Chapter 17.3: Forensic Voice Recognition and Linguistics

Learning Goals and Objectives

Language can inform legal decisions through voice recognition and the specific usage of language in written and spoken forms. In this chapter, you will need to understand the following concepts:

- What are the basics of voice recognition;
- How is voice recognition experimentally done;
- How can forensic linguistics inform legal considerations;
- What are the limitations of forensic voice recognition and linguistics.

Introduction

The ability to speak provides us with the capacity to communicate rapidly and efficiently with others. The production of speech requires the complex and coordinated movements of some rather sophisticated structures in our throats, neck, head and mouth. There are, however, subtle differences in the shapes and sizes of these many sound-producing structures between people that give rise to the individual sound quality of a particular person’s spoken voice. We’ve all experienced hearing the voice of an unseen person, such as on the telephone, radio or television, and knowing instantly the identity of that person from their voice alone. This ability is the basis of the field of forensic voice recognition.

In forensic science, voice recognition and linguistics may help to identify a speaker by their voice or speech patterns. For example, voice recognition analysis can help to identify a person by a threat left on someone’s voicemail or answering machine, a surveillance camera that captured the audio from a robbery, or a kidnapper’s call to make ransom arrangements. Linguistics can also shed important light on the questioned interpretation of the meaning of a recorded confession or on the authenticity of verbal legal instructions. In instances such as these, an analysis of an audio record may provide the crucial information about the identity and intent of the speaker.

Forensic Voice Recognition

Voice recognition for forensic applications usually is considered to have two main uses: voice identification and voice verification. Voice identification focuses upon determining who is the speaker on a recording from among a group of possibilities. The group might consist of just a few suspects or include many possibilities. Voice verification, however, deals with determining if a recorded voice belongs to one particular person and involves comparison of the questioned voice with a known exemplar for that person. Less frequent uses of voice recognition deals with intelligibility enhancement, security authentication, and other

[Image: Some of the organs and tissues involved in voice production (www.thefullwiki.org/Voice_production).]
The physical process of speech requires the carefully choreographed movements of many organs and tissues in the human body, including the brain, nervous system, lungs, larynx, tongue, jaw, and mouth (Figure 17.3.1). The human voice physically begins with the production of a positive air pressure in the lungs to generate a steady stream of air outward through the trachea (windpipe). This causes the thin tissues of the vocal cords in the larynx to rapidly vibrate (Figure 17.3.2), causing alternating slight increases and decreases in air pressure, called sound waves (Figure 17.3.3). The faster the vibration, the higher the perceived pitch of the sound will be. The rate at which these vibrations occur is called the frequency, the number of sound waves to pass a fixed point in a given amount of time. The sound created in the larynx is modified as it passes through the remainder of the vocal tract. The jaw, lips, tongue, palate, nasal cavity and other organs modulate the produced sound waves into the audible patterns that we recognize as speech.

The subtle qualities of a sound, sometimes called timbre, depends largely upon the specific size and shape of the throat, mouth and nasal cavity, something that varies significantly from person to person. As the sound passes through the throat, mouth and nasal cavity, it causes the residual air in these cavities to vibrate more or less based upon the complex natural frequencies of the space. When this happens, some frequencies of the sound are reinforced, and sound louder, while at the same time other frequencies are reduced, leading to a smaller contribution these frequencies to the overall sound. These differences allow us to distinguish sound of a trombone from a piano playing same note. It’s also the reason that we can one person’s voice from another’s. Each sound produced from these instruments or a person’s voice is actually made up a combination of many different sound frequencies, all added together to produce unique sound that we hear, called a waveform. For example, in Figure 17.3.4 is shown the different frequencies and their applications.

![Figure 17.3.2. The Larnx, including the vocal chords and glottis](http://singingolutions.com/blog/singing-anatomy).

![Figure 17.3.3. Formation of increased (compressed) and reduced (rarefied) air pressure by vibration, such as from a tuning fork or the vocal cords, to produce sound](www.cis.rit.edu/research/biomedical/ultrasoundintro/ultraintro.html).

![Figure 17.3.4. Sound frequencies that add together to give the characteristic sound of the didgeridoo](www.didgeridoo-physik.de/E/CADSD/method/method-frameset.htm).
intensities (amplitudes) that add together to give the Australian Didgeridoo its characteristic sound. Changes to the relative amount of each individual frequency that goes into the final mix changes the quality of the sound that we hear.

The differences in the size and shape of a person’s vocal cavities, coupled with the characteristic ways that they modulate and articulate their words, produces a sound pattern that experts contend is unique to each person. The probability that two people both have exactly the same physical make-up in their sound generating structures and employ the same speech articulations is extremely small. Scientists have developed a method to measure and graphically display the differences that exist in the subtleties of our speech, called a sound spectrograph or voiceprint.

A voiceprint displays three key components of speech in a single plot (Figure 17.3.5). The vertical axis is used to display the different frequencies that make up the sound while the horizontal axis shows time elapsed during the measurement. The degree of darkness of the line at a particular point shows the intensity of that particular frequency at that one time — a darker line indicates that the frequency is louder and a lighter line indicates less contribution of that frequency to the overall sound. Putting all of these features together for a given word or sentence produces a voiceprint the captures the individual features of a person’s speech, as illustrated at the bottom of Figure 17.3.5.

Voice spectrograms have been used extensively in legal settings, with well over 7,000 reported instances of voice recognition in forensic investigations and court cases. An example of how this is done is illustrated in Figure 17.3.6. In the top two spectrograms of the Figure, the same speaker (labeled Speaker A) says the phrase “This is a sound spectrogram” on two separate days. The two measured

Figure 17.3.5. A voiceprint of a person from Tennessee saying “She just lives right at about three or four blocks away from here”


Figure 17.3.6. Voice spectrographs of people saying the phrase “This is a sound spectrogram”

(www.section3b.net/research_yme/forensic_voice_identification.pdf).
spectrograms can be compared and forms the basis of a voice verification – the two voice patterns were made by the same person. The bottom plot (labeled Speaker B), however, was made by a different speaker saying the same phrase. Comparing the spectrograms for speakers A and B shows that they are quite different in their frequency patterns. Comparison of a spectrogram from a recording of an unknown speaker, for instance, with spectrograms taken from a group suspects repeating the same message can allow the elimination of some suspects while matching the unknown pattern with that of a single person.

Early work reported that voiceprints might be as individual as a person’s fingerprints. This, however, was shown not to be the case since there are natural variations observed in voiceprints of the same person saying the same phrase at different times. Several large scale studies, including one involving 250 speakers at Michigan State University and one by the FBI using 2,000 test subjects, have reported that high confidence identifications could be reached in just about 35% of the cases examined and that the error rates in these analyses were near 0.3% for false identifications and 0.5% for false eliminations (FBI study). Because of concerns raised with the methods employed in these studies, however, it has been suggested that these error rates should only be considered as the minimum potential error rates. It has been proposed that at least twenty unique indictors are required to make a valid voice identification, although there is also controversy here. Variations in instrumentation, operator skill, analyst experience, potential analyst bias, disguised voices, and other issues often cast doubt on the validity of voiceprint identifications. Voiceprint identifications are currently being challenged in a number of courts and some areas have specific restrictions upon their use in court.

Voice recognition work sometimes includes effort to improve the intelligibility of the speech found on a damaged or noisy recording. Experts have developed a number of computer-based methods to remove some of the noise and artifacts on recordings to enhance the ability to understand the speech on the tape.

**Forensic Linguistics**

Linguistics, the study of language, is increasingly being used in legal settings for understanding legal texts, informing judicial processes, and interpreting the meaning of specific language in criminal cases. Applications may include cases relating to the ability of certain suspects to understand their legal rights, the specific wording of criminal confessions, the meaning of certain statements made by participants during the commission of a crime, and the protection of the commercial rights of corporations for certain words and phrases, to name just a few.

One common application of forensic linguistics involves interpreting the language of the law, including written judgments, statutes, and legal agreements (e.g., contracts, mortgages, wills, etc.). Since linguistics deals with how meaning and intent may be communicated between people, linguists can help to interpret the meaning behind the legal prose in documents that can sometimes be particularly confusing to lay readers. Specific language is also important when communicating with juries regarding their instructions as well as in understanding courtroom testimony. How lawyers phrase questions,

---

**Trademarks to Generic Language**

Linguists have been called to testify as to when a trademark becomes so widely used that it looses its legal protection. A number of common terms in our language today have made the transition to generic usage, including:

- **Aspirin**: meaning acetylsalicylic acid and owned original by Bayer AG;
- **Band-Aids**: Protective skin covers, owned by John and Johnson;
- **Coke**: cola-containing drink, owned by Coca-Cola Corp.;
- **Escalator**: Moving stairs, originally owned by Otis Elevator Co.;
- **Google**: to search the internet, Owned by Google Inc.;
- **Heroin**: an opiate drug, originally a trademark of Bayer AG;
- **Thermos**: An insulating container, originally owned by Thermos GmbH;
- **Yo-Yo**: originally a trademark of Duncan Co.;
- **Zipper**: Interlocking fabric connectors, originally owned by B.F. Goodrich.
and how these are understood by those testifying, can play a key role in the proceedings.

Language plays an important role in aspects of the pre-trial criminal justice system including interrogating witnesses and suspects, collecting statements, and obtaining confessions. This is especially important where bilingual or non-native speakers engage with the system since phrasing, tense, and voice (e.g., active or passive) may have subtly different meanings in across languages.

Forensic linguists are occasionally asked to clarify the meaning of someone’s words or in explaining why a particular phrasing was used. For example, if a police officer asked a cornered criminal to give him their gun, an accomplice might say to the gunman “Give it to him!” But what did the accomplice actually mean – hand the officer the gun or shoot the officer (based on a real case)? The answer to this question would mean the difference between the accomplice advocating for surrender or promoting homicide.

Linguists may also become involved in trademark disputes such as how trademarks and legally protected words may be translated into another language. Additionally, when a trademark becomes part of the common generic language, it loses its legal protection and corporate ownership. For example, many people use the word “Xerox” generically to mean to make a photocopy, when the term was originally intended solely to relate to the copiers produced by the Xerox Corporation. Similar terms might include to “google” as a generic term (to search for something on the internet), have a “coke” (meaning a generic cola), or even Aspirin (meaning acetylsalicylic acid, trademark owned by Bayer).

Forensic aspects of language clearly plays key roles in many types of legal processes and, given the continued evolution of both language and the means that we choose to communicate, the field seems destined to continue to inform legal proceeding in the future.

---

**The Great Escape**

In the 1963 movie *The Great Escape* (The Mirisch Company, United Artists) about the escape of prisoner’s in a German POW camp, a tunnel is constructed from a barracks within the camp to beyond the containment fence. Two of the prisoner’s eventually make their way in disguise to a bus leaving the area. They are captured, however, when a suspicious Gestapo agent wishes them “Good Luck” in English. When one of the prisoner’s unconsciously responds back to the Gestapo agent in English, they are captured. This is an interesting application of forensic linguistics in determining a person’s native language.
Chapter 17 References and Bibliography

Jan Seaman Kelly and Brian S. Lindblom (Eds.), Scientific Examination of Questioned Documents (2nd Ed.), CRC Press, 2006.


# Glossary of Terms

**Age Regression and Progression Photography:** A process of taking a picture of a person and simulating their appearance at either a younger or older age by manipulating the picture.

**Altered writing:** Changes made to a document that including erasures, additions and deletions.

**Block capital writing:** A method of writing that uses all upper case (capital) letters that are upright and separated from one another.

**Chromatography:** A method that uses the varying interactions of different chemical compounds towards a common material to separate the components of a mixture.

**Copybook:** An instructional handwriting model.

**Counterfeit:** An imitation made to be mistaken for an genuine article, such as counterfeit money.

**Cursive writing:** A method of rapid writing in which most of the letters in a word are joined together, leaving spaces only between words.

**Demultiplexing:** The process of separating multiplexed images into the images obtained from a single source (camera).

**Dot-matrix printer:** A printer, similar to a typewriter, that uses an inked ribbon that is struck by a series of very tiny pins that combine to form the image of a letter on the paper.

**Electrostatic detection apparatus** (ESDA): A device that uses a high voltage electrostatic charge to visualize indented writing.

**Exemplar:** A writing samples of known authorship.

**Fictitious signature:** An instance where either or both the document and the signer are imaginary such that there are no exemplars and the signature cannot readily be compared with others samples.

**Frame averaging:** When several adjacent images are added together to reduce random noise and enhance the image on the recording.

**Font:** A complete set of type in one size and design:

**Forgery:** The act of preparing or altering a document, signature, financial certificate, work of art, or other item with the intention to defraud, damage or cheat.

**Frequency:** The number of waves that pass a fixed point per unit of time.

**Graphology:** A pseudoscientific field that tried to form a connection between the features of a person’s handwriting and their personality or character traits.

**Handwriting:** A person’s individual style of writing with an implement.

**Handwriting style:** A basic styles of handwriting such as block, script, and cursive.

**Hyperspectral imaging:** An imaging system that gathers light from a wide band of the electromagnetic spectrum and breaks it into smaller bands for processing – each band covering a relatively small portion of the spectrum.

**Indented writing:** When writing on a piece of paper that lies on top of other sheets causes indentations on the underlying pages.

**Ink-jet printer:** printer that forces ink through a nozzle to form letters on the page.

**Intentional forgery:** When a person intentionally adds forced or unusual features into their own signature such that they can later disclaim the signature.

**Laser printer:** A printer that uses a laser beam to form a positively charged image of what is to be printed on a rotating metal drum that is later transferred to toner and paper.

**Larynx:** The muscular and cartilaginous organ at the upper portion of the trachea, containing the vocal cords, often called the voicebox.

**Linguistics:** The study of language.

**Multiplexed images:** When multiple cameras feed into a single recorder that compile either a sequential of superimposed image.
Non-request exemplar: Handwriting sample of known origin that is written under “natural” conditions, usually before the person’s writing was of any legal interest.

Obliterated writing: Writing where the original text is concealed by adding an opaque substances to block out the ability to visibly read a portion of the document.

Panorama mosaic: A picture that is prepared by aligning multiple images that each capture a portion of the entire object.

Photography: The process of capturing and producing images by the action of light on film or a detector.

Point: A size of a font defined as 72 points per inch.

Questioned document: Any document over which there is some legal dispute regarding its origin, authenticity, or authorship.

Request exemplar: A writing sample provided by a suspect when asked.

Rhythm: The regular repetition or “periodicity” of various writing elements, such as word and letter spacing and the visual “flow” of the words on a paper.

Script handwriting: A method of writing that uses both upper and lower case letters that are not joined together and resembles most a printed text.

Signature: A person’s individualized writing of their name.

Simulation: An attempt by someone to disguise their handwriting.

Sound spectrograph: A device that records the intensities of difference frequencies of sound as a function of time, usually as someone speaks.

Specular highlights: The bright spots that result from strong illumination, such as from the sun or artificial lighting, on the surface of the subject in the picture.

Traced forgery: Use of an authentic signature as a template that is carefully followed in making the forgery.

Trash marks: Dirt and scratches on the window of a photocopier that are seen as small specks on the copy.

Typewriter: A mechanical device that produces print by striking the paper.

Writing style characteristics: Handwriting based upon a particular writing model.

Voiceprint: See Sound spectrograph.

Voice recognition: The application of aspects of voice production and speech that can be used for voice identification and voice verification.
Questions for Further Practice and Mastery

17.1. What is the definition of forgery?
17.2. What gives rise to an individual’s basic style characteristics in their handwriting?
17.3. What are the three basic handwriting styles?
17.4. A counterclockwise circular stroke when writing the letter, O, usually indicates the author is using which hand?
17.5. When examining the content aspect of a document, for what specific features is the Document Analyst looking?
17.6. When comparing a suspected sample of handwriting to a known sample from a given author, the analyst will assign it one of nine categories of certainty. What are those nine different categories?
17.7. What is the significance of the US Supreme court decision in *Gilbert v California, 1967* and *US v Mara*?
17.8. What are non-request exemplars?
17.9. What are some of the problems analyst’s face when collecting requested exemplars of a suspect’s handwriting? What measures are taken to minimize some of these problems?
17.10. What does the term, simulation, mean with respect to handwriting?
17.11. What are comparison tables and how are they constructed?
17.12. What are some of the signs that indicate an exemplar is a simulation?
17.13. When analyzing a signature for authenticity, what is meant by accuracy and fluency?
17.14. What is the difference between a forged signature, a simulated signature and a traced signature?
17.15. What is (a) font (b) a “pica” (c) a ribbon with reference to a typewriter?
17.16. What are the differences between dot matrix, ink jet and laser jet printers?
17.17. What is a possible way to connect a copied document to a particular copy machine?
17.18. What are “trash marks”?
17.19. What are common anti-counterfeiting measures currently in use?
17.20. What might chemical and physical analysis of a questioned document determine about the ink and the paper in that document?
17.21. What methods are used to compare ink samples?
17.22. What is hyperspectral imaging and what is it used for?
17.23. What are some of the common difficulties that forensic scientists face when analyzing surveillance video?
17.24. What is multiplexing?
17.25. What does the acronym, SWGIT, stand for?
17.26. What information must be included with every photograph that is being used as evidence?
17.27. What are specular highlights? How do they help determine if a photograph has been altered?
17.28. What are the two main uses of voice recognition for forensic applications?
17.29. What parts of the body are involved in producing speech?
17.30. What is timbre? What factors contribute to timbre?
17.31. What is a sound spectrograph?
17.32. What three key components of sound are displayed on a sound spectrograph?
17.33. There are several valid concerns over the reliability of voice prints as evidence. What are some of these concerns and what reliability percentage has the FBI assigned to voice print evidence?

Extensive Questions

17.34. Explain the four main categories of individualizing features in handwriting.
17.35. Explain the oblique lighting method of analyzing indented writing as well as the electrostatic detection apparatus method.

17.36. A video of a crime being committed is sent to the lab for analysis. The quality of the video is poor. It is dark with lots of noise, is blurry and has several partial images of the suspect’s get-away vehicle. Explain what the lab can do to make the video much more useful to investigators.