OSAC PROPOSED STANDARD 2023-N-0023
Standard Guide to the Forensic Speaker Recognition Landscape

Speaker Recognition Subcommittee
Digital/Multimedia Scientific Area Committee (SAC)
Organization of Scientific Area Committees (OSAC) for Forensic Science



OSAC Proposed Standard

OSAC 2023-N-0023
Standard Guide to the Forensic Speaker Recognition Landscape

Prepared by
Speaker Recognition Subcommittee
Version: 2.0
December 2024

Disclaimer:

This OSAC Proposed Standard was developed by the Organization of Scientific Area Committees (OSAC) for Forensic Science. It is a completed work product of the OSAC and has been approved for listing on the OSAC Registry. This OSAC Proposed Standard will be submitted to a standard developing organization and is subject to change.

This OSAC Proposed Standard has completed an open comment period and a rigorous technical and quality review by the originating OSAC unit (e.g., subcommittee), their parent body (e.g., scientific area committee), and the OSAC Standards Review Panel.

OSAC encourages the forensic science community to implement this OSAC Proposed Standard into practice as it represents the best guidance available as approved by the OSAC.

There may be references in this OSAC Proposed Standard to other publications under development by OSAC. This OSAC Proposed Standard may be used by the forensic-science community before the completion of such companion publications.

Any identification of commercial equipment, instruments, or materials in the OSAC Proposed Standard is not a recommendation or endorsement by the U.S. Government and does not imply that the equipment, instruments, or materials are necessarily the best available for the purpose.

Visit OSAC's website for more information about the Registry approval process.

Standard Guide to the Forensic Speaker Recognition Landscape

1. Scope

1.1 Forensic speaker recognition, also referred to or covered by the terms forensic speech science, forensic phonetics, and speaker identification, aims to determine whether speakers are likely to be the same person or different people from at least two recordings (e.g., known and questioned recordings). This document provides a landscape of the methods used for analysis in the field of speaker recognition as well as the commonly used interpretation (or opinion) frameworks that is referred to as, Conclusion Frameworks, within the field of speaker recognition. This document also establishes that the wider speaker recognition community has rejected previously held beliefs regarding the scientific validity of voiceprinting. This document is intended to serve as a general overview and reference guide (as it is currently practiced in the field) for forensic speaker recognition.

2. Referenced Documents

2.1 *OSAC Standards:*

2.2 Organization of Scientific Area Committees (OSAC) for Forensic Science Speaker Recognition Subcommittee, "Essential scientific literature for human-supervised automatic approaches to forensic speaker recognition."[1]

[1] Prepared by Scientific Literature Working Group, Forensic Speaker Recognition Subcommittee, "Essential scientific literature for human-supervised automatic approaches to forensic speaker recognition," Organization of Scientific Area Committees (OSAC), Online, Available: https://www.nist.gov/document/essentialscientificliteratureforhuman

3. Significance and Use

3.1 *Definition of Terms Specific to This Standard*

3.2 *Automatic Speaker Recognition (ASR), n.* – as used in this guideline, ASR requires the use of specialized software to compare speech samples, producing a numerical score that is evaluated from the perspective of the same-speaker origin as well as the different-speaker origin.

4. Summary of Practice

4.1 The main objective of forensic speaker recognition is to provide an expert opinion to inform a legal process in determining whether speakers are likely to be the same or different. Forensic science practitioners[2], referred to as "practitioners" in this document, are typically presented with a minimum of two audio recordings and asked to carry out an analysis of those audio recordings. The methods used for analysis in forensic speaker recognition have evolved far past previous traditions of voiceprinting, which has been rejected and discredited by the speaker recognition community. Methods of analysis that are currently in practice include: auditory phonetic analysis, acoustic phonetic analysis, semi-automatic acoustic analysis, automatic speaker recognition, human-assisted speaker recognition, and combined human and automatic speaker recognition analysis. Indeed, just as there are multiple methods for analysis being implemented in speaker recognition, there are also a number of different conclusion frameworks that have also been adopted. Interpretation frameworks (or Conclusion Frameworks) that are currently being utilized by the speaker recognition community include: the binary decision,

[2]ASTM International, 100 Barr Harbor Drive, West Conshohocken, PA 19428, 2024

probability scales, likelihood ratios (both verbal and numerical), the UK Position Statement (**10**)[3], and support statements.

5. Significance and Use

5.1 *Introduction:* Forensic speaker recognition involves the comparison of at least two speech samples (typically from a questioned and known recording) to determine whether speakers are likely to be the same or different. It is common for forensic speaker recognition to be referred to by a few other terms, largely dictated by the field in which the subject is researched and practiced. Within the forensic speech science and forensic phonetics communities, the task is often referred to as forensic speaker comparison or forensic voice comparison. Within the engineering communities, forensic speaker recognition is also sometimes referred to as forensic speaker identification. For clarification purposes, the task of comparing speech samples is referred to in this document as forensic speaker recognition. This document is intended to provide a general overview of forensic speaker recognition. For more detailed information about the human-supervised automatic approaches discussed in this document, please see the OSAC "Essential scientific literature for human-supervised automatic approaches to forensic speaker recognition[4]."

5.2 The objective of the forensic practitioner carrying out forensic speaker recognition is to provide the trier(s) of fact with an informed opinion regarding the probability of obtaining the

[3] The boldface numbers in parentheses refer to the list of references at the end of this standard.
[4] Prepared by Scientific Literature Working Group, Forensic Speaker Recognition Subcommittee, "Essential scientific literature for human-supervised automatic approaches to forensic speaker recognition," Organization of Scientific Area Committees (OSAC), Online, Available: https://www.nist.gov/document/essentialscientificliteratureforhuman

evidence (under the hypothesis that the samples came from the same person, versus under the hypothesis that two different speakers produced each sample). This objective can be reached by practitioners using a variety of methods (i.e., auditory phonetic and acoustic phonetic analysis, semi-automatic acoustic analysis, automatic speaker recognition, human-assisted speaker recognition, or combined human and automatic speaker recognition analysis). While questioned recordings often involve an array of different sounds and speech, the task of forensic speaker recognition is wholly concerned with the speech (and sounds) produced by individuals. Those sounds that cannot be attributed to a person are outside the scope of forensic speaker recognition and fall more into the general area of audio or acoustic forensics. The aim of this document is not to promote or suggest any one method of analysis or interpretation framework over another, but rather to provide a general landscape of the methods used within the speaker recognition community.

5.3 *Voiceprinting:* While there are many ways to conduct forensic speaker recognition, it is important to note here that the method known as "voiceprinting" is not supported by the scientific community and has been discredited. The term "voiceprint"[5] was coined by the author of an article (**28**) which appeared over a half-century ago. The name chosen for that methodology quite transparently implied parallels between a (never-proven) "theory of invariant speech" and the relative invariance of fingerprints.

5.3.1 The so-called “voiceprints” were the product of sound spectrography, a technology carried forward even to the present day, which is still of great utility to speech scientists. The

[5] Gray and Kopp (1944) also used the term voiceprint with the same definition, however, they used the term with a space between the words voice and print. For all intents and purposes this document uses the term voiceprint without a space.

most notable scientific failing of the voiceprint method was that it did not provide examiners with the vocal output (i.e., the audio). This inevitably obscured the phonetic nature of the patterns of acoustic energy and reduced the analysis to a simple pattern-matching exercise. Nevertheless, that exercise was initially heralded with outsized claims of success. The article that introduced the voiceprinting method in 1962 reported that phonetically naïve examiners were able to identify a target voice with 99% accuracy[6], even from a pool of a dozen speakers (**28**). Not surprisingly, this new methodology soon caught the attention of law enforcement and was presented as evidence in a number of criminal prosecutions, in the US and elsewhere.

5.3.2 However, the scientific community remained skeptical. Well-known phoneticians such as Peter Ladefoged and Harry Hollien reported that mere pattern matching (which is all that the young voiceprint examiners were asked to do) was incapable of yielding the astonishing results reported in the 1962 article. Due to the variability present in speech productions from sample to sample, spectrographic template matching is not effective, and it is inconsistent in speaker recognition work. In time, phoneticians began to provide expert testimony against the admissibility of voiceprint evidence, and in consequence, a number of lower-court convictions were eventually overturned. In response to these criticisms, another academic linguist, Oscar Tosi, initiated a more rigorous, and procedurally transparent, voiceprint study (**45**). This yielded less vertiginous, but more scientifically reliable results – 6% false identification errors and 13% false elimination errors, under laboratory conditions. Still, given the high stakes of introducing a still largely unsupported procedure into courts of law, a report issued by the National Research

[6] To drive home his point, Kersta used high school students, who had been given only one week of training, as his examiners. The difficulty of the task was augmented by a forced-choice design; "not sure" was not an option.

Council concluded that the voiceprint method lacked an adequate scientific basis for estimating reliability in many practical situations, pointing out in addition that laboratory evaluations of the voiceprint method showed increasing errors as the conditions for evaluation moved toward real-life situations, such as poor signal-to-noise ratios and dissimilar recording conditions (**36**).

5.3.3 The Federal Rules of Evidence, adopted in 1975, further challenged the voiceprint methodology by shifting the standards for admissibility in favor of practitioners whose "scientific, technical, and other specialized knowledge" can help the trier of fact "to understand the evidence or to determine a fact in issue."[7] Phonetically untrained voiceprint examiners, who sought to identify speakers simply by looking at pictures of their voice signals, were left at a marked disadvantage.

5.3.4 For further historical overviews of the voiceprint approach one can consult (this is not intended to be an exhaustive list, but merely a few selected references):

5.3.4.1 Hollien (1990) (18)

5.3.4.2 Gruber and Poza (1995) (16)

5.3.4.3 Hollien (2002) (19)

5.3.4.4 Rose (2002) (40)

5.4 While voiceprinting has been discredited from the speaker recognition community, it has also been declared inadmissible in court (for further information see U.S. v. Angleton (**46**)).

5.5 Additional methods that are properly applied and under certain circumstances are recognized as appropriate by the community are, in turn, described in detail below. It will become

[7] Federal Rules of Evidence 702. https://www.govinfo.gov/content/pkg/CDOC-118hdoc33/pdf/CDOC-118hdoc33.pdf.

apparent that these methods have historically developed in parallel to one another, as they have grown out of different disciplines. However, it is not uncommon to see some cross-over between the various methods used in forensic speaker recognition. This will be explained further in the sections that follow.

6. Procedure

6.1 There is no one, single method that is used by all practitioners of forensic speaker recognition, and it is sometimes the case that some of these methods are combined when undertaking analysis. The methods most commonly employed in forensic speaker recognition are: auditory phonetic analysis, acoustic phonetic analysis, auditory phonetic + acoustic phonetic analysis, semi-automatic acoustic speaker recognition, automatic speaker recognition, human-assisted automatic speaker recognition, and a combination of auditory phonetic + acoustic phonetic analysis and (human-assisted) automatic speaker recognition. All seven approaches to speaker recognition are detailed below.

6.2 *Auditory Phonetic Analysis (AuPA):* AuPA is defined as the process by which "the expert listens analytically to the speech samples and attends to aspects of speech at the segmental and suprasegmental levels" (**12**). AuPA is very important in the identification of language varieties, such as regional accent, foreign accent or in the detection of linguistic correlates of various social factors. Age, sex, and gender also fall into the category of characteristics most commonly judged auditorily. All of these can be classified as "group-level characteristics," in contrast to "individual-level characteristics" (**22**). Group-level speaker characteristics are crucial in speaker profiling, but they also have their established place in speaker recognition. They can be important in defining

the relevant population. They are also particularly powerful as evidence speaking against speaker identity: if, for example, the known and the unknown voices use two different regional varieties, it is likely that they are different individuals, given that bi-dialectalism is relatively rare. Group-level characteristics can also provide important information that supports inclusion, within the context of the case. They are also particularly helpful in excluding speakers. Part of the tradition of AuPA has been to narrow down dialect to a point that it achieves the status of a rare combination of linguistic parameters only used by a few individuals (and ideally, however, rather unrealistic, just one individual). This can occur when a speaker uses only some features of a dialect (or other language variety) to the exclusion of others, or if features from various language varieties are combined. Discussions of these aspects related to language variety are provided in Jessen (**24**; **25**) and Hughes & Rhodes (**22**).

6.2.1 AuPA is also used for the description and interpretation of various individual-level speaker characteristics. According to the survey by Gold and French (**12**), voice quality is a particularly important one. Voice quality can be measured acoustically (**27**), but given the acoustic limitations typically encountered in forensic casework, most of these methods suffer from information loss or lack of applicability. Auditory analysis, instead, offers more robustness, though auditory analysis is also more subjective. Auditory voice quality assessment in forensics often builds upon the classificatory framework of the phonetician John Laver (**29**), particularly in Europe. A description of that framework and how it is adapted to forensics is found in San Segundo et al. (**42**), and a complete definition of voice quality can be found in McIntyre et al. (**32**). Another classification framework for AuPA-based analysis has been developed for disfluency patterns, which include silent pauses, breathing pauses, filled pauses (utterance such

as uh, and um) or sentence interruptions (McDougall and Duckworth (**31**), de Boer and Heeren (**3**), Hughes et al. (**20**)). Further speaker characteristics observed auditorily are listed in Gold and French (**12**).

6.3 *Acoustic Phonetic Analysis (AcPA):* AcPA is the method by which "the expert analyzes and quantifies physical parameters of the speech signal using computer software. As with AuPA, this is labor intensive, involving a high degree of human input and judgment" (**12**). AcPA traditionally has its strongest focus on speaker characteristics that have an anatomical motivation and that have been known since the 1950s to vary between women, men, and children, but also between individuals within these larger speaker categories. This applies to fundamental frequency and formant frequencies.

6.3.1 Fundamental frequency (f0), which is the frequency of the vibration patterns of the vocal folds, depends on the size of the larynx (especially vocal fold length), but it is to a degree controllable for linguistic purposes. As a way of disregarding locally determined linguistic factors, a common method is to average f0 (mean, mode, or median) across long utterances or the entire recording (**21**). In this process, further irrelevant factors that have a strong influence on f0 must be controlled as much as possible; this is particularly important for the f0-raising effect of vocal effort (that is, speaking loudly) (**23**). Speakers can also differ habitually in terms of how "melodically" they speak (scale from speaking monotonously to highly modulated). Standard deviation of f0 across long passages or the entire recording is a way of capturing these habitual speaker differences. Since mean and standard deviation of f0 are to some extent correlated, variability is sometimes expressed by the coefficient of variation (standard deviation divided by mean), by means of which the correlation almost disappears (**23**).

6.3.2 Vowel formant frequencies, which are characteristic patterns of amplitude peaks in the speech spectrum, are associated with the length of the speaker's vocal tract and other anatomical features. However, formant frequencies – especially the first formant (F1) and the second formant (F2) – are also crucial carriers of linguistic information; they are the main correlates of vowel distinctions in a language, and transitions between successive vowels serve to distinguish any intervening consonants. There is thus a clear need to control for these linguistic factors. One way, which is analogous to the processing of f0, is to average all the formants across long stretches of speech. This method is referred to as long-term formant analysis (**37**). Another method is to measure formant frequencies separately for different vowels. This is the traditional way formants are measured in phonetics. Beyond anatomical restrictions, there are degrees of freedom in transitioning from one target sound to the next. Hence, measuring formant dynamics is a third way of capturing formant information in forensic speaker recognition. It has been shown in many studies that when formant measurements are not limited to targets but, rather, when one takes into account the entire dynamics of the formant movements, speaker recognition capability is improved (see McDougall (**30**) and Morrison (**33**) for some of the early studies).

6.3.3 There are other speaker characteristics that can be based upon acoustic phonetic analysis (**12**). For example, it is possible to measure the spectral energy distribution in fricatives or nasals (**26**) or to make temporal measurements in the domain of rhythm and timing (Dellwo (**5**); Plug et. al. (**38**)). But most actual AcPA casework utilizes f0 and formants.

6.4 *Auditory Phonetic + Acoustic Phonetic Analysis (AuPA+AcPA):* AuPA+AcPA is the combination of both auditory and acoustic analysis in speaker recognition as detailed in §6.2 and §6.3. The combination of AuPA and AcPA has also been referred to as an “auditory-acoustic-

phonetic" method that is carried out by forensic practitioners (**35**)." The term "auditory-acoustic-phonetic by forensic practitioners (qualitative opinion)" reflects how the phonetic data are traditionally interpreted by many practitioners of AuPA+AcPA: though there can be quantification on the feature level (e.g., formant frequencies in Hz; values on a scale of perceived voice qualities), the results are most commonly interpreted qualitatively, e.g., as distances of the values of the unknown and known speaker that is visible in a plot of formant values, or as an experience-based judgment of how frequently a certain speaker characteristic occurs in a relevant population (**35**). Such a qualitative expression of AuPA+AcPA can approach a quantitative likelihood ratio-based method if there are data available of the relevant population, for example of the f0 values of male speakers (**13**). But for full expression of quantitative likelihood ratios, the statistical methodology has to be present, as well as all the necessary data, such as non-contemporary same-speaker and different-speaker data.

6.5 *Semi-Automatic Acoustic Speaker Recognition (SASR):* Semi-Automatic Speaker Recognition refers to a method by which speaker characteristics are derived by phonetic analysis and then compared. The result of the comparison process is a numerical likelihood ratio (**6**). Acoustic phonetic analysis is used to perform SASR, though it is possible to base SASR on auditory phonetic analysis as well (**1**). The prefix "semi" of this term refers to the fact that phonetic analysis is not a fully automatic process. Even if automatic routines such as formant or pitch tracking are used, there needs to be supervision by a phonetic practitioner. The stem "automatic" refers to the fact that procedures after the point of the phonetic feature extraction stage proceed automatically, which includes speaker modeling, distance scoring and calibration. SASR can be

used in casework (Rose (**41**) for a case study), but it is more commonly used as a research method (e.g., Morrison (**34**); Gold (**11**)).

6.5.1 SASR, as defined here, can be viewed as a special case of AcPA, as expressed by Foulkes and French (2012). However, it has as much in common with automatic speaker recognition (ASR) or Human Assisted Automatic Speaker Recognition (HAASR; see below), as it has with AcPA with which it shares feature extraction, and with ASR/HAASR it shares automatic processing of speaker modeling, distance scoring and calibration (though the details of how these three processing steps are carried out might differ). In that sense, SASR is a method that lies between those other methods and merits a standing of its own. This is supported by the classification of methods (analytical approaches) in the "INTERPOL survey of the use of speaker identification by law enforcement agencies" (**35**).

6.6 *Automatic Speaker Recognition (ASR):* ASR requires the use of specialized software to compare speech samples, producing a score that is evaluated from the perspective of the same-speaker origin as well as the different-speaker origin. A typical ASR process involves the extraction of features from a speech sample and the use of these features to create a representation of the speaker's voice. Speaker representations from different samples are then compared to produce a comparison score. Although the underlying algorithms have evolved, this general procedure has remained consistent. In a fully automatic system, the person performing the comparison would simply provide the ASR software with their audio samples and record the comparison score often expressed as a likelihood ratio (LR). It is important to note that intrinsic speaker variability (e.g., variations in a speaker's speech production), extrinsic conditions (e.g., recording conditions, equipment, environmental, communications/channel, etc.) and context

(e.g., specific scenario in which speech communication takes place) all contribute to mismatch. Such mismatch presents challenges in ensuring consistent and effective automatic speaker recognition performance by machine. An overview of recent automatic speaker recognition practices can be found in Hansen and Hasan (**17**). Automatic speaker recognition techniques have been continuously developing over several decades and have evolved from template matching techniques like vector quantization (**15**), to techniques that statistically model the probability density of speech features like Gaussian mixture models (GMM) (**39**), and further to more compressed representations of the speech features like i-vectors (**4**), and x-vectors (**43**) based on deep neural networks. With each of these successive techniques speaker recognition performance has improved significantly, particularly in challenging recording conditions. Although speaker recognition can now be considered a mature technology it is an active area of ongoing research with new algorithms and techniques bringing further improvements in robustness and discrimination. In practice, the researcher will typically use ASR software in a supervisory role, by applying some pre-processing to the audio samples in the case, and by providing the ASR software with additional case-relevant audio samples.

6.7 *Human Assisted Automatic Speaker Recognition (HAASR):* HAASR describes the way in which ASR systems are typically used for forensic speaker recognition in practice. In HAASR, the practitioner assesses the audio samples to inform the appropriate use of the ASR software. The first step in this process is to determine whether the samples satisfy the minimum requirements of the system (with respect to the duration of speech content, the technical quality of the samples in terms of signal-to-noise ratio, etc.). If such requirements are met, the audio samples will usually be pre-processed before providing them to the ASR software. This may involve

removing parts of the audio sample that do not contain speech or extracting the relevant portions of an audio sample in which multiple voices can be heard. When providing pre-processed samples to the ASR system, the researcher may also provide additional case-relevant audio samples; these may be used by the system to adapt to the conditions of the case, and to express the result of the comparison as a likelihood ratio. The selection of such case-relevant data by the researcher should be informed by relevant testing.

6.8 *AuPA+AcPA and (HA)ASR:* This method involves the combination of both human (AuPA+AcPA) analysis and (HA)ASR analysis. The precise nature in which the two methods are combined varies largely by practitioner. Typically, there is an ASR analysis carried out, and those results are reported along with AuPA+AcPA. The AuPA+AcPA analysis will vary in breadth and depth, as the analysis will largely depend on who is conducting the analysis. The human side of the analysis may include the entire gamut of speech parameters analyzed in typical AuPA+AcPA cases, or it may be limited to a smaller subset (**12**).

6.8.1 It is important to note that the term, HASR (Human Automatic Speaker Recognition), has also been used in the speaker recognition community to refer to what is described in both §6.7 and §6.8. The term HASR is a conflated term that does not distinguish the level of human involvement in a given analysis.

7. Report

7.1 *Interpretation Frameworks:* Interpretation Frameworks, known in the speaker recognition community as Conclusion Frameworks, are used to present evaluated evidence. Most Conclusion Frameworks can be adopted regardless of analysis method. Indeed, it is important to

note that different analyses lend themselves much more readily to certain Conclusion Frameworks than others. This section outlines the types of conclusion framework that are utilized by the scientific community: a binary decision, probability scale, likelihood ratio (whether verbal or numerical), the UK Position Statement, and support statements. Each framework is discussed in turn, while making note of which methods of analysis (from §6) lend themselves more readily to which Conclusion Frameworks.

7.2 *Binary Decision:* The most simplistic, but arguably the least common framework, is the binary decision. The binary decision is "a two-way choice that either the [known and the unknown speaker] are the same person or different people" (**12**). Under a binary decision framework, the evidence can only point one way or the other, and there are no means for the practitioner to further indicate the strength of the evidence.

7.2.1 Any method of analysis can be represented as a binary decision. However, for inconclusive evidence, no decision would be given.

7.3 *Probability Scale (PS):* A probability scale is situated within a frequentist (statistical) framework insofar as there is only a single hypothesis being raised. The practitioner is typically asking themselves what the probability of a hypothesis is given the evidence. In practice, this often takes the form of verbal opinion (rather than quantitative ones), such as "it is likely/very likely to be the same (or different) speakers" (**12**). The probability scale avoids the need to make a "categorical judgment about […] voices (e.g., the two voices come from the same speaker)" (**12**). Rather, probability scales allow experts to express the strength of their findings. As a result, "[t]his means that the evaluation of forensic speech samples will not yield an absolute

identification or elimination of the suspect but instead provides a probabilistic confidence measure" (**17**: p.80).

7.3.1 Any method of analysis can be represented as a probability scale, and it is a common Conclusion Framework for forensic speaker recognition, though declining in frequency internationally (**13**).

7.4 *Likelihood Ratio (LR):* As with the probability scale, a framework using the LR "provides a gradient estimation of the strength of the evidence, however, the LR does so based on the ratio of probability (p) of the evidence (E) given (|) the prosecution hypothesis (Hp) to the probability of evidence given the defense hypothesis (Hd)" (**40**). When applied to forensic speaker recognition, the LR "consists of an assessment of the similarity between the known and questioned samples with regard to a given parameter and the typicality of those values within the wider, relevant population" (**14**: p. 293).

7.4.1 When used as a Conclusion Framework, the likelihood ratio can either be presented numerically or verbally. When presented numerically, the LR is expressed as a value centered on one, such that LRs >1 offer support for the hypothesis that the known and unknown voices are the same, and LRs <1 offer support for the hypothesis that the known and unknown voices are not the same. The magnitude of the LR determines how much more likely the evidence would be under the competing hypotheses (**14**). Rose (**40**: p.58) illustrates this with the following example:

7.4.1.1 We assumed that a high degree of similarity was observed between [unknown] and [known] speech samples. This high degree of similarity constitutes the evidence. We assumed further that 80% of paired speech samples with this high degree of similarity have been shown to be from the same speaker. Thus the probability of observing the evidence assuming the

samples are from the same speaker *p*(E | Hp) is 80%. Now, in order to determine the strength of the evidence, we also need to take into account the percentage of paired speech samples with this high degree of similarity that have been shown to come from different speakers. Let us assume it is 10%. The probability of observing the evidence assuming the samples come from different speakers *p*(E | Hd) is thus 10%.

7.4.1.2 Rose (**40**) then uses those two probabilities above, 80% and 10%, respectively, and divides in order to obtain a LR for the evidence that is 80:10 = 8. Rose (**40**: p. 58) provides the interpretation of this result as "the degree of similarity [being] eight times more likely to be observed were the samples from the same speaker than from different speakers."

7.4.1.3 When LRs are presented verbally (see the ENFSI 2015 guidelines for an example of this in practice; **47**), practitioners typically adopt a scale such as that of Champod and Evett (**2**) which associates a verbal expression with a Log10 likelihood ratio. It is also possible for practitioners to arrive at a verbal LR without converting from a Log10 LR. A verbal LR may come about when, for example, parameters in an analysis may be difficult to quantify, when quantitative and qualitative evidence are combined, when reference data may not exist for the exact population under question and a similar one is substituted, etc.

7.4.1.4 Unlike some of the other Conclusion Frameworks, numerical LRs are most readily obtained by those methods that are entirely quantitative in nature (specifically those in §6.3, §6.5, and §6.6). Verbal LRs, on the other hand, are not as restrictive in terms of the methods that need to be employed. All methods in §6 above can be presented as a verbal LR framework.

7.5 *UK Position Statement:* Originally developed for use in the United Kingdom, the UK Position Statement has now been adopted by other countries outside the UK. The UK Position

Statement “involves a potentially two-part decision. The first part concerns the assessment of whether the samples are compatible, or consistent, with having come from the same person. The second part, which only comes into play if there is a positive opinion concerning consistency, involves an evaluation of how unusual or distinctive the features common to the samples may be” (**12**). It is important to note that practitioners in the United Kingdom are no longer using the UK Position Statement as it has been superseded by support statements, (see §7.6).

7.6 *Support Statement:* Support statements were originally proposed by Champod and Evett (**2**) as a way of presenting opinions in a simplified manner that allows more clarity for the trier of fact who is interpreting the evidence. LRs are understandably more complex in communicating to the trier of fact, and support statements were introduced to mitigate this problem. Support statements were adopted in place of the UK Position Statement by almost all UK forensic phoneticians in 2015. The scale of support statements and correspondences with both verbal and numerical LRs is set out in French (**9**).

7.6.1 The results of any method of analysis can be represented as a support statement, and as of 2022 it is the United Kingdom’s preferred method for providing opinions in speaker recognition cases.

8. Common Practices:

8.1 Several comprehensive surveys of practitioners of forensic speaker recognition around the world (Gold and French (**12**); Morrison et al. (**35**); Gold and French (**13**)) have determined that their formal methodologies and conclusion frameworks vary. The most recent of these surveys (**13**) found that around 60% of practitioners report using the combined AuPA and AcPA approach, though the percentage of practitioners using ASR has risen

considerably since the previous survey in 2011.

9. Keywords

9.1 speaker recognition; forensic speaker recognition; forensic speech science; forensic phonetics; speaker identification; voiceprinting; methodology; auditory phonetic analysis; acoustic phonetic analysis; semi-automatic acoustic speaker recognition; automatic speaker recognition; human-assisted automatic speaker recognition; conclusion frameworks; binary decision; probability scale; likelihood ratio; UK Position Statement; support statement.

References

1. Aitken, C., and Gold, E., “Evidence Evaluation for Discrete Data,” *Forensic Science International*, Vol 230, Issues 1-3, 2013, pp. 147-155.
2. Champod, C., and Evett, I. W., “Commentary on A. P. A. Broeders (1999) ‘Some Observations on the Use of Probability Scales in Forensic Identification,’ Forensic Linguistics 6(2): 228–41,” *International Journal of Speech, Language and the Law*, Vol 7, No. 2, 2000, pp. 239–243.
3. De Boer, M., and Heeren, W., “The Speaker-Specificity of Filled Pauses: A Cross-Linguistic Study,” *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 2019, pp. 607-611.
4. Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., and Ouellet, P., “Front-End Factor Analysis for Speaker Verification,” *IEEE Transactions on Audio, Speech, and Language Processing*, Vol 19, No. 4, May 2011, pp. 788–798.
5. Dellwo, V., Leeman, A., and Kolly, M.-J., “Rhythmic Variability Between Speakers: Articulatory, Prosodic, and Linguistic Factors,” *Journal of the Acoustical Society of America*, Vol 137, No. 3, 2015, pp. 1513–28.
6. Drygajlo, A., Jessen, M., Gfroerer, S., Wagner, I., Vermeulen, J., and Niemi, T., *Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition*, Frankfurt, Germany: Verlag für Polizeiwissenschaft, 2015, Online, Available: http://enfsi.eu/wp-content/uploads/2016/09/guidelines_fasr_and_fsasr_0.pdf.
7. Foulkes, P., and French, P., "Forensic Speaker Comparison: A Linguistic–Acoustic Perspective," in *The Oxford Handbook of Language and Law*, Oxford University Press, Oxford, 2012, pp. 557–72.
8. Gray, C., and Kopp, G., “Voice Print Identification,” Bell Telephone Laboratories, Inc., 1944.
9. French, P., “A Developmental History of Forensic Speaker Comparison in the UK,” *English Phonetics*, Vol 21, 2017, pp. 255–270.
10. French, P., and Harrison, P., “Position Statement concerning use of impressionistic likelihood terms in forensic speaker comparison cases, with a foreword by Peter French & Philip Harrison.” *International Journal of Speech, Language and the Law*, Vol 14, 2007, pp. 137-144.
11. Gold, E., “Calculating Likelihood Ratios for Forensic Speaker Comparisons Using Phonetic and Linguistic Parameters,” PhD diss., University of York, 2014, Online, Available: https://etheses.whiterose.ac.uk/6166/.
12. Gold, E., and French, P., “International Practices in Forensic Speaker Comparison,” *International Journal of Speech, Language and the Law*, Vol 18, No. 2, 2011, pp. 293–307.

13. Gold, E., and French, P., “International Practices in Forensic Speaker Comparison: Second Survey,” *International Journal of Speech, Language and the Law*, Vol 26, No. 1, 2019, pp. 1-20.
14. Gold, E., and Hughes, V., “Issues and Opportunities: The Application of the Numerical Likelihood Ratio Framework to Forensic Speaker Comparison,” *Science and Justice*, Vol 54, No. 4, 2014, pp. 292-299.
15. Gray, R. M., “Vector Quantization,” *IEEE ASSP Magazine*, Vol 1, No. 2, April 1984, pp. 4–29.
16. Gruber, J. S., and Poza, F. T., “Voicegram Identification Evidence”, *American Jurisprudence Trials*, Vol 54, 1995.
17. Hansen, J.H.L., and Hasan, T., “Speaker Recognition by Machines and Humans: A Tutorial Review,” *IEEE Signal Processing Magazine*, Vol 32, No. 6, November 2015, pp. 74-99.
18. Hollien, H., *The Acoustics of Crime, The New Science of Forensic Phonetics*, Plenum Press, New York, NY, 1990.
19. Hollien, H., *Forensic Voice Identification*, Academic Press, San Diego, CA, 2002.
20. Hughes, V., Wood, S., Foulkes, P., “Filled Pauses as Variables in Forensic Voice Comparison,” *International Journal of Speech, Language and the Law*, Vol 23, No. 1, 2016, pp. 99−132.
21. Hudson, T., de Jong, G., McDougall, K., Harrison, P., and Nolan, F., “F0 Statistics for 100 Young Male Speakers of Standard Southern British English,” *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 2007, pp. 1809–12.
22. Hughes, V., and Rhodes, R., “Questions, Propositions and Assessing Different Levels of Evidence: Forensic Voice Comparison in Practice,” *Science and Justice*, Vol 58, 2018, pp. 250–57.
23. Jessen, M., Köster, O., and Gfroerer, S., “Influence of Vocal Effort on Average and Variability of Fundamental Frequency,” *International Journal of Speech, Language and the Law*, Vol 12, No. 2, 2005, pp. 174–213.
24. Jessen, M., “The Forensic Phonetician: Forensic Speaker Identification by Experts,” *The Routledge Handbook of Forensic Linguistics*, Routledge, London, UK, 2010, pp. 378–394.
25. Jessen, M., “Speaker Profiling and Forensic Voice Comparison: The Auditory-Acoustic Approach,” *The Routledge Handbook of Forensic Linguistics*, Routledge, London, UK, 2021, pp. 382–399.
26. Kavanagh, C.M., “New Consonantal Acoustic Parameters for Forensic Speaker Comparison,” PhD thesis, University of York, 2012, Online, Available: https://etheses.whiterose.ac.uk/3980/
27. Keating, P., Garellek, M., and Kreiman, J., “Acoustic Properties of Different Kinds of Creaky Voice,” *Proceedings of the 17th International Congress of Phonetic Sciences*, August 2015, Vol 2015, No. 1, pp. 2-7.

28. Kersta, L.G., “Voiceprint Identification,” *Nature*, Vol 196, No. 4861, December 1962, pp. 1253-1257.
29. Laver, J., *The Phonetic Description of Voice Quality*, Cambridge University Press, Cambridge, UK, 1980.
30. McDougall, K., “Dynamic Features of Speech and the Characterization of Speakers: Towards a New Approach Using Formant Frequencies,” *International Journal of Speech, Language and the Law*, Vol 13, No. 1, 2006, pp. 89–126.
31. McDougall, K., and Duckworth, M., “Individual Patterns of Disfluency Across Speaking Styles: A Forensic Phonetic Investigation of Standard Southern British English,” *International Journal of Speech, Language and the Law*, Vol 25, No. 2, 2018, pp. 205–30.
32. McIntyre, D., Lesley J., Evans, M., Price, H., and Gold, E., *The Babel Lexicon of Language*, Cambridge University Press, Cambridge, UK, 2022.
33. Morrison, G. S., “Likelihood-Ratio Forensic Voice Comparison Using Parametric Representations of the Formant Trajectories of Diphthongs,” *Journal of the Acoustical Society of America*, Vol 125, No. 4, 2009, pp. 2387-2397.
34. Morrison, G. S., “A Comparison of Procedures for the Calculation of Forensic Likelihood Ratios from Acoustic-Phonetic Data: Multivariate Kernel Density (MVKD) Versus Gaussian Mixture Model - Universal Background Model (GMM-UBM),” *Speech Communication*, Vol 53, No. 2, February 2011, pp. 242-256.
35. Morrison, G. S., Sahito, F. H., Jardine, G., Djokic, D., Clavet, S., Berghs, S. and Goemans Dorny, C., “INTERPOL survey of the use of speaker identification by law enforcement agencies,” *Forensic Science International*, Vol 263, June 2016, p p. 92–100.
36. National Research Council, *On the Theory and Practice of Voice Identification*, the National Academies Press, Washington, DC, 1979, Online, Available: https://doi.org/10.17226/19814.
37. Nolan, F. and Grigoras, C., “A Case for Formant Analysis in Forensic Speaker Identification,” *International Journal of Speech, Language and the Law*, Vol 12, No. 2, 2005, pp.143–73.
38. Plug, L., Lennon, R., and Gold, E., “Articulation Rates’ Inter-Correlations and Discriminating Powers in an English Speech Corpus,” *Speech Communication*, Vol 132, September 2021, pp. 40-54.
39. Reynolds, D. A., “A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification,” PhD thesis, Georgia Institute of Technology, Atlanta, Georgia, 1992.
40. Rose, P., *Forensic Speaker Identification*, Taylor & Francis, London, UK, 2002.
41. Rose, P., “Where the Science Ends and the Law Begins: Likelihood Ratio-Based Forensic Voice Comparison in a $150 Million Telephone Fraud,” *International Journal of Speech, Language and the Law*, Vol 20, No. 1, 2013, pp. 277-324.

42. San Segundo, E., Foulkes, P., French, P., Harrison, P., Hughes, V., and Kavanagh, C., "The Use of the Vocal Profile Analysis for Speaker Characterization: Methodological Proposals," *Journal of the International Phonetic Association*, Vol 49, No. 3, December 2019, pp. 353–80.
43. Snyder, D., Garcia-Romero, D., Sell, G., Povey, D., and Khudanpur, S., "X-Vectors: Robust DNN Embeddings for Speaker Recognition," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 5329-5333.
44. Solan, L., and Tiersma, P., "Hearing Voices: Speaker identification in court," *Hastings Law Review*, Vol 54, No. 2, 2003, pp. 373-435.
45. Tosi, O., Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J., and Nash, E., "Experiment on Voice Identification," *Journal of the Acoustical Society of America*, Vol 51, No. 6b, 1972, pp. 2030-2031.
46. *US v. Angleton*, 269 F. Supp. 2d 878 (S.D. Tex. 2003).
47. Willis, S., Ligertwood, A., Molina, J.J., Berger, C., Zadora, G., Nordgaard, A., Rasmusson, B., Lunt, L., Champod, C., Biedermann, A., Hicks, T., Taroni, F., Zhu, X., "ENFSI Guideline for Evaluative Reporting in Forensic Science," European Network of Forensic Science Institutes, 2015.