards and best practice among those working in the field. In the following section we consider some of the methods analysts apply in casework involving samples of recorded speech.

3. Technical speaker identification

The majority of work undertaken by forensic speech analysts, at least in the UK, is of two main types: speaker profiling and speaker comparison. Profiling is carried out when no suspect has yet been identified, as for example when recordings of anonymous phone calls from kidnappers or bomb hoaxers are produced. Its purpose is to narrow down the population of possible suspects by identifying linguistic features associated with certain geographical and social groups, and any unusual pronunciations that may be attributable to exceptional anatomical or pathological characteristics.

Ellis's (1994) study of the 'confession' recorded by an individual claiming to be the serial killer 'The Yorkshire Ripper' is a well-known example of speaker

profiling. The speaker on the tape was obviously accentually from north-eastern England, but through careful listening and consultation of published sources on accent variation Ellis identified features such as (h)-dropping in the word *having*, a diphthongal /uː/ vowel, and the use of [ai] in *strike*, that would place the talker's origin more specifically in the city of Sunderland. Next, comparison of the recording against reference samples from other Sunderland males narrowed down the speaker's likely provenance to the northern suburbs of Southwick and Castletown. The correctness of Ellis's accent profile of 'Wearside Jack', whose confession turned out to be a hoax, was confirmed in 2005 after the arrest of John Humble, a man who had grown up just one mile from Castletown, and who readily confessed to having prepared the hoax tape more than twenty years earlier (French et al. 2007). Humble's speech was still remarkably similar to that on the tape, in spite of the effects of age and alcohol abuse.

Speaker comparison, as the name implies, is based upon close comparison of two speech samples with a view to estimating the likelihood that the samples were produced by the same person. The expert's task is to look for points of similarity and difference between the voice of the speaker in the 'questioned' or 'disputed' sample and that of a speaker whose identity is known. The disputed sample could have been made covertly by the police, or seized as evidence, for example from a video camera or an answerphone belonging to a suspect. Or it might have been made incidentally, for example by a bystander who used a mobile phone to record an assault on a third party. The 'known sample' is typically a recording of a suspect in police custody, but it may also be a recording of a telephone call (say, to a bank) in which the caller's identity is not in question.

In the UK, and in many other jurisdictions around the world, an analysis procedure based on a combined auditory-acoustic method is most frequently used. Repeated and careful listening to samples is undertaken alongside detailed instrumental scrutiny of digitised copies using dedicated acoustic phonetic software. As a first step, transcriptions of a range of segmental (vowel and consonant) features and notes on observations of the prosodic characteristics of the samples (intonation, rhythm, tempo, voice quality) are made using IPA symbols and other specialised notation. Any other relevant linguistic information - hesitation markers, dysfluencies, non-standard grammar, unusual lexis such as dialect words or slang terms, and so forth - is also noted, as it may be of evidential value.

Acoustic properties of the samples are then measured. Software packages such as *Praat* (Boersma and Weenink 2009) allow extraction of statistics relating to the fundamental frequency of a talker's voice (mean, range, standard deviation), and measurements of features such as vowel formants and voice onset times of stop consonants are also generally straightforward if the recorded material is of adequate quality. The speaker-discriminant potential of vowel formant trajectories is currently being assessed by groups in the UK and Australia (McDougall and Nolan 2007; Morrison 2008), and the relative stability of formants over extended stretches of speech is also considered to have particular value in forensic speaker identification (Nolan and Grigoras 2005). Speech articulation rate can be expressed in syllables per second (Jessen 2007) and rhythmic properties can be quantified using indices such as the Pairwise Variability Index (Low et al. 2001). Voice quality variations may be related to characteristic patterns in the harmonic spectrum (Gobl and Ní Chasaide 1992; Nolan 2005), though as yet the forensic tools for impressionistic labelling and acoustic measurement of this particular aspect of the speaker's voice are

comparatively underdeveloped. This is perhaps surprising considering that experts are frequently presented with speech samples which are segmentally very similar but markedly divergent in terms of voice quality.

As time goes on it is becoming increasingly common for recordings to be subjected to automated analysis by machine only, and indeed in some continental European jurisdictions the method is preferred. For this task, programs like *Loquendo ASR* (www.loquendo.com) and *BATVOX* (www.agnitio.es) have been developed. Impressively high accuracy rates are claimed for these packages by their manufacturers. Speaker verification systems are becoming more commonly used for other applications - for example in computing, banking and building-access security systems - and there are proposals to include speech samples as part of the biometric data stored on individuals by government security agencies (Woodward et al. 2003). Voice data retained for security purposes is the only form of biometric information not directly related to measurements of visible features of the human body, but the currency of the popular term 'voiceprint' encourages the misperception that individuals possess vocal profiles that are at some level as immutable as physical attributes like fingerprints or facial features.

It may come as a surprise to some readers that at present, in spite of the aforementioned technological developments and the research that underpins them, we know of no one speech feature - analogous to a fingerprint - that can be used to single out an individual from a sample of sociolinguistically-comparable speakers. Just as the presence of a particular pronunciation in a person's speech (say, a dentalised 'lisped' /s/ and /z/, or the use of labiodental [ʋ] for /r/) may contribute to the distinctiveness of his or her voice, so too may the absence of a feature. It may be the case, for instance, that on the basis of what is typically heard in the social or regional

accent of the speaker one would expect to observe features which in fact do not occur, or are found only sporadically. An example might be the absence of linking and intrusive /r/ in phrases like *you're about* or *pizza instead* in the speech of a talker using an accent in which (like most non-rhotic British accents of English) the majority pattern is to produce an overt rhotic consonant at the word boundary. Especially problematic is the fact that the envelope of variability defining a single person's speech ('intraspeaker variation') will almost certainly overlap with those of other speakers ('interspeaker variation'). This must be taken into account when assessing whether the differences we inevitably observe between two samples are likely to indicate that the samples were spoken by two different talkers or the same one. That is, are the differences sufficient in nature and in number to allow us to rule out the possibility that they arose as a result of intraspeaker variability - which can in some cases be on quite a considerable scale - and thereby to eliminate the known suspect as the talker in the disputed sample?

If we consider multiple phonological features in combination we can generally identify what makes Speaker A's voice different from the other voices in our sample, but it does not follow that we can then say with any certainty that Speaker A is linguistically unique among the population at large. The number of speakers who may share that same set of features is unknown. For this reason, any judgement we make about the degree of correspondence between two speech samples not known in advance to have been produced by the same person should, where feasible, be cast in terms of the likelihood ratio (LR) of the Evidence. The evidence is the observed difference(s) between the suspect and offender speech samples. The LR is then the probability of these differences assuming the prosecution hypothesis (same talker) is correct, relative to the probability of the differences assuming that the defence

hypothesis (different talkers) is correct (Rose 2006). Where it is not possible to express an opinion in this way - which is in reality almost always, because in most cases we lack population statistics on the distribution of speech features even in well-described languages like English - the use of likelihood statistics should be avoided altogether. The position statement published by a working group of UK-based forensic speech scientists in 2007 (see http://www.forensic-speech-science.info) recommends instead that the expert's decisions be expressed in terms of the *consistency* and *distinctiveness* of samples. If analysts find similarities between two samples that, in their opinion, are sufficient to satisfy them that the samples are consistent with one another - that is, they could have been produced by the same talker - the question then becomes one of how distinctive the combination of features heard in both samples is in the context of the wider population. At the low end of the distinctiveness scale we have 'not distinctive' (the samples are consistent but there are no features of special note) while at the high end is 'exceptionally distinctive', a label used when 'the possibility of this combination of features being shared by other speakers is considered to be remote'.

A good deal of ongoing research aims to identify speech parameters which would help forensic experts to link recordings to individual talkers more reliably than is presently possible. It is true that significant advances in this area have been made in recent years, and we should not devalue the methods currently used given the success with which they have often been applied in criminal investigations. However, analysts are duty-bound to inform legal professionals, jurors, and the general public of the limitations of these methods, an obligation necessitated further by the so-called 'CSI effect', whereby laypeople's expectations are raised to unrealistic levels by the misleading portrayal of forensic speech analysis on television and in film (Schweitzer

and Saks 2007). This misconception stems in part from the notion that because human listeners can identify individuals by voice the task must be one that machines can do at least as easily, and probably very much faster and more accurately. After all, modern computers are, by any standards, capable of some extremely impressive feats. As we saw in section 2, however, human listeners are in fact not as good as we might like to believe at speaker recognition, and even the best machines available are at present unable to accomplish what we see them do in the movies. As yet they cannot cope sufficiently well with factors such as channel mismatch (telephone speech versus recordings made in quiet conditions, for example), differences in voice quality and pitch brought about by emotional state or by the Lombard background-noise-compensation reflex (Hirson et al. 1995), and other intraspeaker variability exhibited by talkers in forensically realistic situations.

4. Conclusions

As should be clear from the brief overview presented in this chapter, it is prudent given our current state of knowledge to approach the idea of a one-to-one mapping between individual people and voices with some scepticism. It would be true to say that at a general level people do have distinct voices - professional impersonators make a living on this basis, and an underlying assumption made by forensic speech analysts and computer programmers working on ASR and speaker verification is that although there is always a chance that two people will share precisely the same vocal characteristics, the odds of this actually occurring in the scenarios of central concern to professionals in these areas are typically very slim.

Nevertheless, the experimental work on earwitnessing and memory for voices, which shows that we are often not especially good at identifying even familiar voices

- including our own - compels caution. Consistency in the methods used to elicit statements about perpetrators' voice characteristics from victims and witnesses is lacking virtually everywhere at present, it seems, and the extent to which verbal overshadowing may influence the quality of earwitness evidence is still largely unknown. Further research should be done on the latter before attempting to address the former. There is also much to do in terms of convincing laypeople, police officers and legal professionals of the non-existence of the 'voiceprint', despite what is claimed by some software manufacturers and reinforced by unrealistic representations of forensic speech science in the popular media.

It should be pointed out, lest these observations strike the reader as reasons for alarm or pessimism, that while the field of forensic speech analysis is still relatively young it is rapidly maturing into a branch of forensic science that bears comparison with areas more firmly established in public consciousness, such as fingerprinting, DNA profiling, toxicology or ballistics. Considerable levels of research effort and resources are being committed internationally to improving and standardising procedures and analytical methods in forensic speech science, and tighter controls over who is permitted to practise it and to present expert evidence in law courts are being imposed in many countries. Qualms about levels of governmental surveillance of private citizens of the sort voiced recently by the British House of Lords (2009) certainly give grounds for serious concern. But we can at least weigh any curtailment of personal freedoms that new intelligence-gathering measures may entail against the knowledge that improved, more reliable speaker identification methods will result in fewer errors of impunity and wrongful convictions, and a greater number of correct decisions pertaining to the identities of individuals recorded or overheard committing an offence.

References

Beck, Janet Mackenzie (1999), 'Organic variation of the vocal apparatus', in William J. Hardcastle and John Laver (eds.), *The Handbook of Phonetic Sciences*, Oxford: Blackwell, pp. 256-289.

Blatchford, Helen and Paul Foulkes (2006), 'Identification of voices in shouting', *International Journal of Speech, Language and the Law*, 13(2): 241-254.

Boersma, Paul and David Weenink (2009), Praat: doing phonetics by computer (Versions 5.1). [Computer program]. Retrieved January 31, 2009, from http://www.praat.org/.

Broeders, Ton and Toni Rietveld (1995), 'Speaker identification by earwitnesses', in Angelika Braun and Jens-Peter Köster (eds.), *Studies in Forensic Phonetics*, Trier: Wissenschaftlicher Verlag, pp. 24-40.

Bull, Ray and Brian R. Clifford (1984), 'Earwitness voice recognition accuracy', in Gary L. Wells and Elizabeth F. Loftus (eds.), *Eyewitness Testimony: Psychological Perspectives*, Cambridge: Cambridge University Press, pp. 92-123.

Bull, Ray, Harriet Rathborn and Brian R. Clifford (1983), 'The voice recognition accuracy of blind listeners', *Perception* 12: 223–226.

Burgess, Melinda and George E. Weaver (2003), 'Interest and attention in facial recognition', *Perceptual and Motor Skills* 96(2): 467-480.

Cambier-Langeveld, Tina (2007), 'Current methods in forensic speaker identification: Results of a collaborative exercise', *International Journal of Speech, Language and the Law* 14(2): 223-243.

Chaski, Carole E. (2005), 'Who's at the keyboard? Authorship attribution in digital evidence investigations', *International Journal of Digital Evidence* 4(1): 1-13.

Clark, J. Alan, P. Dee Boersma and Dawn M. Olmsted (2006), 'Name that tune: call discrimination and individual recognition in Magellanic penguins', *Animal Behavior* 72(5): 1141-1148.

Clifford, Brian R., Harriet Rathborn and Ray Bull (1981), 'The effects of delay on voice recognition accuracy', *Law and Human Behavior* 5: 201 -208.

Coadou, Marion and Abderrazak Rougab (2007), 'Voice quality and variation in English', *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, pp. 2077-2080.

Cook, Susan and John Wilding (2001), 'Earwitness testimony: effects of exposure and attention on the Face Overshadowing Effect', *British Journal of Psychology* 92: 617-629.

Dudai, Yadin (1997), 'How big is human memory, or on being just useful enough', *Learning and Memory* 3: 341-365.

Eladd, Eitan, Sima Segev and Yishai Tobin (1998), 'Long-term working memory in voice identification', *Psychology, Crime and Law* 4(2): 73-88.

Ellis, Stanley (1994), 'The Yorkshire Ripper enquiry: part I', *Forensic Linguistics* 1(2): 197–206.

Foulkes, Paul and Anthony Barron (2000), 'Telephone speaker recognition amongst members of close social network', *Forensic Linguistics* 7: 180-198.

French, J. Peter, Philip T. Harrison and Jack Windsor-Lewis (2007), 'R -v- John Samuel Humble: the Yorkshire Ripper Hoaxer trial', *International Journal of Speech, Language and the Law* 13(2): 255-273.

Garrido, Lúcia, Frank Eisner, Carolyn McGettigan, Lauren Stewart, Disa Sauter, J. Richard Hanley, Stefan R. Schweinberger, Jason D. Warren and Brad Duchaine (2009), 'Developmental phonagnosia: a selective deficit to vocal identity recognition', *Neuropsychologia* 47(1): 123-131.

Gobl, Christer and Ailbhe Ní Chasaide (1992), 'Acoustic characteristics of voice quality', *Speech Communication* 11: 481-490.

Grant, Tim (2008), 'Approaching questions in forensic authorship analysis', in John Gibbons and M. Teresa Turell (eds.), *Dimensions of Forensic Linguistics*, Amsterdam: John Benjamins, pp. 215-231.

Hammersley, Richard and J. Don Read (1985), 'The effect of participation in a conversation on recognition and identification of the speakers' voices,' *Law and Human Behavior* 9(1): 71-81.

Hill, Russell A. and Robin Dunbar (2003), 'Social network size in humans', *Human Nature* 14(1): 53-72.

Hirson, Allen, Peter French and David Howard (1995), 'Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics', in Jack Windsor-Lewis (ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J.D. O'Connor*, London: Routledge, pp. 230-240.

Hollien, Harry (2002), *Forensic Voice Identification*, San Diego, CA: Academic Press.

Hollien, Harry, Wojciech Majewski and E. Thomas Doherty (1982), 'Perceptual identification of voices under normal, stress, and disguised speaking conditions', *Journal of Phonetics* 10: 139-148.

Hollien, Harry and Reva Schwartz (2001), 'Speaker identification utilizing noncontemporary speech', *Journal of Forensic Science* 46(1): 63-67.

House of Lords Select Committee on the Constitution (2009), *Surveillance: Citizens and the State, vol. I: Report*, London: The Stationery Office. URL: http://www.publications.parliament.uk/pa/ld200809/ldselect/ldconst/18/18.pdf

Jessen, Michael (2007), 'Forensic reference data on articulation rate in German', *Science and Justice* 47: 50-67.

Johnson, Keith and Misty Azara (2000), *The perception of personal identity in speech: evidence from the perception of twins' speech*. Unpublished manuscript. URL: http://johnsonazara.notlong.com

Kerstholt, José H., Noortje Jansen, Adri Van Amelsvoort and Ton Broeders (2004), 'Earwitnesses: effects of accent, retention and telephone', *Applied Cognitive Psychology* 20(2): 187-197.

Künzel, Hermann J. (1989), 'How well does average fundamental frequency correlate with speaker height and weight?', *Phonetica* 46: 117-125.

Künzel, Hermann J. (2000), 'Effects of voice disguise on speaking fundamental frequency', *International Journal of Speech, Language and the Law* 7(2): 150-179.

Ladefoged, Peter and Jenny Ladefoged (1980), 'The ability of listeners to identify voices', *UCLA Working Papers in Phonetics* 49: 43-51.

Laver, John (1980), *The Phonetic Description of Voice Quality*, Cambridge: Cambridge University Press.

Low, Ee-Ling, Esther Grabe and Francis Nolan (2001), 'Quantitative characterisations of speech rhythm: syllable-timing in Singapore English', *Language and Speech* 43(4): 377-401.

Macaskill, Mark (2008), 'Blind taught to 'see' like a bat', *The Sunday Times*, 10th February 2008. URL: http://www.timesonline.co.uk/tol/news/uk/article3341739.ece

Mann, Virginia A., Rhea Diamond and Susan Carey (1979), 'Development of voice recognition: parallels with face recognition', *Journal of Experimental Child Psychology* 27: 153-165.

Mäntylä, Timo (1997), 'Recollections of faces: remembering differences and knowing similarities', *Journal of Experimental Psychology: Learning, Memory and Cognition* 23: 1203-1216.

Masthoff, Herbert (1996), 'A report on a voice disguise experiment', *Forensic Linguistics* 3: 160-167.

Meudell, Peter, Bernice Northen, Julie S. Snowden and David Neary (1980), 'Long term memory for famous voices in amnesic and normal subjects', *Neuropsychologia* 18(2): 133-139.

McCarty, Christopher, Peter D. Killworth, H. Russell Bernard and Eugene Johnsen (2001), 'Comparing two methods for estimating network size', *Human Organization* 60(1): 28-39.

McClelland, Elizabeth (2008), 'Voice recognition within a closed set of family members', Paper presented at the International Association for Forensic Phonetics and Acoustics 2008 Conference, Lausanne, Switzerland, July 2008.

McDougall, Kirsty and Francis Nolan (2007), 'Discrimination of speakers using the formant dynamics of /u:/ in British English', *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, pp. 1825-1828.

McGehee, Frances (1937), 'The reliability of the identification of the human voice', *Journal of General Psychology* 17: 249-271.

Morrison, Geoffrey S. (2008), 'Forensic voice comparison using likelihood ratios based on polynomial curves fitted to the formant trajectories of Australian English /aɪ/', *International Journal of Speech, Language and the Law* 15(2): 247-264.

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

Nolan, Francis (1993), 'Auditory and acoustic analysis in speaker recognition', in John Gibbons (ed.), *Language and the Law*, London: Longman, pp. 326-345.

Nolan, Francis (2005), 'Forensic speaker identification and the phonetic description of voice quality', in William J. Hardcastle and Janet Mackenzie Beck (eds.), *A Figure of Speech*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 385-411.

Nolan, Francis and Catalin Grigoras (2005), 'A case for formant analysis in forensic speaker identification', *International Journal of Speech, Language and the Law* 12(2): 143-173.

Nolan, Francis and Tomasina Oh (1996), 'Identical twins, different voices', *Forensic Linguistics* 3: 39-49.

Parker, Janat F., Elizabeth Haverfield and Stephanie Baker-Thomas (2006), 'Eyewitness testimony of children', *Journal of Applied Social Psychology* 16(4): 287-302.

Perfect, Timothy J., Laura J. Hunt and Christopher M. Harris (2002), 'Verbal overshadowing in voice recognition', *Applied Cognitive Psychology* 16(8): 973-980.

Philippon, Axelle C., Julie Cherryman, Ray Bull and Aldert Vrij (2007), 'Lay people's and police officers' attitudes towards the usefulness of perpetrator voice identification', *Applied Cognitive Psychology* 21(1): 103-115.
Pollack, I., J.M. Pickett and W.H. Sumby (1954), 'On the identification of speakers by voice', *Journal of the Acoustical Society of America* 26(3): 403-406.
Quiroga, Rodrigo Q., L. Reddy, Gabriel Kreiman, Christof Koch and Itzhak Fried (2005), 'Invariant visual representation by single neurons in the human brain', *Nature* 435: 1102-1107.
Remez, Robert (forthcoming), 'Spoken expression of individual identity and the listener', in Ezequiel Morsella (ed.), *Expressing Oneself/Expressing One's Self: A Festschrift in Honor of Robert M. Krauss*, London: Taylor and Francis.
Rose, Philip (2006), 'Technical forensic speaker recognition: evaluation, types and testing of evidence', *Computer Speech and Language* 20: 159-191.
Schweitzer, Nicholas J. and Saks, Michael J. (2007), 'The CSI Effect: popular fiction about forensic science affects public expectations about real forensic science', *Jurimetrics* 47: 357-364.
Shirt, Marion (1984), 'An auditory speaker-recognition experiment', *Proceedings of the Institute of Acoustics* 6(1): 101-104.
Stuart-Smith, Jane (1999), 'Glasgow: accent and voice quality', in Paul Foulkes and Gerry Docherty (eds.), *Urban Voices: Accent Studies in the British Isles*, London: Arnold, pp. 201-220.
Vanags, Thea, Marie Carroll and Timothy J. Perfect (2005), 'Verbal overshadowing: a sound theory in voice recognition?', *Applied Cognitive Psychology* 19: 1127-1144.
Van Lancker, Diana R., Jeffrey Cummings, Jody Kreiman and Bruce Dobkin (1988), 'Phonagnosia: a dissociation between familiar and unfamiliar voices', *Cortex* 24(2): 195-209.
Windsor-Lewis, Jack (1994), 'The Yorkshire Ripper Enquiry: part II', *Forensic Linguistics* 1: 207-216.
Woodward, John D., Nicholas M. Orlans and Peter T. Higgins (2003), *Biometrics: Identity Assurance in the Information Age*, Berkeley, CA: McGraw Hill Osborne Media.
Yarmey, A. Daniel (2004), 'Eyewitness recall and photo identification: a field experiment', *Psychology, Crime and Law* 10: 53-68.
Yarmey, A. Daniel, Meagan J. Yarmey and Leah Todd (2008), 'Frances McGehee (1912-2004): the first earwitness researcher', *Perceptual and Motor Skills* 106(2): 387-394.