6th International Criminal Law Congress, Melbourne 1996.
Forensic Issues Session, October 11th.

Observations on Forensic Speaker Identification

Phil Rose
Senior Lecturer in Phonetics and Chinese Linguistics
Dept. Linguistics (Arts)
Australian National University
philip.rose@anu.edu.au.

AIM

To clarify some major points in current **Forensic Speaker Identification** practice using a list of "weaknesses in the science" as perceived by the legal profession (Hayne & Crockett, 1995: 2,3).

MODELS OF INFORMATION CONTENT IN SPEECH

Simple model: transmission of information in "Speech Chain": idea -> production (how acoustic disturbances are produced by speaker) -> acoustics (physical structure of transmitted speech wave) -> perception (how listener decodes acoustic structure to understand its information content.)

Fact - Acoustic output of a speaker is uniquely determined by the speaker's anatomy.

Fact - it is relatively easy, even for medium quality recordings, to extract and quantify acoustic parameters from recorded speech, and to use these to characterise the speaker *as they are speaking on that particular occasion.*

Fact - It is anything but easy to apply this information forensically.

PROBLEMS IN APPLICATION

1) Fact - Complexity of both: a) information content in speech and b) the way it is encoded. Adequate model of information content and relationship to speech (Nolan 1983, 1996:2/3): two types of content: **VOLITIONAL** (linguistic message; self-presentation; attitude to listener;...) & **NON-VOLITIONAL** (age; sex; physique; health; socio-economic background;...).

These different types of information are not encoded in separate channels, so it is difficult (but not impossible) to untangle.

E.g. 1. Speaker's sex/physique will be encoded in fundamental frequency, (acoustical correlate of rate of vibration of vocal cords, approx. = pitch), but so will *inter alia* their psychological state (non-volitional), their attitude towards speaker (volitional), and aspects of the linguistic message (question vs. statement etc.). So the same (high) fundamental frequency measurement can reflect speaker A with small vocal cords speaking normally, or speaker B with large cords speaking excitedly. Same acoustics therefore do not necessarily indicate same speaker.

E.g. 2 Speaker's sex/physique will be encoded in vocal-tract resonance or 'F-pattern' (acoustical correlate of size and shape of vocal tract), but so will their different vowels (volitional), and their choice of vowels to reflect sociolinguistic factors like formality of utterance. So speaker A's F-pattern for his vowel in car will be different when speaking formally from speaking informally (the difference will also be audible to a trained ear). Different acoustics do not necessarily therefore reflect different speakers.

Forensic Phonetic Point - Forensic phonetic analysis must be done by someone who can demonstrate understanding of the complexity of the relationship between speech (as output of vocal mechanism) and its information content. This knowledge is part of Linguistics (the 'science' of Language), not specific to forensic phonetics. Typical qualifications would be *Ph.D* in speech science (phonetics, signal processing, engineering), with strong background in Linguistics.

2) Although acoustics are determined by anatomy, the vocal tract is plastic, and capable of a wide range of shapes, which are used (volitionally or non-volitionally) to encode content. Consequently a speaker is associated with a correspondingly wide RANGE of acoustic values (not a set of quasi invariaerprints).
Variation is inherent in the system, both contemporaneously and (especially) non-contemporaneously. *Voices are not like fingerprints. There are no invariant cues to speaker identity*.

Fact - *it has not yet been established that everybody speaks differently from everybody else (that everyone has their own idiolect.)*

We do not know whether an individual inhabits their own range of variation within multidimensional acoustic space (whether between-speaker variation is greater than within-speaker variation), but there are some indications that this is so. It is a separate question whether it is possible to make use of this putative idiolectal region of acoustic variation under the considerable limitations imposed by typical forensic circumstances. It is probably the case that between-speaker variation is not always greater than

within-speaker variation. Rose (1996) and Rose & Duncan (1995) show that 6 male speakers, similar sounding enough to be misidentified by close relatives, all differ significantly from each other providing access is had to a wide enough range of acoustic information. Restricting this information to ranges typical for forensic circumstances means that from 1 to 3 out of the 15 pairs of similar sounding speakers cannot be distinguished acoustically.

3) Since a typical forensic situation involves an effectively open class verification, probabilistic approaches are required. *But there is almost a complete lack of statistical information on distribution of acoustic parameters in relevant populations (similar-sounding male speakers of general-broad Australian?). This also applies to auditory characteristics*. This means we are as yet unable to quote probabilities of two recordings coming from the same vocal tract using classical statistical estimation procedure (but see Jones 1994) .

TESTING

It may be concluded from the degree of difficulty in **FSI** outlined above that testing is crucial. Haynes & Crockett (1995) (H&C) correctly caution (point #6):

'Sixthly, there is no system in place for testing the tester. An individual may apply and follow their own technique. That individual is not tested to determine their ability, nor their accuracy, nor their competence.'

Two points must be made with respect to testing. **FSI** practitioners are not tested, although the **IAFP** (International Association for Forensic Phonetics) announced early this year (1966) that an anonymous test will be conducted in which members can participate. It is also possible that blind tests will be implemented by the Forensic Speaker Identification Standards Committee of **ASSTA** (Australian Speech Science and Technology Association) in the near future. 2). More importantly (since individuality of method is subordinate to the demonstration of efficacy) *few practitioners have conducted tests on the basis of which they can report on the reliability of their method(s)*. It seems reasonable to assume that the court is entitled to be told at the least: 1) the number of false positives a particular expert's method generates (when it is claimed that two recordings come from the same speaker and they don't), and (2) the number of false negatives.

ACOUSTIC vs. AUDITORY METHODS IN FSI

H&C (point #6) mention different techniques in **FSI.** They comment further (point #8):

'... used the auditory method only. She found that the acoustic method -

that is the mechanical method - could not be utilised, because of the quality of the video. However, as she put it, 'It is very nice to have the acoustic method as corroboration, that is confirmation of the findings one makes from the auditory method'. It was not able to be employed in this case, to either confirm or disagree, or cast doubt on the auditory findings.

There are several misleading points here, the most important of which concerns the relationship between the two different methods of analysis.

The overwhelming consensus among practitioners, for solid theoretical reasons, is that both auditory and acoustic approaches are indispensable in **FSI**. H&C were therefore correct in their criticism. Most practitioners will first listen carefully to recordings, and make a narrow (detailed) phonetic transcription. This fulfills several functions. It is to judge the quality of the material for forensic examination; to listen for dialectal differences between samples which can indicate different speakers; to listen for inconsistencies in articulation which may indicate disguise; to listen for 'idiosyncrasies' e.g. speech defects, unusual articulations or co-articulations. Auditory analysis is, however, usually to gain an idea of how much comparable material exists that can be used for an acoustic analysis.

Fact - *Neither approach on its own is adequate, on principle*.

It is well known since Nolan (1990) that auditory methods are inadequate. Nolan (1990) reports a case where two prosecution experts had identified the suspect's voice acoustically as the criminal voice, with a very high degree of certainty. Acoustic analysis showed it was extremely unlikely that the same speaker was involved. The perceptual mechanism can resolve two different acoustic patterns from different speakers in the same way, such that it is not possible to hear the difference between the two. This demonstrates conclusively that **FSI** must involve acoustic as well as auditory analysis. In addition, acoustic analysis lends itself to quantification and statistical analysis in a way that auditory analysis does not.

Acoustics alone are inadequate for the following reason: it is a commonplace that the same component of a physical signal can correspond to two different linguistic units. For example, in a tone language, like Cantonese, a male high tone can have the same fundamental frequency (so many Hertz) as a female low tone. The human perceptual mechanism, which is used in auditory analysis, 'knows' to allow for such differences.

Since most (Australian) cases do not involve voices with very different accents, or constellations of orthogonal speech defects, the relationship between the two approaches is usually one of auditory analysis constituting an essential preliminary to acoustic analysis. Acoustics should definitely not be considered as a corroboratory adjunct to auditory analysis as

implied in the quote above.

ON FSI AS AN EMERGING DISCIPLINE

H&C comment (point #3):

'...there are precious few publications in Australia or overseas that form the basis of learning and instruction in the field. Exchange of information in the scientific field is an essential basis for research and fine tuning and recognition of and elimination of error, and fallacies in any discipline. And so you have medical conferences held all over the world, legal conferences held over the world, conferences in all sorts of scientific matters. People have become members of or fellows of particular organisations.'

They continue (point #4):

'..there are international gatherings, conferences, fellowships, an integral part of the development of sciences and an international and national basis of such organisations are very much in the voice identification field in their fledgling formative stages.'

These comments are somewhat off the mark There are, to be sure, few Australian publications exclusively devoted to **FSI**. Possible exceptions are the speaker recognition sections in the biennial proceedings of the Australian Speech Science and Technology Association published since 1986. (Speaker recognition research is currently pursued for obvious technological reasons in many centres in Australia, and world-wide). However, there exists an international Journal of Forensic Linguistics in which papers on **FSI** are regularly published, and the literature on speaker recognition is now voluminous (see Hollien (1990) for a bibliography). So it is incorrect to assume that the field is characterised by a dearth of or little exchange of information. Moreover, the International Association for Forensic Phonetics has been established since 1991, and has just held its international conference in Wiesbaden. The last International Conference on Forensic Linguistics, which included papers and a forum on **Forensic Phonetics**, was held in Armidale in 1995.

The latest development is the setting-up (in 1995) of a Forensic Speaker Identification Standards Committee by the Australian Speech Science and Technology Association. This committee is now considering proposals for accreditation, and blind testing, (both of these in collaboration with the Police), and other aspects of code of practice. This should help bring about an improvement in standardisation of methodology, and terminology. In sum, the discipline is therefore somewhat more mature than is appreciated in legal circles.

CAN IT BE DONE?

In view of the difficulties involved in **FSI**, it is sensible to ask whether it is in fact possible. The answer appears to be a qualified "yes, but as yet asymmetrically". There have been some cases which phonetic evidence has been instrumental in resolving, for example Nolan (1991), and Labov and Wendell (1994). It is notable that most of these cases have involved proving the defence case, that the accused was not in fact the speaker. Jones (1994: 354-355) mentions a case where acoustic phonetic analysis contributed to (an apparently correct) conviction. Nevertheless, it remains the case that the conditions under which most practitioners would be prepared to say that different speakers are involved are clearer than those under which they would say that the same speaker is involved.

H&C correctly observe (point #2):

'...the techniques of voice identification are still under development'

Until more information becomes available on the distribution of acoustic parameters in varieties of (Australian) English, and testing is carried out on reliability of method, the most that the courts can reasonably expect is a statement that the evidence (acoustics/auditory impression) is consistent with recordings having from the same speaker - not a probability judgment that the same speaker is involved. It is difficult to imagine at the moment that **FSI** evidence alone would constitute a sufficient basis for a case (auditory identification by a familiar non-expert is another matter). Nevertheless, both kinds of statement that **FSI** is at present capable of are not without corroborative value.

==========

SELECT ANNOTATED BIBLIOGRAPHY/REFERENCES

In addition to the references in the text, this contains a short annotated list of useful, or frequently cited items. * = recommended for legal practitioners.

Baldwin and French (1990) Forensic Phonetics. Pinter [Often mentioned, but although interesting, and despite blurb, not a text that will be of much use to the legal profession. It is anecdotal, superficial, and self-contradictory - see review by Nolan (1991).

John Gibbons (1994) Language and the Law. Longmans [Accessible volume containing three of the papers mentioned in this bibliography].

Alex Jones (1994) 'The limitations of voice identification' In John Gibbons (ed.) (1994):346-361. [Somewhat muddled account of between-speaker differences, but contains examples of forensic application of a classical statistical technique in acoustic analysis which is now considered somewhat problematic - see Robertson and Vigneuax (1995)].

Hayne J.A. & Crockett A.J.A. (1995) Supreme Court of Victoria Court of Appeal Report 325/94 **® v Bell**).

Harry Hollien (1990) The Acoustics of Crime. Plenum [Serious, reliable and up-to-date, with a good research bibliography, but too wide-ranging, and not focussed enough to be very useful from a legal perspective.]

*William Labov and Wendell A. Harris (1994) 'Addressing social issues through linguistic evidence'. In John Gibbons (ed.) 1994: 287-302 ('The Prinzivalli case'). [A nice example of straightforward auditory dialect ology, rather than acoustics, resolving a case.]

*F. Nolan (1983) The phonetic bases of speaker recognition. Cambridge Studies in Speech Science and Communication, Cambridge U.P. [Based on author's Cambridge Ph.D. Builds on Laver (1980) to provide a comprehensive model of phonetic between-speaker differences. **FSI** experts should be expected to have a good knowledge of the contents. Introduction, and chapters 1 & 2 give a good, detailed and clear explanation of the complexity of the problem of speaker recognition]

**F. Nolan (1990) 'The limitations of auditory speaker identification'. In H. Kniffka (ed.) Texte zu Theorie und Praxis forensischer Linguistik. Max Niemayer: 457-479 [A vital paper, explaining why auditory approaches are inadequate. Proponents of exclusively auditory approaches should have to justify this in view of Nolan's argument].

F. Nolan (1991) 'Forensic Phonetics' (Review of Baldwin and French (1990) Journal of Linguistics 27 483-493.

F. Nolanciences.
Blackwell. [Very clear overview of the state of the art, but not out yet.]

*F. Nolan (1994) 'Auditory and acoustic analysis in speaker recognition'. In John Gibbons (ed.) (1994): 326-345. [Includes a summary of information in Nolan (1990)].

F. Nolan & T. Oh (1996) 'Identical twins, different voices' Journal of Forensic Linguistics 3.

F. Nolan (1996) 'Forensic Phonetics'. Notes from the two-week course given at the 1996 Australian Linguistics Institute, Australian National University, Canberra.

J.M.D. Laver (1980) The phonetic description of voice quality. *Cambridge U.P.* [Another standard text that **FSI** experts should know. Provides a framework for auditory (and acoustic) description of between-speaker differences which has been partially adopted in speech therapy.]

*Peter Ladefoged (1983) A Course in Phonetics. Harcourt Brace Jovanovich 3rd ed. [University introductory text-book. Chapter 8 'Acoustic Phonetics' contains nice introduction to acoustic phonetics.]

Bernard Robertson & G.A. Vigneaux (1995) Interpreting Evidence. *Wiley.*

Phil Rose & Sally Duncan (1995) 'Naive auditory identification and discrimination of similar voices by familiar listeners' *Journal of Forensic Linguistics 2/1: 1-17.*

Phil Rose (1996) 'Speaker verification under realistic forensic conditions' To appear in Proc. 6th International Australian Conference on Speech Science and Technology **ASSTA.**

Phil Rose Ph.D. (Cambridge), M.A. (Manchester), Dip. International Phonetics Association (London), Zwischenpruefung (Erlangen).
Member, Forensic Speaker Identification Standards Committee, Australian Speech Science and Technology Association.
Member, International Association of Forensic Phonetics.
Former Member of Council, International Phonetics Association.