

Handling the “Voiceprint” Issue

Jonas Lindh

Department of Linguistics

Göteborg University

What was and is “Voiceprint”/Aural-Spectrographic Identification?

Throughout the years, going all the way back to the Second World War, speaker identification by spectrograms has had an influence on Forensic Speaker Identification. The method has been criticized and embraced by different people at different points in time. This paper describes the issues diachronically, using the most important references as a thread. Questions are asked to the phonetic science community concerning their beliefs and their actions regarding the critics of the method and what the method has for a kind of influence on today’s forensic voice identification. It seems as if a common view towards the early stages in the development of “voiceprint” identification, is that not many phoneticians contributed with a view or opinion (Hollien, 1977). However, this paper presents several well-known scientists and their papers regarding the issue at an early stage and throughout the years. It also discusses the proponents’ arguments and credibility.

Chronological Background

Towards the end of the 1930-ies and the beginning of the 1940-ies the sound spectrograph was developed as a means of visualizing the speech signal. During the Second World War the work on the spectrograph was classified as a military project because the military saw the possibility of using the method as a way of identifying enemy spies from intercepted radio communications and telephone exchanges. (Grey and Kopp, 1944)

The focus on speaker identification decreased for a period of time immediately after the war, but found new interest when the New York Police started to receive reports of bomb threats to different airlines. They then turned to Bell Laboratories and asked if spectrograms could be used as a means of identifying the callers. Lawrence Kersta, an engineer at Bell, was assigned the task of investigating the matter. (McDermott et. al., 1996)

After studying the matter for two years,

Kersta had become convinced that spectrograms could indeed be used to identify speakers. In his paper “Voiceprint Identification” (1962) he refers to the method as “voiceprint” in direct analogy with the term “fingerprint”. His paper touches the surface of this visual pattern matching procedure, but concentrates on presenting the results from an experiment that he conducted using high school girls as a demonstration of the simplicity in the procedure. They performed the identification of subjects very successfully (see Table 1).

What happened next?

Several investigations followed, trying to establish how well one can actually identify a speaker using the method. Experiments on disguise were performed with differing results, most of them criticizing Kersta and his method. Bricker and Pruzansky (1966) discovered that context dependence is important to be able to perform identification. They also suggested that an aural perspective might enhance the identification rates.

Young and Campbell (1967) also tested the effects of different contexts and made a summary of the experiments that had been performed up to date.

Table 1. Summary from Young/Campbell (1967)

Exp	Pollack/ Pickett/ Sumbly (1954)	Kersta (1962)	Bricker/ Pruzansky (1966)	Young/ Campbell (1967)
Res (%)	84-92	99-100	81-87/89	78/37,3
Meth	Short words, isolation	Short words isolation and context	Words isolation and context	Short words isolation and context

The differences between studies were substantial. However, the results are difficult to compare due to lack of documentation.

The suggestion from Bricker and Pruzansky to use an aural approach as well was tested the following year (Williams et. al., 1968). It was

discovered that in these tests subjects had an error rate of 6% with an aural approach and 21% errors using visual pattern matching. Probably this was a first indication of including aural examination.

The Question of Admissibility

In the case *United States v. Fry* (1923), an early version of a so called lie detector test was presented as evidence, the court dismissed the evidence saying that for an expert testimony to be admissible the method on which it is based must be "sufficiently established to have gained general acceptance in the particular field in which it belongs". This principle has not been applied in all states and the interpretation of exactly what it means for a method to "have gained general acceptance" has often been disputed but the ruling has nevertheless often been used as a motivation for not accepting voiceprint testimonies. The first attempts at introducing voiceprint testimonies were all dismissed with reference to the Fry ruling. (McDermott et. al., 1996)

Disputes

Opinions quickly became divided after the introduction of the voiceprint technique. Proponents (some scientists, but mostly laymen) defended the technique, regarded it as highly reliable and appeared as expert witnesses in various criminal cases. Most scientists, however, were skeptical regarding it as not sufficiently tested (e.g. Stevens) or dismissed it completely (e.g. Hollien).

Bolt et. al. (1969, 1973) published two papers criticizing the method and presented several relevant questions:

- "When two spectrograms look alike, do the similarities mean "same speaker" or merely "same word spoken"?"
- "Are the irrelevant similarities likely to mislead a lay jury...?"
- "How permanent are voice patterns...?"
- "How distinctive are they for the individual?"
- "Can they be successfully disguised or faked?"

A response concluded more or less "It is our contention that opinions based on feelings other than in actual experience are of little value irrespective of the scientific authority of those who produce such an opinion" (Black et. al., 1973). This paper did not contain any scientific evidence supporting the method though.

In the midst of this heated debate the IAVI (International Association of Voice Identification, later called VIAAS, Voice Identification and Acoustic Analysis Subcommittee) was founded (1971) on initiative from Kersta, presenting guidelines for the practitioners of aural/spectrographic identification. The IAVI stipulated that to become a practitioner, one needs two years apprenticeship supervised by an authorized examiner. Five levels of identification were used: positive identification/elimination, probable identification/elimination and "no opinion". (McDermott et. al., 1996)

The critique continued during this period presenting results showing change as a function of age as well as disguise by imitators (Endres et. al., 1971) and emotional states (Williams and Stevens 1972).

Other researchers, however, published results that seemed to lend support to Kersta's method. A group of researchers led by Tosi at Michigan State University tested Kersta's methods in a very extensive study that produced results which very closely matched Kersta's (Tosi et al., 1972). If they excluded all "uncertain", there were only 2% false identification and 5% false elimination, this supporting the "no opinion" criterion given by IAVI.

Several papers from IAVI members followed supporting the method. Hall (1975) stated, "Variability does not exist". Lt. Smrkovski (1975) published a paper on the importance of experts performing the aural-visual identification.

The proponents' activity had increased as a response to the earlier critique and it was time for a new response.

Hollien and McGlone (1976) tested the method and concluded "...even skilled auditors such as these [faculty] were unable to match correctly the disguised speech to the reference (normal) samples as much as 25% of the time...".

Hollien (1977) also published a "Status report..." explaining the continuous use of the method, by making four claims: (1) "Proponents of voiceprints are rarely opposed...", (2) "They claim their method meets the Frye test", (3) "They claim uniqueness...", (4) "They claim that their research demonstrates reliability...".

The Greatest Proponent

Michigan State University presented several interesting papers regarding the method, all proposing it. The ox pulling the scientific carriage was Oscar Tosi. At an early stage of developing the method he had testified against the use of spectrographic comparison, but in the case "State ex. rel. Trimble v. Heldman" (1971), the Supreme Court held that "spectrograms ought to be admissible at least for the purpose of corroborating opinions as to identification by means of ear alone". Tosi had impressed the court now stating extreme reliability of the technique after testing it (McDermott et. al., 1996).

Later Tosi (1979) published a book where he gives an up to date reference to all areas involved in forensic phonetics. It also criticizes the authors mentioned earlier in this paper opposing aural-spectrographic examination. Even though the book gives a complete picture of speech acoustics and its reflection in spectrographic representations his conclusions and opinions are questionable.

Professor Tosi can easily be appointed as the greatest proponent of the method along with his colleagues at the Michigan State University, where he was also one of the founders of the Forensic Science program. Thanks to him the term "voiceprint" was excluded as a term as he, opposing Kersta, did not propose the method's infallibility.

In 1980 the IAVI was absorbed by the IAI (International Association for Identification).

What happened next, dispute II?

A period of status quo followed until it was revealed in 1986 that the FBI was using the method. By this time it had been used for investigative purposes since the 1950:s.

Koenig (1986) reported error rates for the spectrographic voice identification technique under forensic conditions stating: "The survey revealed that decisions were made in 34,8% of the comparisons with a 0,31% false identification error rate and a 0,53% false elimination error rate." The report/survey by Koenig was rather thin in explaining the figures.

In a reply (Shipp et. al., 1987) relevant questions were asked. First of all, critique was aimed at the methodology. What procedures do they (the FBI) actually use employing the method? The results were based solely on the feedback from verdicts, contradicting them-

selves, since the technique employed might determine guilt or innocence and that verdict then was used to verify the results of the technique.

The extended answer to this reply cleared out some of the blurs in the first paper, but reported no more evidence to why the method was employed at all.

Standards for Voiceprint Examinations

According to the FBI survey, voice identification examiners at the FBI had to have a minimum of two years of experience and completed at least a hundred voice comparison cases. Combined aural visual examination was employed and decisions used were: very similar, very dissimilar, no decision (low confidence) (Koenig, 1986). Further, they had to have "...a minimum of a Bachelor of science Degree in a basic scientific field", completed a two-week course in spectrographic analysis ("...or equivalent") and pass a yearly conducted hearing test. (McDermott et. al., 1996)

The VCS (Voice Comparison Standards) of the VIAAS (Voice Identification and Acoustic Analysis Subcommittee, 1991) were very similar and obviously not independent. The criteria for the VIAAS examiners included a high school diploma instead of a Bachelor, ten-word comparison vs. twenty and they did not demand the recording to be an original. (McDermott et. al., 1996)

The VCS also contain general descriptions on what to look for in the visual comparisons such as "...general formant shaping, and positioning, pitch striations, energy distribution, word length, coupling (how the first and second formant are tied to each other) and a number of other features such as plosives, fricatives and inter formant features". (Rose, 2002)

Up to date

The greatest issue today is perhaps the common opinion expressed by media. "The CIA, FBI and National Security Agency have computers that use special programs to identify voiceprints. The idea is that every voice has a unique pattern like a fingerprint." (CNN website, December 2002, in conjunction with the Bin Laden voice affair) (Rose, 2003)

Aural-spectrographic voice identification evidence can still be found. FBI is using it for investigative purposes and so are the Japanese police. It is still admitted as evidence in US courts and at least one case involved voiceprint evidence in Australia 2002. (Rose, 2002)

Conclusions

The Frye-test concludes basically that new scientific evidence should have gained "general acceptance" in the relevant scientific field. In 1993 the Daubert-case set a new standard of interpretation now accepted by several American courts of law, basically stating "good grounds" in validating an expert's testimony.

Aural/spectrographic/voiceprint identification has several methodological problems that have not been dealt with in the literature. What is it that one is supposed to look for? What signals identification? When are spectrograms similar enough to indicate same speaker? When are they dissimilar enough?

Even if several experiments have shown that spectrograms are not reliable to verify identity, none of the papers conclude how representative they are for a speaker's voice. Can one make a reliable decision using spectrograms? Finally, has the method ever been an accepted one, according to Frye, which might be a proper measure of a method's reliability? That depends, though, on how one is supposed to interpret it.

References

- Black, J. W., W. Lashbrook, et al. (1973). "Reply to "Speaker identification by speech spectrograms: some further observations". " Journal of the Acoustical Society of America 54: 535-537.
- Bolt, R. H., F. S. Cooper, et al. (1969). "Identification of a speaker by speech spectrograms: How do scientists view its reliability for use as legal evidence?" Science(Oct 17): 338-343.
- Bolt, R. H., F. S. Cooper, et al. (1973). "Speaker identification by speech spectrograms: some further observations." Journal of the Acoustical Society of America 54: 531-534.
- 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.
- Grey, G. and G. Kopp (1944). "Voiceprint identification." Bell Telephone Laboratories Report: 1-14.
- Hall, M. C. (1975). Spectrographic Analysis of Interspeaker and Intraspeaker variables of Professional Mimicry, Michigan State University.
- Hollien, H. and R. McGlone (1976). The effect of disguise on voiceprint identification. Car-nahan Crime Countermeasures Conference, Lexington, KY, University of Kentucky Press.
- Hollien, H. (1977). "Status report of "voiceprint" identification in the United States." Occasionally 2: 29-40.
- Kersta, L. G. (1962). "Voiceprint identification." Nature 196: 1253-1257.
- Koenig, B. E. (1986). "Spectrographic voice identification: A forensic survey." Journal of the Acoustical Society of America 79: 2088-2090.
- Koenig, B. E. (1986). "Review article: Spectrographic voice identification." Crime Laboratory Digest 13: 105-118.
- McDermott M, McDermott F and Owen T (1996). Voice Identification, Ltd. Owl Investigations, Inc.
- Pollack, I., Pickett, J. M. et al. (1954). "On the identification of speakers by voice." Journal of the Acoustical Society of America 26: 403-412.
- Rose, P. (2002). Forensic Speaker Identification. New York, Taylor & Francis.
- Rose, P. (2003) The Technical Comparison of Forensic Voice Samples. Chapter 99, *Expert Evidence* (I. Freckelton & H. Selby eds.), Sydney: Thomson Lawbook Co.
- Shipp, T., E. T. Doherty, et al. (1987). "Some fundamental considerations regarding voice identification." Journal of the Acoustical Society of America 82: 687-688.
- Smrkovski, L. (1975). "Collaborative study of speaker identification by the voiceprint method." Journal of AOAC international 58: 453.
- Stevens, K. N., Williams C. E., 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., H. Oyer, et al. (1972). "Experiment on voice identification." Journal of the Acoustical Society of America 51: 2030-2043.
- Tosi, O. (1979). Voice Identification: Theory and Legal Applications. Baltimore, University Park Press.
- Williams, U., & Stevens, K.N. (1972). Emotions and Speech: Some Acoustical Correlates. JASA 52, 1238-1250.
- Young, M. A. and Campbell, R. A (1967). "Effects of context on talker identification." Journal of the Acoustical Society of America 42: 1250-1254.