
FORENSIC AUDIO LAB

AUDIO FORENSICS TECHNOLOGY WHITE PAPER

January 2013

SCOPE

Speech Technology Center is the leading manufacturer of products for forensic audio investigations. Its Forensic Audio Workstation has a long history dating back in 1993 when a group of audio experts and software developers joined to create a powerful tool for signal analysis.

INTRODUCTION

In recent years audio recording tools thanks to their reliability, small sizes and simplicity are widely used in everyday life and for security purposes. Needless to say that sometimes audio recording may be the only evidence of a security threat or crime and therefore may become a key element in the case analysis or subsequent court trial.

From a legal point of view forensic audio analysis allows to prove some facts of criminal activity which might take place in private, without witnesses. For this reason, sound recordings are widely used in criminal and civil proceedings.

The expertise in the field of criminology, acoustics, sound equipment, mathematics, linguistics, phonetics and theory of speech production make up the scientific basis of forensic audio. Using tools and techniques developed in various sciences for forensic audio analysis allows experts to solve a wide range of audio analysis challenges. Each challenge is associated with a particular investigative situation arising in the course of the investigation. The most frequently encountered challenges are the following:

SPEECH ENHANCEMENT: Digital and analog processing to restore verbal clarity which makes audiotapes and files more intelligible in a courtroom.

SPEECH DECODING: Methods that can be used to extract human speech from a noisy track and convert it to a reasonably accurate and complete transcript and to a final hard copy.

AUDIO AUTHENTICATION: Aural, electronic and physical examination of an audio evidence to prove that it has not been tampered, altered, or otherwise changed from its original state. Another common challenge is to determine tapes' authenticity by checking whether a particular tape was indeed made on a particular machine.

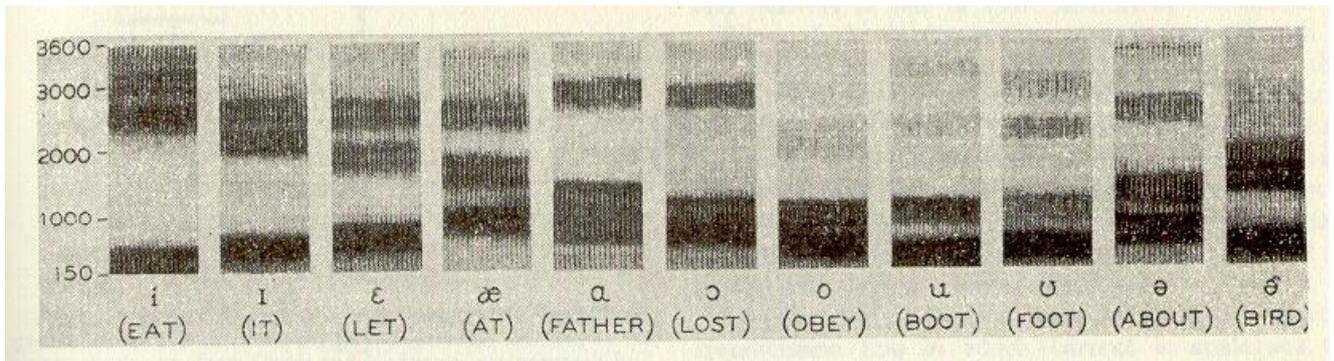
VOICE IDENTIFICATION: Voice ID is the science that attempts to determine whether the recorded voice belongs to the suspect or not. Voice ID is based on the theory that voice of each person is as unique as fingerprints or DNA and depends on the individual features of speech production organs, the shape of vocal tract, mouth cavity, pronunciation skills, regional accent etc.

A LOOK INTO HISTORY

To determine the exact "birth date" of audio forensic science is hardly possible. The traces of the first discussions of the admissibility of "aural-perceptual" (i.e., hearing) testimony go to a few centuries ago in England, where in 1660 the witness identified the defendant by voice.

However, only in the middle of last century this branch of forensic science has evolved. The reasons for this are three factors:

- Law enforcement group inquired about what help they could get in combating telephoned bomb scares to airlines and public buildings. Their particular interest was in being able to identify the voice of the perpetrator of such crimes.
- The studies carried out in the Bell Telephone Laboratories have pointed up the truly remarkable uniqueness of an individual human voice.
- Sound spectrograph which acted as an automatic wave analyzer recording the acoustic patterns of speech in the dimensions of time, frequency, and intensity. The acoustic patterns called voiceprints permitted side-by-side visual comparison of speech sounds, instead of requiring that an investigator listen to the sounds one after another with uncertain dependence on memory.



Visual graph of speech as a function of time (horizontal axis), frequency (vertical axis), and voice energy (gray scale or color differences). The plots received with the first analog spectrograph machine.

(From: *The calculation of vowel resonances, and an electrical vocal tract* by H.K. Dunn 1950, *J. Acoust. Soc. Amer.*, **22**, pp 740-753)

In the USA the first known case when voice spectrograms (voiceprint) were presented in the courtroom as an ID method was recorded in 1966. In the early days of this identification technique there was little research to support the theory that human voices are unique and could be used as a means for identification. There was also no standardization of how identification was reached, or even training or qualifications necessary to perform the analysis. Voice comparisons were made solely on the pattern analysis of a few commonly used words. Due to the newness of the technique there were only a few people in the world who performed voice identification analysis and were capable of explaining it to a court. Gradually the process became known to other scientists who voiced concerns, not as to the validity of the analysis, but as to the lack of substantial research demonstrating the reliability of the technique. They felt that the technique should not be used in the courtroom without more documentation. Thus the battle lines were drawn over the admissibility of voice identification evidence with proponents claiming a valid, reliable identification process and opponents claiming more research must be completed before the process should be used in courtrooms.

Today voice identification analysis has matured into a sophisticated identification technique, using the latest technology science has to offer. The research, which is still continuing today, demonstrates the validity and reliability of the process when performed by a trained and certified examiner using established, standardized procedures. Voice identification experts are found all over the world. No longer limited to the visual comparison of a few words, the comparison of human voices now focuses on every aspect of the words spoken; the words themselves, the way the words flow together, and the pauses between them. Both aural and spectrographic analyses are combined to form the conclusion about the identity of the voices in question.



THE TYPICAL AUDIO ANALYSIS PROCEDURE

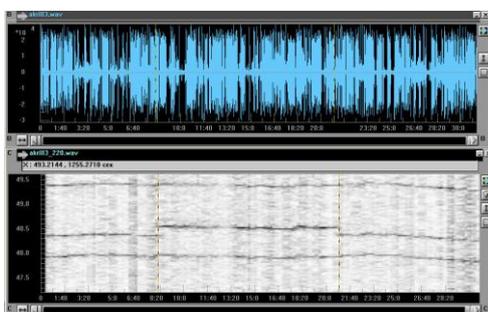
PREPARATION

At this stage an examiner should check the documents relating to procedural and organizational side of the examination, clarify the circumstances of the case and the questions posed to the expert. The investigated evidences should be visually inspected and described in details.

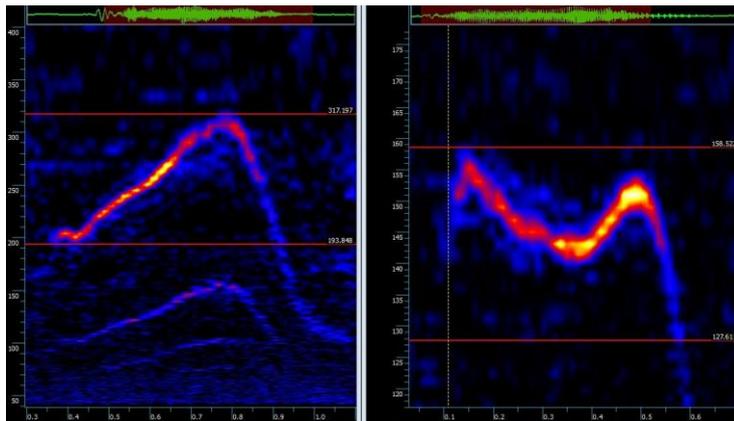
Additional information and materials related to the case should be requested if required.

PRELIMINARY EXAMINATION

The whole sound material received for examination should be listened to. Then an examiner should determine the location of the speech signal in the whole recording. Sound samples and investigated recordings should be assessed in terms of their suitability for forensic identification. The authenticity analysis should be carried out to establish whether a recording is original and whether it has been tampered with. This task is considered to be the most complicated one in audio forensics and requires very specialist skills and equipment. If the question of audio authenticity was not posed to the examiner this type of analysis can be omitted. However, one should remember that artificially created or modified recordings can contain false information about the content of conversations, facts or the participants allegedly fixed in the audio document in the moment of its recording. This kind of recordings cannot be considered as authentic piece of evidence and must be excluded by the court from consideration.



Audio evidence with the traces of editing detected by a forensic audio examiner.



Melodic pattern of the word "Hello!" pronounced by two different speakers.

VOICE IDENTIFICATION

The foundation of voice identification is on the premise that every individual voice is uniquely characteristic enough to distinguish it from all others. The theory of the premise lies in the fundamental processes of human speech. There are two general factors involved. The first factor in determining voice uniqueness lies in the shape of the vocal tract, length and thickness of vocal folds, the sizes of the oral and nasal cavities and other individual voice traits caused by anatomic peculiarities. The second factor of voice uniqueness is a speech production skill which every individual acquires since childhood. Each person has his/her own dictionary of frequently used words, style, grammar patterns, phonetic features which all together make up an individual speech behavior. Thus, the unique combination of physiological and behavioral voice and speech characteristics makes the good potentialities of voice ID.

Voice identification can be started with aural analysis or critical listening. At this stage an examiner assesses and describes the general impression of compared voices: loud, dull, deep, distinct, bright, monotonous, hoarse, staccato, constrained, strong, snuffling, casual, uneducated etc.

Audio forensics is sometimes referred as audio phonetics. This term proves that as far as speech is concerned linguistic analysis of voice and speech should be carried out as one of the phases of ID examination. At this stage the examiner scrutinizes voice and speech of a person as a united system functioning at different levels:

- at the phonemic level - how individual pronounces different vowel and consonant sounds and their conjunctions;
- at the prosodic level - melodic and intonation patterns, rhythmical structure, pauses;
- at the level of vocabulary – words used in speech;
- at the level of syntax and grammar - grammatical structures used for utterances and their correctness.

SpeechPro's SIS II is the most used forensic audio software in the world. Nowadays it is used in more than 350 labs in over 36 countries worldwide.

As above-said, anatomic structure of a human speech apparatus influence the speech it produces. The vocal cavities are resonators, much like organ pipes, reinforcing some of the overtones produced by the vocal folds, and producing spectral peaks or formants. Both research and practice demonstrate that formants correlate directly with anatomic and geometrical sizes and structures of speech apparatus and its live tissues.

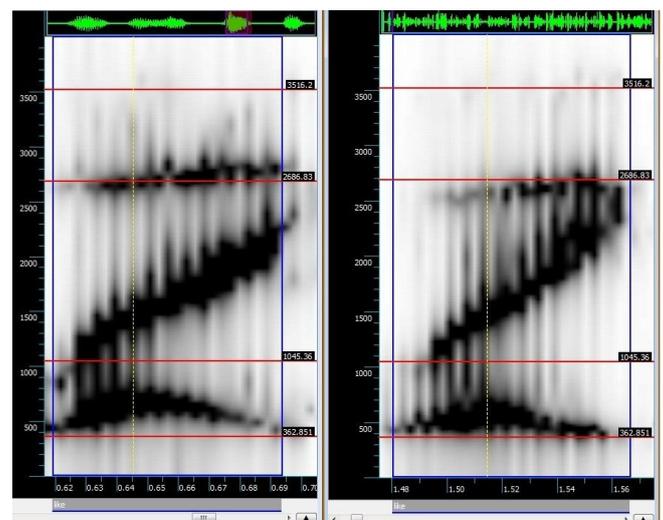
Spectrographic analysis is performed for detailed examination of these resonances. Thus, the third step of voice ID procedure is called spectrographic or instrumental. Nowadays computer-based spectrographs have completely ousted analog

spectrograph machines. The sophisticated software provides high fidelity signal acquisition, high- speed digital signal

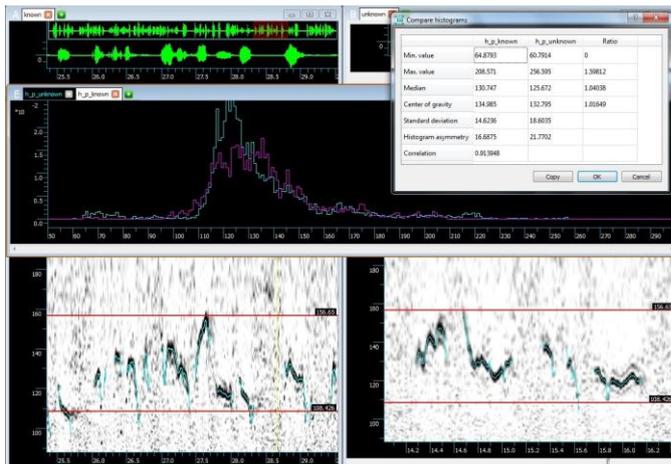
processing for quick and flexible analysis, and CD-quality playback. The computerize-based systems accomplish all the same tasks of the analog systems, but with the computer-based systems the examiner gains a host of comparison and measurement tools not available with the analog equipment. The computer-based systems are capable of displaying multiple sound spectrograms, adjusting the time alignment and frequency ranges and taking detailed numeric measurements of the displayed sounds. With these advances in technology, the examiner widens the scope of the analysis to create a more detailed picture of the voice or sound being analyzed.

Using spectrograms of the recordings of known and unknown speakers an examiner compares visual presentation of similar words, syllables and sounds. Matching of all formants and their curves for similar sounds results in positive identification.

Additionally, the length of the similar stressed vowels, gaps between consonants and vowels, spectrums of consonant sounds can be compared.



Two similar words "like" of know and unknown speakers compared.



Pitch and pitch histograms compared for known and unknown speakers.

A special type of spectrographic signal presentation called cepstrum or cepstrum allows for detail pitch analysis. An examiner compares minimal, medium and maximal pitch values for both samples. When pitch curves are extracted from the signals they can be compared using overlaid histograms. As far as speech is a skill comparing melodic patterns for similar phrases is also a good practice.

Fundamental frequency or pitch also can and must be thoroughly analyzed. Fundamental frequency refers to vibration of vocal folds. Long and thick vocal folds produce less oscillation. Owners of such vocal folds are normally men. And visa versa short and thin tissues yield high women voices. Exactly like guitar strings.

When the analysis is complete the examiner integrates his findings from both the aural and spectrographic analyses into one conclusion.