

Person Authentication by Voice: A Need for Caution

Jean-François Bonastre,^{1,3} Frédéric Bimbot,^{1,4} Louis-Jean Boë,^{1,5}
Joseph P. Campbell,^{2,6*} Douglas A. Reynolds,^{2,6*} Ivan Magrin-Chagnolleau^{2,7}

(1) Association Francophone de la Communication Parlée (AFCP)

jfb@lia.univ-avignon.fr, frederic.bimbot@irisas.fr, boe@icp.inpg.fr

(2) Speaker and Language Characterization (SpLC) SIG

j.campbell@ieee.org, dar@ll.mit.edu, ivan@ieee.org

(3) LIA, Université d'Avignon, BP 1228, 84911 Avignon CEDEX 9, France

(4) IRISA, Pièce A 123, Campus Universitaire de Beaulieu, 35042 Rennes CEDEX, France

(5) ICP, Université Stendhal, BP 25, 38040 Grenoble CEDEX 09, France

(6) MIT Lincoln Laboratory, Lexington, Massachusetts 02420, USA

(7) DDL, CNRS & Université Lyon 2, 14 Avenue Berthelot, 69363 Lyon CEDEX 07, France

Abstract

Because of recent events and as members of the scientific community working in the field of speech processing, we feel compelled to publicize our views concerning the possibility of identifying or authenticating a person from his or her voice. The need for a clear and common message was indeed shown by the diversity of information that has been circulating on this matter in the media and general public over the past year. In a press release initiated by the AFCP and further elaborated in collaboration with the SpLC ISCA-SIG, the two groups herein discuss and present a summary of the current state of scientific knowledge and technological development in the field of speaker recognition, in accessible wording for nonspecialists. Our main conclusion is that, despite the existence of technological solutions to some constrained applications, *at the present time, there is no scientific process that enables one to uniquely characterize a person's voice or to identify with absolute certainty an individual from his or her voice.*

1. Introduction

There has long been a desire to be able to identify a person on the basis of their voice. For many years, judges, lawyers, detectives, and law enforcement agencies have wanted to use forensic voice authentication to investigate a suspect or to confirm a judgment of guilt or innocence [1, 2].

Despite the fact that the scientific basis of person authentication by his/her voice has been questioned by researchers (e.g., by scientists in 1970 [3], British academic phoneticians in 1983 [4], and the French speech communication community from 1990 to today [5]), there is a perception by the general public that it is a straightforward task. As shown in [5], this misunderstanding partially began in 1962 in an article by Kersta appearing in Nature [6]. This paper introduced the misleading term – “Voiceprint identification”, which is still in vogue in daily newspapers,

televised police dramas, and spy films. This term, *voiceprint*, leads many people to believe that a graphical representation of the voice, via a spectrogram, is just as reliable as the structure of the ridges and minutiae of the fingertips or genetic fingerprints (e.g., DNA) and that it allows reliable identification of the original speaker.

With the developments in automatic speaker recognition over the last decade (e.g., [7, 8]), there is increased need to distinguish between its appropriate and inappropriate uses in various forensic voice authentication contexts and to differentiate between common versus forensic speaker recognition applications. This need was highlighted during recent world events. This paper is intended to offer clear information to journalists, as well as the general public, concerning the possibility of identifying or authenticating a person by his or her voice.

Specialists in the field of speech science and technology from the French-speaking scientific community (*Association Francophone de la Communication Parlée*¹ - AFCP) have in recent years attempted to highlight to public and legal bodies, to the general public and to the media, the limitations of the techniques used for the identification of individuals based on their voice characteristics. Their position has been outlined in a number of official statements [9, 10], scientific publications [5, 11], and several legal proceedings. The AFCP in collaboration with the *Speaker and Language Characterization* (SpLC) ISCA-SIG² present a summary of the current state of scientific knowledge and technological development in the field of speaker recognition, in accessible wording for nonspecialists. This common position is presented in Section 2 of this paper and joint conclusions are presented in Section 3.

2. Voice identification: a set of processes not all supported by a scientific approach

The ability of humans to identify speakers from their voice for forensic applications is of great legal interest [1, 2]. Given the current state of knowledge, there are no methods, either

* These authors are sponsored by the United States Government's Technical Support Working Group and the Federal Bureau of Investigation under Air Force Contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

¹ The AFCP was initiated by the Groupe Francophone de la Communication Parlée (GFCP) of the Société Française d'Acoustique (SFA) in November 2001: <http://www.afcp-parole.org/>.

² The SpLC is a special interest group of the International Speech Communication Association (ISCA): <http://www.spic-isca.org/>.

automatic or based on human expertise, that enable one to state with certainty that a person is (or is not) the speaker in a particular recording [9, 12, 13, 14, 15, 16]. This is particularly true when one is trying to authenticate a short utterance with strong background noise, recorded with poor quality equipment over unknown channels by a speaker who may have disguised or artificially modified his or her voice [10, 17]. Internationally recognized studies conducted with a high level of scientific rigor, such as those appearing in scientific publications, support this statement [18, 19].

2.1. Aural recognition

As humans, we are able to listen to and recognize speakers based on their voice alone with varying degrees of success. This natural ability is the basis of aural (auditory) speaker recognition. People, however, have varying abilities to recognize talkers [20, 21] and various factors influence the reliability of this method. Some factors that can increase the reliability of aural recognition are familiarity with the speaker [22, 23, 24], duration of the sample [25], context [26], contemporaneous samples [27, 28], lack of vocal stress and disguise [29, 30], and training [31]. A voice examiner typically performs an aural analysis in addition to other types of speaker recognition, which we will now discuss.

2.2. Spectrogram recognition

A *voiceprint*, which is properly referred to as a spectrogram [3, 32], is frequently misused in common everyday language and in works of fiction. A *voiceprint* is simply a spectrogram of the voice signal that can be printed. It displays the signal in three dimensions of time vs. frequency vs. intensity. Spectrograms are useful engineering and voice analysis tools; however, their print connotation has nothing at all to do with fingerprints.

- The term *voiceprint* gives the false impression that voice has characteristics that are as unique and reliable as fingerprints or genetic imprints. This is absolutely not the case. Currently, scientific research has not reached a stage in which it is possible to state that voice characteristics permit unique identification of individuals.
- Voice differs from fingerprints and genetic imprints in several major aspects:
 - Voice changes over time, either in the short-term (at different times of the day), the medium term (times of the year), or in the long-term (with age). Voice is also affected by the speaker's health or emotional state.
 - Long-term noncontemporary samples represent a challenge.
 - Voice can be altered voluntarily (e.g., impersonators) and can be easily disguised using existing technology.

Fingerprint examination has the benefits of long history and large databases relative to voice evaluation databases/corpora, which do not have sufficient numbers of speakers, languages, and recording conditions to measure speaker recognition accuracy at the level of confidence desired for high-reliability forensic identification.

In addition, there are documented peculiar legal cases involving spectrograms, e.g., [33].

Furthermore, the term *voiceprint* is sometimes used to refer to the entire field of manual and automatic speaker recognition, regardless of whether spectrograms are used or not.

2.2.1. IAI standards for aural and spectral comparisons

The Voice Identification and Acoustic Analysis Subcommittee of the International Association for Identification established standards for the comparison of recorded voice samples [34]. These standards aim to provide reliable and uniform spectrographic voice comparisons. Using the IAI method, the known and unknown samples must contain spoken words that are comparable. Following IAI standards, an examination can only produce one of seven decisions: (1) Identification, (2) Probable Identification, (3) Possible Identification, (4) Inconclusive, (5) Possible Elimination, (6) Probable Elimination, or (7) Elimination. The IAI defines these decisions based on aural and spectral (spectrographic) comparisons of comparable words in the samples that match or do not match. As one might expect, extreme decisions (1 and 7) are rare. There are cases where the IAI standards cannot be met and, choosing to err conservatively, no decision can be made [16].

2.2.2. Daubert Factors: spectrogram admissibility?

In their 1993 *Daubert v. Merrell Dow Pharmaceuticals* decision (509 U.S. 579), the U.S. Supreme Court ruled that five conditions should be met for evidence to be admissible as *scientific* in a court of law [35]:

- The theory or technique has been or can be tested.
- The theory or technique has been subjected to peer review and publication.
- The existence and maintenance of standards controlling use of the technique.
- General acceptance of the technique in the scientific community.
- A known or potential rate of error (that is acceptable).

With respect to the Daubert Factors, here are some questions to be answered regarding the spectrographic method:

- Is spectrographic evidence to be considered scientific and is the method an exact science?
- How is objectivity established during the comparison processes?
- Can other researchers and examiners replicate the results?
- What is the known potential rate of error?
- What are the components of error due to technique and due to analysis?
- Which scientific community accepts the spectrographic technique?

2.3. Forensic phonetics speaker recognition

The forensic phonetics method of recognizing speakers uses a linguistic approach [4, 16, 36, 37]. Information gleaned from systematic study of the sounds of language is used by a phonetic expert to give evidence as to the likelihood that a recorded voice sample was produced by a particular person or, more correctly, to estimate how many times more likely it

is to observe the differences between the samples assuming that they have come from the same, rather than different, speakers (a so-called Bayesian likelihood ratio [16]). This is in contrast to aural identification performed by lay listeners (e.g., in a voice lineup). Forensic phonetics actually predates the spectrographic method and was the basis for the first widely published study on comparing voices [38], which stems from the very famous court case of the Lindberg baby kidnapping. Some limitations of forensic phonetics include the limited availability of qualified phoneticians who are native in the language of the samples, complications arising from samples in different languages, inadequate reference data, and insufficient resources to determine how typical a voice is in general.

2.4. Automatic speaker recognition

Voice contains information that partially characterizes a particular speaker. The scientific field that uses this information to recognize a speaker by machine is called *automatic speaker recognition* [39]. The most common applications are access protection to physical premises or securing remote services (especially telephone-based services).

State-of-the-art automatic speaker recognition techniques rely on similarity measures across a set of recordings. These measures are based on acoustic parameters extracted with signal analysis techniques. They can take into account the statistical distributions for a particular speaker, the content of the message, and information about the environment and recording medium.

To provide a reasonable level of performance in speaker recognition applications, the following prerequisites are usually required:

- Speakers must not try to disguise their voice.
- The recording conditions and signal processing techniques are known or controlled.
- Speech, recorded in similar conditions to those in which the test signal is recorded, is available to register a speaker in the system.
- Reference values for similarity measures must have been established in similar conditions to those in which the test signal is recorded. Decision thresholds must have been calibrated from these reference values and tuned as a function of a specific application.

Applying additional constraints can result in improved performance:

- Speakers must be willing to be recognized and cooperate with the system.
- Potential impersonators must be prevented from using sophisticated technology to modify or disguise their voice.
- The use of speech synthesis devices is not allowed.
- The linguistic content of the message includes words already known to the system, so that the similarity between different voices can be calculated on the basis of similar contents.

Many of these constraints can be somewhat enforced by inherent ergonomics of the system. As the field progresses, fewer of these constraints might be necessary to provide satisfactory performance for various applications.

2.4.1. Judicious uses

Despite the preceding limitations and cautions on the use of automatic speaker recognition technology, it should be noted that these techniques are under continual and vigorous research, development, and evaluation [19]. Current research is focused on the practical limitations of automatic speaker recognition (e.g., [7, 8]) and on the presentation, reporting, and interpretation of the results of automatic speaker recognition systems [16, 18].

Progress is being made, so judicious uses may become more reliable. Provided they have been carefully evaluated beforehand, automatic systems can be useful in augmenting other methods to aid in directing investigative efforts when critical voice evidence is available [15]. However, the influencing factors listed in Section 2.4 must be kept in mind, as they limit the interpretation of the output of automatic systems.

3. Conclusions

Currently, it is not possible to completely determine whether the similarity between two recordings is due to the speaker or to other factors, especially when: (a) the speaker does not cooperate, (b) there is no control over recording equipment, (c) recording conditions are not known, (d) one does not know whether the voice was disguised and, to a lesser extent, (e) the linguistic content of the message is not controlled. Caution and judgment must be exercised when applying speaker recognition techniques, whether human or automatic, to account for these uncontrolled factors. Under more constrained or calibrated situations, or as an aid for investigative purposes, judicious application of these techniques may be suitable, provided they are not considered as infallible.

At the present time, there is no scientific process that enables one to uniquely characterize a person's voice or to identify with absolute certainty an individual from his or her voice.

4. Acknowledgements

Thank you to V. Hazan for her assistance with the manuscript and to her and R. Sock for the AFCP press release translation. We are grateful to R. Schwartz, H. Nakasone, J. Wayman, P. Rose, H. Fraser, and P. Higgins for helpful consultations.

5. References

- [1] Bolt, R. H., Cooper, F. S., Green, D. M., Hamlet, S. L., McKnight, J. G., Pickett, J. M., Tosi, O., Underwood, B. D., Hogan, D. L., (1979) *On the Theory and Practice of Voice Identification*, National Research Council, National Academy of Sciences: Washington, D.C.
- [2] Tosi, O. (1979) *Voice Identification: Theory and Legal Applications*. University Park Press: Baltimore, Maryland.
- [3] Bolt, R. H., Cooper, F. S., David, E. E. Jr., Denes, P. B., Pickett, J. M., Stevens, K. N. (1970) "Speaker Identification by Speech Spectrograms: A Scientists' View of its Reliability for Legal Purposes." *Journal of the Acoustical Society of America* 47, 2 (2), 597-612.
- [4] Nolan, J. F. (1983) *The Phonetic Bases of Speaker Recognition*, Cambridge University Press: Cambridge.

- [5] Boë, L. J. (2000) "Forensic voice identification in France," *Speech Communication*, Elsevier, Volume 31, Issues 2-3, June 2000, pp. 205-224 ([http://dx.doi.org/10.1016/S0167-6393\(99\)00079-5](http://dx.doi.org/10.1016/S0167-6393(99)00079-5)).
- [6] Kersta, L. G. (1962) "Voiceprint Identification." *Nature* 196, pp. 1253-1257.
- [7] Reynolds, D. A., Andrews, W. D., Campbell, J. P., Navrátil, J., Peskin, B., Adami, A., Jin, Q., Klusáček, D., Abramson, J. S., Mihaescu, R., Godfrey, J. J., Jones, D. A., Xiang, B. (2003) "The SuperSID Project: Exploiting High-level Information for High-accuracy Speaker Recognition," *Proc. International Conference on Acoustics, Speech, and Signal Processing*, IEEE, Hong Kong, pp. 784-787.
- [8] Reynolds, D. A., (2002) "An Overview of Automatic Speaker Recognition Technology," *Proc. International Conference on Acoustics, Speech, and Signal Processing*, IEEE, Orlando, Florida, pp. 300-304.
- [9] Pétition pour l'arrêt des expertises vocales, tant qu'elles n'auront pas été validées scientifiquement. Pétition du GFCP de la SFA, 1999 (<http://www.afcp-parole.org/doc/petition.pdf>).
- [10] Motion adoptée à l'unanimité par le Bureau du GCP (Groupe de la Communication Parlée) de la SFA, reconduite intégralement par le GFCP de la SFA en 1997 et par l'AFCP en 2002 (http://www.afcp-parole.org/doc/MOTION_1990.pdf).
- [11] Boë, L. J., Bimbot, F., Bonastre, J. F., Dupont, P. (1999) "De l'évaluation des systèmes de vérification du locuteur à la mise en cause des expertises vocales en identification juridique," *Langues*, Vol. 2, n°4 Décembre, pp 270-288 (<http://www.afcp-parole.org/doc/Article-Langue.pdf>).
- [12] Bolt, R., Cooper, F., David, E., Jr., Denes, P., Pickett, J., Stevens, K. (1973) "Speaker Identification by Speech Spectrograms," *J. Acoust. Soc. Am.* 54, pp. 531-537.
- [13] Doddington, G. (1985) "Speaker Recognition: Identifying People by their Voices," *Proc. IEEE* 73, pp. 1651-1664.
- [14] Braun, A., Künzel, H. J. (1988) "Is forensic speaker identification unethical? — or can it be unethical not to do it?," *Forensic Linguistics*, Vol 5, No. 1, pp. 10-21.
- [15] Nakasone, H., Beck, S. (2001) "Forensic Automatic Speaker Recognition," *Proc. 2001: A Speaker Odyssey, The Speaker Recognition Workshop*, ISCA, Chania, Crete, Greece, 18-22 June 2001.
- [16] Rose, P. (2002) *Forensic Speaker Identification*, Taylor & Francis: London and New York.
- [17] Reich, A., Moll, K., Curtis, J. (1976) "Effects of Selected Vocal Disguises upon Spectrographic Speaker Identification," *J. Acoust. Soc. Am* 60, pp. 919-925.
- [18] Champod, C., Meuwly, D. (2000) "The Inference of Identity in Forensic Speaker Identification," *Speech Communication*, Elsevier, Volume 31, pp. 193-203 (<http://www.unil.ch/ipsc/pdf/science.pdf>).
- [19] Martin, A., Przyboc, M. (2000) "The NIST 1999 Speaker Recognition Evaluation — An Overview," *Digital Signal Processing*, v. 10, n. 1-3. January/April/July, pp. 1-18 (<http://dx.doi.org/10.1006/dspr.1999.0355>).
- [20] Ladefoged, P., Ladefoged, J. (1980) "The Ability of Listeners to Identify Voices," *UCLA Working Papers in Phonetics* 49, pp. 43-51.
- [21] Schmidt-Nielsen A., Crystal, T. H. (2000) "Speaker Verification by Human Listeners: Experiments Comparing Human and Machine Performance Using the NIST 1998 Speaker Evaluation Data," *Digital Signal Processing* vol. 10, no. 1-3. January/April/July, pp. 249-266 (<http://dx.doi.org/10.1006/dspr.1999.0356>).
- [22] Van Lancker, D., Kreiman, J., Emmorey, K. (1985) "Familiar voice recognition: Patterns and parameters—Recognition of backward voices," *J. Phonetics* 13, pp. 19-38.
- [23] Papcun, G., Kreiman, J. Davis, A. (1989) "Long-term memory for unfamiliar voices," *J. Acoust. Soc. Am.* 85, pp. 913-925.
- [24] Yarmey, A. D., Yarmey, A. L., Yarmey, M. J., Parliament, L. (2001) "Commonsense Beliefs and the Identification of Familiar Voices," *Appl. Cognit. Psychol.* 15, pp. 283-299.
- [25] Compton, A. (1963) "Effects of Filtering and Vocal Duration upon the Identification of Speakers, Aurally," *J. Acoust. Soc. Am.* 35, pp. 1748-1752.
- [26] Young, M., Campbell, R. (1967) "Effects of Context on Talker Identification," *J. Acoust. Soc.* 42(6), pp. 1250-1254.
- [27] Hollien, H., Schwartz, R. (2001) "Speaker Identification Utilizing Noncontemporary Speech," *J. Forensic Sci.*, 46, pp. 63-67.
- [28] Saslove, H., Yarmey, A. D. (1980) "Long Term Auditory Memory: Speaker Identification," *J. Applied Psychol.* 65, pp. 111-116.
- [29] Hollien, H., Majewski, W., Doherty, E. T. (1982) "Perceptual identification of voices under normal, stress and disguise speaking conditions," *J. Phonetics* 10, pp. 139-148.
- [30] Reich, A. R., Duke, J. E. (1979) "Effects of Selective Vocal Disguise Upon Speaker Identification by Listening," *J. Acoust. Soc. Am.* 66, pp. 1023-1028.
- [31] Schiller, N. O., Koster, O. (1998) "The Ability of Expert Witnesses to Identify Voices: A Comparison Between Trained and Untrained Listeners," *Forensic Linguistics*, 5, pp. 1-9.
- [32] Stevens, K. N., Williams, C. E., Carbonell, J. R., Woods, B. (1968) "Speaker Authentication and Identification: A Comparison of Spectrographic and Auditory Presentations of Speech Material," *J. Acoust. Soc. Am.*, 44(6), pp. 1596-1607.
- [33] Hollien, H. (1974) "Peculiar Case of Voiceprints," *J. Acoust. Soc. Am.* 56, pp. 210-213.
- [34] Voice Identification and Acoustic Analysis Subcommittee of the International Association for Identification (1991) "Voice Comparison Standards," *Journal of Forensic Identification*, Vol 41, No. 5, pp. 373-392.
- [35] *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, 113 S. Ct. 2786, 125 L. Ed. 2d 469 (1993).
- [36] Bricker, P., Pruzansky, S. (1966) "Effects of Stimulus Content and Duration on Talker Identification," *J. Acoust. Soc. Am.* 40, pp. 1441-1450.
- [37] Pollack, I., Pickett, J. M., Sumbly, W. H. (1954) "On the Identification of Speakers by Voice," *J. Acoust. Soc. Am.* 26(3), pp. 403-412.
- [38] McGehee, F. (1937) "The Reliability of the Identification of the Human Voice," *J. Gen. Psychol.*, 17, pp. 249-271.
- [39] Campbell, J. P. (1997) "Speaker Recognition: A Tutorial," *Proceedings of the IEEE*, 85, pp. 1437-1462.