Tutorial on *Forensic Speech Science*

Part I: Forensic Phonetics

Anders Eriksson
Department of Linguistics, Gothenburg University,
Gothenburg, Sweden

1. Introduction

In this part of our tutorial we will cover three areas.

First we will position forensic speech science in its historical context. We will briefly describe the development from its modern beginnings with *Voiceprints* in the USA and *Phonoscopy* in Stalin's USSR to the present situation by giving you a glimpse of the advances as well as the controversies.

Secondly we will present some fundamental issues in forensic phonetics research and see how our present knowledge can be applied in forensic fieldwork. In this part of the tutorial, the focus will be on human voice recognition and discrimination and how factors like memory, familiarity, language, disguise etc. influence these abilities.

Lies and deception are age-old problems in forensic investigations. It is therefore not surprising that many people have tried to find ways of detecting deception. We will describe attempts to use the information conveyed by the human voice in lie detection and recent experiments using brain scan methods (fMRI) to approach the problem. In the description of these efforts we will also make an attempt to draw the line between methods, the use of which in forensics, can be justified on scientific grounds and methods which are speculative at best and outright bogus in the worst case.

For each topic we will present a list of suggested reading.

2. Background reading

Several textbooks on forensic phonetics have been published during the last decades, many of which will be found in university libraries. Here we will only mention two recent books which provide good and comprehensive introductory reading. *Forensic Voice Identification* by Hollien, is an introductory textbook, which sketches a historical background of the field and covers topics like automatic speech recognition, memory and voice lineup procedures. The book is fairly non-technical and does not require any in depth phonetic knowledge. The book by Rose, *Forensic Speaker Identification*, is considerably more technical in nature. It deals with automatic speaker identification and covers some of the techniques used, like cepstrum analysis, in some depth. It is an excellent introduction to the field of automatic methods for the reader who has a reasonable background in speech technology or acoustic phonetics. The book also covers statistical problems and methods involved in speaker verification evaluation like Bayesian statistics. Both books are highly recommended reading.

Hollien, Harry. (2002). *Forensic Voice Identification*. San Diego, CA: Academic Press.
Rose, Phil. (2003). *Forensic Speaker Identification*. New York: Taylor & Francis.

3. Historical background

The use of voice identification in criminal cases has a longer history than one might think. It has probably been used to identify suspects who were heard but not seen committing a crime for thousands of years. In more recent times there are many records of the use of speaker

identification as evidence in courts. A well known early case is the trial of William Hulet in 1660. Hulet was accused of having executed King Charles I. A witness, Richard Gittens, testified that he had heard the executioner, whose face was obscured, beg the king's forgiveness and that he knew that it was Hulet "by his speech". The jury found Hulet guilty of high treason and he was sentenced to death. But this case is also one of the first known cases of speaker misidentification. Before the death sentence was carried out it was found that the actual murderer was the ordinary hangman, who later confessed and Hulet was, as a consequence, set free. The details of the Hulet case are typical of many cases even today almost 350 years later. The perpetrator is heard but not seen. The witness may feel certain that the identification is correct, but as this and many other cases show that is no guarantee. The heard voice may be known or unknown. In the Hulet case the witness thought the voice belonged to a person known by him, but this was obviously a mistake.

Even though it is quite common that part of the evidence in a case is a report by an earwitness (who may or may not at the same time be the victim), in most cases the voice of the perpetrator does not belong to someone known by the witness. It is also quite common that a period of time on the order of weeks or more has passed between the crime and a later attempt at deciding whether the voice of a suspect is the same as that of the perpetrator. When this is the case, the accuracy of the witness' memory of the voice becomes a crucial issue. An important question in this context is how the memory of a voice decays over time. The first attempts to answer this question (see section 4.2!) were inspired by a testimony in the Lindbergh case. The son of the famous aviator Charles Lindbergh was kidnapped on March 1, 1932. A ransom letter was found in the boy's room where the kidnapper demanded $50.000. Negotiations followed by letter and by advertisements in a local newspaper. The Lindberghs agreed to pay the ransom, and on the night of April 2, 1932, Lindbergh drove his negotiator, Condon, to a cemetery where the ransom money was delivered. Lindbergh was waiting in his car and could hear (but not see) the kidnapper calling Condon saying: "Here, Doctor. Over here! Over here!" Five weeks later the boy was found dead. The police eventually tracked down a suspect, Bruno Hauptmann, and arrested him. In September 1934, 29 months after hearing the voice in the cemetery, Lindbergh in disguise at the DA's office, heard Hauptmann repeat the call heard in the cemetery. Lindbergh said he recognized the voice as the one he had heard. At the trial, in January 1935, he testified under oath that he recognized Hauptmann's voice.

The invention of telephones and recording equipment opened new areas for forensic phonetics. Analysis tools suitable for acoustic analysis of speech were also developed. A milestone in the latter development was the invention of the spectrograph. Most of this work was done at the *Bell Telephone Laboratories* from the late thirties and onwards. The construction of their spectrograph was based on ideas suggested by Steinberg (*JASA*, vol 8, 1934, pp. 16–24). A spectrograph of basically the same type was later produced by *Kay Elemetrics* and sold commercially under the brand name *Sonagraph*.

The original motivation behind the development of the spectrograph was the phonetic study of speech – "a method of approach to studies of speech production and measurement" (Steinberg). A real time spectrograph called *Direct Translator* where the spectrogram was displayed on a florescent screen was also produced. Its intended use was as an aid in pronunciation training for the deaf and foreign language students.

As we all know today, the spectrograph has been, and is, a very valuable tool for phonetics research. Its usefulness as an aid for the deaf or pronunciation training in foreign language education never reached the initial expectations but these expectations did at least generate quite a bit of relevant research. It may therefore come as a surprise to find that hardly anything was published during the first years (the late thirties to 1945) of this development and research. The explanation for this is the fact that the project was rated as a war project. In

one of the first publications after the war Potter (1945) who was a researcher at Bell Labs writes: “The work here described was begun before the war. Because of related war interests it was given official rating as a war project, and has progressed far enough during the war period to justify its being brought now to public attention”. Now it is highly unlikely that the American military would view the development of a training aid for the deaf as a war project. Exactly what it was that they hoped to get out of it we do not know much about but people have speculated that what the military really wanted was a reliable method for speaker identification. Meuwly (2003) remarks: “La participation des États-Unis à la deuxième guerre mondiale a donné naissance à un projet d'application militaire du spectrographe sonore: l’identification de navires ennemis par l'intermédiaire de la voix de leurs opérateurs radio.“ This sounds reasonable enough, but as we have said, hardly anything has been published about this part of their research efforts. It is worth noting, however, that when two of the researchers at Bell, Grey and Kopp, published an in-house report towards the end of the war they used the term “voiceprint” to refer to spectrograms with obvious metaphorical reference to fingerprints and the same term is used in their first published paper (Grey and Kopp, 1944). Voiceprint is also the term that came to be used in connection with the somewhat infamous history of speaker identification by spectrograms that was to follow some years later initiated by a former engineer at Bell by the name of Lawrence Kersta. (See 4.1)

It is also interesting to note that when the people at Bell started to publish again after the war, little mention of speaker identification is to be found, but instead all the previously mentioned topics: “the deaf will benefit greatly ... and students of foreign language ... It offers an objective means of verifying existing phonetic concepts and of extending our knowledge of the spoken language” (Kopp and Green, 1946). It is also worth noting that they seem to stress inter-speaker similarities more than individual differences: “Visible patterns of the same words and sentences spoken by different individuals, show that the similarities in diction are much greater than the differences”. Potter (1945), while recognizing individual differences, also emphasizes the similarities in the patterns for different speakers.

If the people at Bell Labs sponsored by the military, secretly worked on “voiceprints” for speaker identification purposes, as we have good reasons to believe, then the early history of the “voiceprints” follows very parallel tracks in the Soviet Union, including the fact that we know very little about it. The only (?) account of the Soviet efforts we have is the novel *The First Circle* by Solzhenitsyn. The plot of the novel takes place within a time-span of only three days during the Christmas Holiday of 1949. The setting is the Mávrino prison at the outskirts of Moscow were the Stalinist regime held “unreliable” scientist imprisoned, and Solzhenitsyn was one of them. The prison, which was divided into several sections with their own specialties, held around 300 prisoners. It was in *Number Seven*, considered as the most important and prestigious sector, were the work on various speech technology related projects took place. It had its own *Acoustics Laboratory* and a laboratory called the *Clipped Speech Laboratory*, where the work was focussed on finding efficient ways of coding speech so that it would be difficult or impossible to decode by “the enemy”. One day the focus shifted, at least temporarily, from voice “clipping” to voice “recognition”. A “criminal” telephone call from an unknown speaker believed to be working in the Ministry of Foreign Affairs to a professor of medicine warning him against sharing his research results with foreign colleagues had been intercepted and taped. Five people at the ministry were prime suspects but the police did not know which of them had made the call. So, the scientists at the Acoustics Laboratory were given a tape that contained the recorded call and recorded voice samples of the five suspects and were given the task of identifying one of the suspects as the caller. They were given only two days to complete the task, with Siberia as a likely alternative option. Solzhenitsyn’s description of the technology and analysis tools used is not detailed enough to draw any definite conclusions, but it is obvious from the vocabulary used (e.g.

Voiceprints, Vocoder) that the scientists were well aware of corresponding work done outside the Soviet Union. They did not use a spectrograph in the modern sense of the word, but it is likely to have been something similar to the model described in Steinberg's (1934) paper.

Given the alternatives it is hardly surprising that they did indeed succeed with their voice recognition efforts, although they could only narrow down the number of suspects to two out of the five, something they found disturbing. It is obvious, though, that they viewed the possibility of reliable voice recognition quite optimistically. Here is a quote from the novel which gives you a feeling of the mild euphoria that characterized their initial success. Rubin, the one responsible for the experiment gives a first report to the officer who has come to be informed about the results.

Rubin: *Only the beginning! Only the most tentative deductions, Adam Veniaminovich!.*
Officer: *And what are they?*
Rubin: *They're are open to dispute, but one thing is incontrovertible. The science of phonoscopy, born today, December 26th 1949, does have a rational core.*

And the future looked no less promising:

"*They envisioned the system, like fingerprinting, which would someday be adopted: a consolidated audio-library with voiceprints of everyone who had at one time or another been under suspicion. Any criminal conversation would be recorded, compared, and the criminal would be caught straight off, like a thief who had left his fingerprints on the safe door*."

The novel ends here and what happened next, we do not know. No (?) accounts of the further development of speaker identification during that era are available. It is worth noting, however, that the term *phonoscopy*, used to refer to forensic phonetics and coined by the people at Mávrino is still used in Russia and in many former East European countries

Suggested reading

Grey, G. and G. A. Kopp (1944). "Voiceprint identification." Bell Telephone Laboratories Report: 1–14.

Hollien, Harry. (2002). *Forensic Voice Identification*. San Diego, CA: Academic Press. (Ch. 2)

Kopp, G. A. and H. C. Green (1946). "Basic phonetic principles of visible speech." Journal of the Acoustical Society of America **18**: 74–89.

Meuwly, D. (2003). "Le mythe de « L'empreinte vocale » (I)." Revue internationale de criminologie et de police technique et scientifique **56**(2): 219–236.

Potter, R. (1945). "Visible patterns of speech." Science **November**: 463–470.

Solzhenitsyn, A. I. (1968). *The First Circle* (T. P. Whitney, Transl.). Evanston, Ill: Northwestern University Press.

Steinberg, J. C. and N. R. French (1946). "The portrayal of visible speech." Journal of the Acoustical Society of America **18**: 4–18.

4. A detailed look at some fundamental issues.

In the following sections we will present a selection of important issues in forensic phonetics and suggest relevant reading for those of you who want to dig a little deeper. The books and papers we refer to will in most cases be available through reasonably well-equipped university libraries.

4.1 Voiceprints

There was nothing controversial about the use of the word "voiceprint" when it first appeared in the paper by Grey and Kopp (1944) and it is quite understandable that they were inspired by the patterns they saw in the spectrograms to make comparisons with fingerprints and speculate about the possibility of using spectrograms to identify speakers just as fingerprints are used to identify individuals. The controversy arose much later when an engineer at Bell Labs by the name of Lawrence Kersta started to use voiceprints in forensic applications. We

do not know much about Kersta's scientific involvement in the early work on the spectrograph at Bell. He was an engineer and not a researcher, but was or became the head of the research lab so he must have been acquainted with the scientific questions as well.

As we have seen in the previous paragraph, researchers at Bell did not promote the idea of voice fingerprints when they started to publish after the war but focused on traditional phonetic research questions and applications in education and therapy. If they also worked on speaker identification, we do not know. It may well be the case that that part of the work was still classified. At any rate there was total silence from Bell about speaker identification from the first publications in 1944 until 1962 when Kersta published a paper in *Nature* titled V*oiceprint identification* and later the same year gave a talk at the annual meeting of the Acoustical Society of America titled *Voiceprint-identification infallibility*. His claims regarding the accuracy of speaker identification by voiceprints were extraordinary. Based on visual comparison of key words, examiners achieved no less than 99% correct identification or better. In 1966 he left Bell and started his own company called *Voiceprint Laboratories Corporation* and started to offer his services in criminal investigations and to train people in voiceprint identification. Up until that point in time his claims remained largely unchallenged by the scientific community. He therefore enjoyed some initial success and his testimonies were accepted as evidence by courts in some, but not all, states.

But he soon began to meet with resistance. Subjects in a study by Young and Campbell (1967), using the voiceprint technique, obtained 78.4% correct identifications for a training material consisting of two words spoken in isolation but when the same words taken from different contexts were used, identification went down to 38.3%. Stevens *et al*. (1968) let subjects perform speaker identification both aurally via headphones and visually from spectrograms. The error rate for the aurally presented stimuli was 6% compared to 21% for visual identification. False alarm rate was also high in the visual test. In both studies there was considerable variation in the identification scores for individual speakers, some speakers being much more difficult to recognize than others. Numerous other studies gave basically the same results, error rates for voiceprint identification were high, in many cases very high.

It would be unfair, however, not to mention that Kersta's method also had supporters. Most of the supporters were not researchers in any relevant field, but at least one of them, Tosi, was a qualified phonetician. Tosi set up a lab in his department and started to test Kersta's ideas in numerous experiments. After two years of work he published a paper (Tosi *et al*., 1972) where the results were summarized. The reported error rates were typically 5–15%.

The controversy continued until the late eighties with Koenig (1986) representing FBI and others on the defending side and Shipp *et al*. (1987) and others on the critical side. Voiceprinting is still done by private detectives and other non-academic "experts" but nobody in the speech science community believes in its usefulness for forensic purposes any more.

For those of you who read French, an excellent overview of the voiceprint controversy may be found in Meuwly (2003a,b).

For an amusing account of an enthusiastic layman's view, I recommend the book by Block if you can find it.

Suggested reading

Block, E. B. (1975). *Voiceprinting: How the Law Can Read the Voice of Crime.* New York: David McKay Company, Inc.

Kersta, L. G. (1962). "Voiceprint identification." Nature **196**: 1253–1257.

Kersta, L. G. (1962). "Voiceprint-identification infallibility." Journal of the Acoustical Society of America **34**: 1978.

Koenig, B. E. (1986). "Spectrographic voice identification: A forensic survey." Journal of the Acoustical Society of America **79**: 2088-2090.

Meuwly, D. (2003a). "Le mythe de « L'empreinte vocale » (I)." Revue internationale de criminologie et de police technique et scientifique **56**(2): 219–236.

Meuwly, D. (2003b). "Le mythe de « L'empreinte vocale » (II)." Revue internationale de criminologie et de police technique et scientifique **61**(3): 361–374.

Shipp, T. E. T. Doherty and H. Hollien. (1987). "Some fundamental considerations regarding voice identification." Journal of the Acoustical Society of America **82**: 687-688

Stevens, K. N. *et al*. (1968). "Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material." Journal of the Acoustical Society of America **44**: 1596–1607.

Tosi, O. *et al*. (1972). "Experiment on voice identification." Journal of the Acoustical Society of America **51**: 2030–2043.

Young, M. A. and R. A. Campbell (1967). "Effects of context on talker identification." Journal of the Acoustical Society of America **42**: 1250–1254.

4.2 Auditory voice recognition and memory

As we mentioned in section 3, the Lindbergh case raised questions about voice recognition accuracy and memory. A researcher who questioned whether it would be possible to accurately remember an unknown voice over a period of more than two years was a psychologist by the name of Francis McGehee. She performed two studies (McGehee 1937, 1944) where voice recognition as a function of the time interval between first hearing the voice and a later attempt at recognizing the voice in a voice line-up was tested. In the first experiment the listeners heard (but did not see) a speaker read a 56-word passage. The listeners were then assigned to groups who heard the speaker as one of the speakers in a voice line-up with five foils at intervals of 1, 2, and 3 days, 1, 2 and 3 weeks and 1, 3, and 5 months respectively. Recognition rate varied as a function of time starting at a little over 80% correct identifications after a lapse of 1 day or 1 week. After 2 weeks the recognition rate had fallen to 69%, after a month to 57%, after 3 months to 35% and after 5 months it was down to 13%, which is less than chance. The results of this experiment on voice recognition are in general agreement with other studies of memory decay over time. The second series of experiments presented in McGehee 1944 differed from the first one mainly in that recorded voices were used instead of live voices behind a screen as was the case in the first series of experiments. But the results in the two experiments are very similar. She also made sub-studies were other factors were varied like familiarity with the language but these results will not be discussed here. Later studies have in general confirmed her findings although the precise decay rate may vary from study to study. References to some of the later studies are found in the reading list.

Suggested reading

Clifford, B. R., H. Rathborn and R. Bull. (1981). The effects of delay on voice recognition accuracy. *Law and Human Behavior,* **5**, 201–208.

Papcun, G., J. Kreiman and A. Davis. (1989). Long-term memory for unfamiliar voices. *Journal of the Acoustical Society of America,* **85**, 913–925.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology,* **17**, 249–271.

McGehee, F. (1944). An experimental study of voice recognition. *Journal of General Psychology,* **31**, 53–65.

Saslove, H. and A. D. Yarmey. (1980). Long-term auditory memory: speaker identification. *Journal of Applied Psychology,* **65**, 111–116.

4.3 Non-contemporary speech samples

The term refers to speech samples, which are obtained at different points in time and later used in an identification process. We know, of course, that the human voice changes over time. But the change is normally rather slow. The relevant question in the context of forensic phonetics is at what separation in time between speech samples, change over time becomes a problematic factor. Over very long periods of time we have reasons to expect marked changes. For obvious reasons, there are few longitudinal studies of this kind. In the one by Endres *et al* (1971), recordings of 7 speakers sampled over a time interval of up to 29 years were compared. The authors found a downward trend, as a function of increasing age, for

fundamental frequency and formant frequencies. In forensic cases it is unusual, to say the least, that a time span on the order of decades separates a suspect recording and a later attempt at identifying the speaker using a recent recording. But time spans of a year or more are not unusual. It is therefore important to know if voice changes that take place over a period of one or a few years may affect the accuracy of speaker recognition. This question has been addressed in a series of studies by Hollien and Schwartz (2000, 2001). In their experiments they tested latencies (between recordings) from 4 weeks up to 20 years. There was a drop in correct identification from around 95% for contemporary samples to 70–85% for latencies from 4 weeks to 6 years (with no observable time trend in the interval). For the 20-year latency, however, a sharp drop down to 35% could be observed. Two factors, other than latency in time between recordings, were also tested – listener experience and similarity between voices. As might be expected, experienced phoneticians performed markedly better than students. For the phoneticians correct ID was as high as 76% for the 20-year latency. Similarity between voices had a dramatically degrading effect, however. Performance dropped from just under 95% for contemporary samples to just over 40% for samples recorded 4 weeks later. For the latencies we normally have to work with in forensic investigations, non-contemporary speech samples thus seem to affect identification only marginally, at least if trained phoneticians are used as listeners and voices are not too similar.

Suggested reading

Endres, W., W. Bambach and G. Flösser. (1971). Voice spectrograms as a function of age, voice disguise, and voice imitation. *Journal of the Acoustical Society of America,* **49**, 1842–1848.

Hollien, H. and R. Schwartz. (2000). Aural-perceptual speaker identification: Problems with noncontemporary samples. *Forensic Linguistics,* **7**, 199–211.

Hollien, H. and R. Schwartz (2001). "Speaker identification utilizing noncontemporary speech." Journal of Forensic Sciences **46**: 63–67.

4.4 Other issues involving the speech sample

Factors that may influence identification accuracy are primarily sample duration and acoustic quality. If we first consider the influence of sample duration, we may observe that in real life investigations samples may be very short, often just a few words or a phrase or two which means that sample duration is on the order of a few seconds. In an early study by Pollack *et al*. (1954) the authors observed that identification accuracy increased as sample size (for monosyllabic words) increased, but only up to about 1.2 seconds. For longer samples they claim that phonetic variation takes over as the most important factor. They conclude that "we believe that the duration of the speech sample *per se* is relatively unimportant, except insofar as it admits a larger or smaller statistical sampling of the speaker's speech repertoire". This somewhat surprising finding has, however, been confirmed in other studies. In a study by Compton (1963), 15 recorded segments of the vowel [i] for each of 9 speakers, familiar to the listeners, were presented. The segments differed only in duration (25–2500 ms). For segments longer than about 75 ms, there was no increase in recognition rate as a function of duration. Bricker and Pruzansky (1966) presented stimuli which varied in duration as well as phonemic variation. They found that identification rate increased with duration only if the longer stimuli also contained more phonemic variation and that "Identification accuracy improved directly with the number of phonemes in the sample even when duration was controlled". In a study by Orchard and Yarmey (1995) correct identification rate was substantially higher for 8 minute stimuli compared with 30 second stimuli. No attempt was made, however, to estimate the respective contributions of duration and phonological variation, but it is likely that phonological variation must have been higher in the longer stimuli.

It is important to point out, however, that while an increase in correct identifications is desirable it is equally desirable to keep the number of false alarms down. In an earlier study

by Yarmey (1991) recognition of voices recorded over the telephone was studied. In this study the number of correct identifications increased as the durations of the samples increased from 3.2 and 4.3 minutes to 7.8 minutes. But so did the number of false alarms. The same type of trade off between hits and false alarms was observed in a similar study by Yarmey and Matthys (1992): “The facilitating effect on identification of longer voice-sample durations was counteracted by the high false alarm rates in both suspect-present and suspect-absent line-ups”. To minimize false alarms is, of course, very important in real-life forensic situations but is nevertheless often overlooked.

Acoustic quality is a wide topic and can be thought of in many different ways. The most common quality problems in the forensic phonetics are background noise and bandwidth of recordings or transmissions. By background noise we mean everything except the part of the speech signal, which contains information about the speaker we are interested in. It thus includes such diverse things as random noise as well as someone talking or a radio playing in the background. There is a wealth of studies in the area of automatic speech recognition and dialog systems where these things have been investigated, but there is not much work done on how human speaker recognition is influenced by background noise. While recognizing the importance of this topic we will not have anything more to say about it here.

A large proportion of threats are done over the telephone and criminals often use telephones when they plan or coordinate crimes. Telephone quality speech has therefore received attention in forensic phonetics studies. Telephone lines have limited bandwidth. Most of the frequencies relevant for speech transmission are covered, but not all. Frequencies below 300 Hz are filtered out for example. With mobile phones, problems related to speech coding are introduced. These effects are particularly noticeable for female voices.

Important questions in the forensic context are whether the poorer sound quality of recorded telephone conversations adversely affects voice identification and if so to what extent and how. Also, from a methodological point of view one would like to know whether one should only use voices recorded over the telephone in lineups where the incriminating call is recorded over the telephone. There are surprisingly few studies that address this question, but there are some results which indicate that the problem might not be as serious as one might expect. For example Rathborn, Bull and Clifford (1981, cited in Yarmey, 1991) “failed to find any significant differences in voice identification of a target voice heard originally over the telephone and tested using a taped lineup over the telephone, in contrast to voice identification heard originally over the telephone and tested directly with a taped lineup.

A question that has received some attention lately is the influence of the band-pass filtering that occurs in telephone transmissions on acoustic analysis of voice samples. In a recent study, Künzel (2001) found that the relatively high (300 Hz) lower cut-off frequency had the effect of shifting F_1 in German vowels upwards compared to the corresponding tokens in a simultaneous DAT-recording. The average size of the shift was 6.6% for male and 6.1´% for female speakers and all the differences were significant at the 5% level or better. Other, but minor, artefacts were observed as well. As a consequence, Künzel warns against using formant data for speaker identification purposes if the recordings were made from telephones. His results have not been questioned, but his total rejection of the use of formant data in speaker identification based on telephone recordings has been challenged by Nolan (2002).

Suggested reading

Bricker, P. D. and S. Pruzansky (1966). "Effects of stimulus content and duration on talker identification." Journal of the Acoustical Society of America **40**: 1441–1450.

Compton, A. J. (1963). "Effects of filtering and vocal duration upon the identification of speakers aurally." Journal of the Acoustical Society of America **35**: 1748–1752.

Künzel, H. J. (2001). "Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies." Forensic Linguistics **8**: 80–99.

Nolan, F. (2002). "The 'telephone effect' on formants: A response." Forensic Linguistics **9**: 74–82.

Orchard, T. L. and A. D. Yarmey (1995). "The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification." Applied Cognitive Psychology **9**(3): 249–260.

Pollack, I., J. M. Pickett, *et al*. (1954). "On the identification of speakers by voice." Journal of the Acoustical Society of America **26**: 403–412.

Yarmey, A. D. (1991). "Voice identification over the telephone." Journal of Applied Social Psychology **21**: 1868–1876.

Yarmey, A. D. and E. Matthys (1992). "Voice identification of an abductor." Applied Cognitive Psychology **6**: 367–377.

4.5 Familiarity with the speaker

We have all experienced recognition of a familiar voice and recognition is often fast and accurate. Even short non-linguistic stimuli like coughs are often enough to recognize a familiar person. But these informal observations are not enough. In forensic phonetics we must be more precise about the influence of familiarity on voice recognition accuracy. There are at least a few studies, which have addressed this question. Hollien *et al*. (1982) studied speaker identification as a function of speaker familiarity under three different speaking conditions, normal, disguised and stressed. Listeners who were familiar with the speakers performed significantly better under all conditions. These results have generally been confirmed in other studies (e.g. Schmidt-Nielsen and Stern, 1985).

It is important to point out, however, that although recognition rates are generally high for familiar speakers, recognition is by no means always perfect. For individual speakers and listeners the error rate can run as high as 30–40% if the utterances are short and belong to a fairly large open set (Ladefoged and Ladefoged, 1980). An influence of utterance length on the recognition of familiar speakers has also been found in other studies. In a series of experiments reported by Rose and Duncan (1995), recognition of familiar speakers varied from chance level to nearly perfect as a function of utterance length.

It has been generally assumed that in voice recognition, discrimination constitutes the initial step with recognition occurring as a later phase in a single process. But Van Lancker *et al*. (1985) have shown that this does not seem to be that case, but that "discrimination and recognition are not stages in one coherent process, but are dissociated, unordered abilities". It is therefore entirely possible that a listener who is good at recognizing familiar speakers may perform badly if the task is to discriminate between unfamiliar speakers.

Suggested reading

Hollien, H., W. Majewski and E. T. Doherty. (1982). "Perceptual identification of voices under normal, stress, and disguise speaking conditions." Journal of Phonetics **10**: 139–148.

Ladefoged, P. and J. Ladefoged (1980). "The ability of listeners to identify voices." UCLA Working Papers in Phonetics **49**: 43–51.

Rose, P. and S. Duncan (1995). "Naive auditory identification and discrimination of similar voices by familiar listeners." Forensic Linguistics **2**: 1–17.

Schmidt-Nielsen, A. and K. R. Stern (1985). "Identification of known voices as a function of familiarity and narrow-band coding." Journal of the Acoustical Society of America **77**: 658–663.

Van Lancker, D. and J. Kreiman (1985). "Unfamiliar voice discrimination and familiar voice recognition are independent and unordered abilities." UCLA Working Papers in Phonetics **62**: 50–60.

4.6 Disguise

Voice disguise, to the extent that it is used, may be a serious problem for speaker identification. In the extreme end of the spectrum we find electronic manipulation or even communicating via speech synthesis, which would make speaker identification virtually impossible. In the world of real forensic work, however, voice disguise tends to be of a rather unsophisticated nature. Künzel (2000) notes, based on experience from BKA (the German Federal Police Office), that "falsetto, pertinent creaky voice, whispering, faking a foreign accent, and pinching one's nose" are the most common types. Basically the same observations

have been made in experimental studies. In a study by Masthoff (1996) where undergraduate students served as subjects, the majority of the chosen disguises (35%) were phonation level disguises (whisper, raised pitch or lowered pitch). Articulation level disguises (dialect mimicry, foreign accent etc.) were also used (20%). The remaining disguises were combinations of two types. Electronically manipulated messages are still rare, but Künzel notes that there has been an increase in recent years, mainly in the form of editing recorded voices.

Even if the used types of disguise in most cases are rather unsophisticated, disguise may nevertheless have a considerable detrimental effect on speaker identification. In a study by Reich and Duke (1979) where various types of disguise were used, all types produced significantly fewer correct identifications. Hypernasality produced the greatest effect but there were in most cases no significant differences between the different types. Whisper, one of the more common types, resulted in markedly fewer correct identifications in a study by Orchard and Yarmey (1995) if whispered samples were compared with phonated samples. If both the reference and the test samples were whispered the difference was less pronounced.

Voice disguise is not as common as one might think. Künzel (2000) reports that:" Over the last two decades, between 15 and 25 per cent of the annual cases dealt with at the BKA speaker identification section exhibited at least one kind of disguise".

Suggested reading

Künzel, H. (2000). "Effects of voice disguise on speaking fundamental frequency." Forensic Linguistics 7: 149–179.

Masthoff, H. (1996). "A report on a voice disguise experiment." Forensic Linguistics **3**: 160–167.

Orchard, T. L. and A. D. Yarmey (1995). "The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification." Applied Cognitive Psychology 9(3): 249–260.

Reich, A. R. and J. E. Duke (1979). "Effects of selected vocal disguises upon speaker identification by listening." Journal of the Acoustical Society of America **66**: 1023–1028.

4.7 Foreign accents, and foreign languages

The influence of foreign accents or foreign languages on speaker identification has been investigated in a number of studies. It is generally found that foreign accent makes identification more difficult, but the difference is usually small and not always present. In the study by McGehee (1937) mentioned above, a study of the influence of foreign accent was included as a substudy. The recognition of a speaker of English with a German accent was tested. No difference in recognition rate was found. "An unfamiliar foreign (German) voice was recognized by approximately the same percentage of auditors as an unfamiliar American voice when each occurred under similar conditions." In other studies, however, differences have been found. In a study by Doty (1998), native speakers of English from the US and England and speakers of English as a foreign language from France and Belize were recorded reading English sentences. With native speakers of English as listeners, recognition rate was dramatically higher for other native speakers than for speakers with a foreign accent. The results from a study by Goldstein, *et al.* (1981) fall somewhere in between: "With relatively long speech samples, accented voices were no more difficult to recognize than were unaccented voices; reducing the speech sample duration decreased recognition memory for accented and unaccented voices, but the reduction was greater for accented voices". As may be seen, the results are somewhat ambiguous, but we may perhaps conclude that there is a tendency for accented voices to be less well recognized, although the difference is often small. It is also highly likely that experienced professionals, like linguists, are better at recognizing accented voices than lay listeners.

The influence of foreign language has also been the subject of many studies. In a study by Thompson (1987), six bilingual male students recorded messages in English, Spanish, and English with a strong Spanish accent. The lineup message was delivered in the same language and accent as the initial message. Voices were best identified (by monolingual English speaking listeners) when speaking English and worst when speaking Spanish. Identification accuracy was intermediate for the accent condition. Schiller and Köster (1996) tested Americans with no knowledge of German, Americans who knew some German, and native German speakers using recordings of six native German speakers. Subjects with no knowledge of German made significantly more identification errors than other subjects. Subjects who knew some German performed similarly to native German speakers.

Köster and Schiller (1997) duplicated the experiment by Schiller and Köster using Spanish and Chinese listeners. “It was found that the Spanish and Chinese listeners who were familiar with German showed better recognition rates than Spanish and Chinese listeners with no knowledge of German, whereas the Spanish and Chinese listeners with a knowledge of German performed measurably worse than the German and English listeners with a knowledge of German”.

We may summarize the results by saying that listeners with no knowledge of a language perform worse on voice recognition than listeners with some knowledge or native speakers, while listeners with some knowledge of the language tend to perform on the same level as native speakers.

Suggested reading

Doty, N. D. (1998). "The influence of nationality on the accuracy of face and voice recognition." American Journal of Psychology 111: 191–214.

Goldstein, A. G., P. Knight, *et al*. (1981). "Recognition memory for accented and unaccented voices." Bulletin of the Psychonomic Society **17**: 217–220.

Köster, O. and N. O. Schiller (1997). "Different influences of the native language of a listener on speaker recognition." Forensic Linguistics **4**: 18–28.

Schiller, N. O. and O. Köster (1996). "Evaluation of a foreign language speaker in forensic phonetics: A report." Forensic Linguistics **3**: 176–185.

Thompson, C. (1987). "A language effect in voice identification." Applied Cognitive Psychology **1**: 121–131.

4.8 Earwitnesses

Factors, which are relevant for speaker recognition in general, like memory, familiarity, disguise etc. described above are also relevant when we talk about earwitnesses, but there are some additional factors about which we presently do not know as much as we would like. As Bull and Clifford (1984) point out “the majority of (the relatively few) studies of earwitnessing bear little resemblance to real-life witnessing circumstances. Most have used nonstressful situations with prepared subjects participating in laboratory situations”.

Firstly, the stress that witnesses may experience in a real life situation can never be fully recreated in a laboratory experiment. Neither can we, or the witness, have much experience to draw on that will help us determine just how and how much the capabilities of a traumatized victim to recognize a voice or discriminate between voices may be affected. Secondly, “personal experience of voice recognition, is always of familiar voices – the voices that are *not* usually those to be identified in criminal situations” (Bull and Clifford). And as we know from the work by Van Lancker and Kreiman (see 4.5), recognizing a familiar voice and discriminating between unfamiliar ones are independent abilities. And thirdly, whereas subjects in a laboratory experiment are, to a higher or lesser degree, prepared for the situation, real life witnesses are in most cases not. Studies have shown (e.g. Clifford and Denot, 1982, cited in Bull), that voice identification accuracy under unprepared conditions is much lower. Witness confidence is of no great help either. Bull reports significant correlations between

accuracy and confidence, but other studies (e.g. Yarmey, 2001) have not found such correlations.

Suggested reading

Broeders, A. P. A. and A. C. M. Rietveld (1995). Speaker identification by earwitnesses. In A. Braun and J.-P. Köster (Eds.), Studies in Forensic Phonetics. Trier: Wissenschaftlicher Verlag, 24–40.

Bull, R. and B. R. Clifford (1984). Earwitness voice recognition accuracy. Eyewitness Testimony: Psychological Perspectives. G. L. Wells and E. F. Loftus. Cambridge, Cambridge University Press**:** 92–123.

Yarmey, A. D. (2001). "Earwitness descriptions and speaker identification." Forensic Linguistics **8**: 113–122.

4.9 Earwitness line-ups

An earwitness lineup (or 'voice parade') is meant to be the auditory equivalent of an eyewitness lineup. It is used when a person has heard but not seen the perpetrator. As was pointed out in 4.8, the voice is unknown to the witness in the normal case. In the lineup, recordings of a suspect's voice and the voices of a number of foils are presented for the witness whose task it is to compare the recorded voices with the memory of the heard voice and determine if any of the recorded voices matches the memory of the perpetrator's voice.

Two important questions in connection with earwitness lineups are 1) how many voices should be present in the lineup? and 2) how similar to the suspect's voice should the voices of the foils be?

It has been found that with few voices in a lineup, there may be marked position effects. It has also been found that the number of correct identifications decreases as lineup size increases. So the question is if there is an optimal size where the position effect is minimized and the decrease in correct identifications has bottomed out. There are a number of studies, which have addressed these questions, but here we will only cite one. Bull and Clifford (1984) tested the influence of lineup size on performance in two experiments. In the first experiment 5 or 11 foils were used, and in the second experiment 4, 6, or 8. There were significant differences between the results for 4 foils compared to 6 or 8, but otherwise the differences were minimal. The results thus indicate that, as a rule of thumb at least, 5 or 6 foils should be used. They also found an effect of target position only when the target came first in the array.

How similar to the target should the foils be? This is of course a difficult question with many complications, but at least two extremes must be avoided. The target voice must not stand out as different from all the rest. The speakers must be reasonably matched with respect to general characteristics like speaker age, dialect etc. On the other hand they should not be sound-alikes. When Rothman (1977, cited in Hollien, 2002) used sound-alikes (brothers, fathers, sons) identification dropped from 94% (ordinary foils) to 58% (sound-alikes). Similar results were obtained by Hollien and Schwartz (2000, see 4.3). Thus foils should be chosen so as to represent a reasonable degree of variation but avoiding the extremes.

Suggested reading

Broeders, A. P. A. and A. Amelsvoort (1999). Lineup construction for forensic earwitness identification. Proceedings of the 7th International Congress of Phonetic Sciences. San Francisco**:** 1373–1376.

Hollien, H. *et al*. (1995). "Criteria for earwitness lineups." Forensic Linguistics **2**: 143–153.

Hollien, H. (1996). "Consideration of guidelines for earwitness lineups." Forensic Linguistics **3**: 14–23.

4.10 Lie detection using fMRI and other techniques

Attempts have been made recently to use brain scanning methods in order to study the possibility of detecting consistent differences in brain activity patterns which may be used to separate lie or deception from truthful statements. Although this research is only in its infancy, some highly interesting results have been obtained. We will only touch upon this

research marginally here, however, since this tutorial is about forensic phonetics and few of the results in brain research are directly relevant with respect to forensic phonetics as such.

Langleben *et al*. (2002) used Functional Magnetic Resonance Imaging (fMRI) to see if they could detect any differences in brain activity when their subjects told a lie compared to when they told the truth. And their results indicate that there was indeed a difference: "This finding indicates that there is a neurophysiological difference between deception and truth at the brain activation level that can be detected with fMRI". Similar results have been obtained in other studies (e.g. Lee *et al*., 2002).

Interesting work using the fMRI technique in search for neural correlates of voice perception (Belin *et al*., 2004) and person (voice and face) familiarity (Shah *et al*., 2001) is also under way.

High resolution thermal imaging which can detect minor regional changes in the blood flow in the face for example has also been used in an attempt to develop methods to detect lie and deception (Pavlidis and Levine, 2002).

But we should be aware that, as at least some of the authors of the research papers point out, these are very preliminary results. We must also always keep in mind that results like the ones reported here are the results of laboratory experiments, often highly sophisticated, time consuming and costly! When, and indeed if, these methods can be put to use in forensic fieldwork we will not know for many years to come. We must also be aware that there may be a very long way to go between research results and reliable field applications. Unfortunately this is not always the case. "Unproven technologies are becoming increasingly attractive to US law enforcement and security agencies … Laboratory tools – from infrared sensors to eye trackers – are being converted into lie detectors" (Knight 2004).

Suggested reading

Belin, P., S. Fecteau and C. Bédard. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences,* **8**, 129–135.

Knight, J. (2004). The truth about lying. *Nature,* **428**(15 April 2004), 692–694.

Langleben *et al*. (2002). "Brain activity during simulated deception: an event-related functional magnetic resonance study." NeuroImage **15**(3): 727–732.

Lee, T. M. C., *et al*.. (2002). Lie detection by functional magnetic resonance imaging. *Human Brain Mapping,* **15**, 157–164.

Pavlidis, I. and J. Levine. (2002). Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine,* **21**(6), 56–64.

Shah *et al*. (2001). "The neural correlates of person familiarity." Brain **124**(4): 804–815.

4.11 Overgeneralization, charlatanry and outright fraud

The possibility of detecting lie or deception with some kind of automatic lie detector is of course something to be wished for by the police and other investigators. In the world of films and comic strips, perfectly reliable lie detectors have been around for a long time. In real reality we have not come quite that far, however. The most well known "lie detector" is the so called *Polygraph*. Its first appearance, in a rather preliminary form, can be dated back to 1917. A more refined version produced in the beginning of the twenties, was used in a court case in 1923 and Polygraphs have been used ever since with some refinements. The basic idea behind the Polygraph when used as a "lie detector" is that lying increases the level of stress in the person who is lying and if you can accurately register the involuntary reactions we know to be correlated with stress like respiration, pulse, blood pressure, galvanic skin response, you can also detect lies. The problem with using the Polygraph as a lie detector lies in the interpretation. Correlations between stress levels and pulse for example are found as group results. To generalize from group results to individuals is, of course, not a valid step. Neither is it a valid step to conclude that a person who experiences stress must necessarily be lying.

People who defend the use of the Polygraph avoid calling it a lie detector, but they nevertheless use it as if it were one, which pretty much amounts to the same thing.

The Polygraph as such is only marginally relevant in the context of forensic phonetics but the general principle of trying to find some reliably measurable correlate to lie and deception is fundamental also in attempts to construct lie detectors based on voice analysis. The idea is that some voice property can be used as a cue to lie or deception. It has been suggested that micro tremor in the voice can be used to detect deception and a number of analysers based on micro tremor analysis have been marketed. There is, however, no scientific basis for the claims that these analysers can detect deception. Hollien (1987) surveyed the literature and concluded that: “the ability of voice analyzers to detect stress from speech–or to identify spoken deception–have been negative or “mixed” in nature”. He also performed tests of his own, using commercial voice analyzers which turned out to perform at chance level: “stress/nonstress identifications occurred only at chance levels; the lie/nonlie identification scores were quite similar”. Shipp and Izdebski (1981) have tested the idea using hooked-wire electrodes inserted into the laryngeal muscles, but no micro tremor patterns at all were found. Nevertheless these products are still in use by private detectives and even in some cases by the police. Given the weak or non-existent scientific basis underlying these gadgets one feels justified in calling the use of them charlatanry at best.

But there is also outright fraud. An Israeli based company markets the most wonderful tools including both lie detectors and love detectors. The technique behind the lie detector is said to be something called Layered Voice Analysis (LVA) and the assumption is that

every “event” that passes through the brain will leave its “finger prints” on the speech flow. LVA Technology ignores what your subject is saying, and focuses only on his brain activity. In other words, the “how” it is said is crucial and not the “what”.

They are careful not to explicitly call the gadget “lie” detector, but there is absolutely no question that that is what they want us to believe it is: “LVA is capable of detecting the intention behind the lie, and by so doing can lead you in identifying and revealing the lie itself”.

As any one with even the slightest knowledge of voice analysis will know, there is not a shred of evidence for a relationship between voice and brain activity of the proposed kind. And a thorough scrutiny of the description of the method in the American patent documents confirms the suspicion that the method is pure nonsense, perhaps best described as statistics based on digitization artefacts.

You would think that a company that markets brain finger-printers and love detectors would give rise to suspicion or at least caution in prospective customers, but that does not in general seem to have been the case. The company is a million dollar business with among others some UK and US insurance companies as customers. There are also reports that its products are used by police departments in the US and perhaps elsewhere.

We may learn something from earlier experience, namely that there is a certain danger in completely ignoring charlatans. Laymen may wrongly interpret the silence as acceptance no matter how outrageous, even ridiculous, the claims may seem to an expert in the field. On the other hand it can be quite time consuming to expose them, time that will have to be taken from other, scientifically more important things. Herein lies a dilemma we must come to grips with.

Suggested reading

Hollien, H. (1987). "Voice stress evaluators and lie detection." Journal of Forensic Sciences **32**: 405–418.

Shipp, T. and K. Izdebski (1981). "Current evidence for the existence of laryngeal macrotremor and microtremor." Journal of Forensic Sciences **26**: 501–505.