ISSN: 2957-3874 (Print)
Journal of Al-Farabi for Humanity Sciences (JFHS)
https://iasj.rdd.edu.iq/journals/journal/view/95
مجلة الفارابي للعلوم الإنسانية تصدرها جامعة الفارابي

The Contribution of Forensic Linguistics to Speaker Identification: Cues, Reliability, and the Evaluation of Voice Evidence

Lect. Anees Izzat Salman
م.أنيس عزت سلمان
Place of work:The General Directorate for Education of Diyala-Al-Naqaa
Secondry school for Valedictorian boys.
مكان العمل: المديرية العامة لتربية ديالى .ثانوية النقاء للمتفوقين
Phone Number: 07734265064
Email: aneesizzat77@gmail.com

Abstract

Few questions in forensic science are asked as often, or answered as loosely, as whether two recordings carry the same voice. This paper reviews the contribution that forensic linguistics, together with its close relative forensic phonetics, makes to speaker identification, and it sets out to organise a field that has grown quickly but unevenly. The review has three aims: to distinguish the analysis of voice from the analysis of written language, both of which shelter under the label "forensic linguistics"; to gather what is known about where a speaker's individuality actually resides and how well each source of it survives disguise and poor recording conditions; and to place all of this inside the evaluative logic that now governs responsible practice, the likelihood ratio. Out of that synthesis I propose the Levels–Robustness–Evaluation (LRE) framework. LRE holds that the evidential value of a voice comparison depends on three things read together: the linguistic level at which an individuating cue sits, the robustness of that cue to disguise and signal degradation, and the way similarity is weighed against typicality in a relevant population. This paper is a conceptual review of the above. No new experimental data are described; it synthesises from the literature and provides a heuristic framework, propositions, and agenda for testing. The reasoning behind it is that both of these types of errors in speaker identification are weakly measured with lower bounds, and that the myth (the "voiceprint") — that every voice has one-of-a-kind acoustic like a fingerprint when identifying a specific individual by their respect to other individuals(W), is still the scientifically-worst false inference made about such a phenomenon, outside the absence of any similarity judgments for non-ancient humans.

Keywords: *forensic linguistics; forensic phonetics; speaker identification; voice comparison; likelihood ratio; voice disguise; authorship attribution; idiolect*

ملخص

قلّما نجد سؤالاً في العلوم الجنائية يُطرح بتكرار –أو تُقدم له إجابات تفتقر إلى الدقة المنهجية– مثل السؤال عما إذا كان تسجيلان صوتيان يعودان للشخص نفسه. تستعرض هذه الورقة مساهمة "اللسانيات الجنائية" –إلى جانب تخصصها الشقيق "الصوتيات الجنائية"– في مجال تحديد هوية المتحدث، وتهدف إلى تنظيم هذا المجال الذي شهد نمواً سريعاً ولكنه غير متوازن. وللمراجعة ثلاثة أهداف: التمييز بين تحليل

الصوت وتحليل اللغة المكتوبة (وكلاهما يندرج تحت مظلة "اللسانيات الجنائية")؛ وتجميع المعارف المتاحة حول مكامن التفرد الصوتي لدى المتحدث ومدى صمود كل مصدر من مصادر هذا التفرد أمام محاولات التمويه وظروف التسجيل الرديئة؛ ووضع كل ذلك في سياق المنطق التقييمي الذي يحكم الممارسات المهنية المسؤولة حالياً، ألا وهو "نسبة الاحتمالية (likelihood ratio). "وانطلاقاً من هذا الطرح التجميعي، أقترح إطار عمل يرتكز على ثلاثة محاور: المستويات(Levels) ، والمتانة(Robustness) ، والتقييم(Evaluation) ، ويُشار إليه اختصاراً بـ (LRE). يرى هذا الإطار أن القيمة الإثباتية لمقارنة الأصوات تعتمد على ثلاثة عناصر تُقرأ مجتمعة: المستوى اللغوي الذي تقع فيه السمة المميزة للهوية، ومدى متانة هذه السمة وقدرتها على الصمود أمام التمويه وتدهور جودة الإشارة الصوتية، وكيفية الموازنة بين التشابه والنمطية (أو الشيوع) ضمن المجموعة السكانية ذات الصلة. تُعد هذه الورقة مراجعةً مفاهيميةً لما سبق ذكره؛ فهي لا تعرض بيانات تجريبية جديدة، بل تجمع وتصيغ ما ورد في الأدبيات العلمية لتقدم إطاراً استرشادياً، ومجموعة من الفرضيات، وخطة عمل للاختبار المستقبلي. ويستند هذا الطرح إلى حقيقة أن كلا نوعي الخطأ في تحديد هوية المتحدث يُقاسان بضعفٍ (من خلال تقدير الحدود الدنيا للخطأ)، وأن الخرافة القائلة بوجود "بصمة صوتية" فريدة لكل شخص –على غرار بصمة الإصبع– تظل أسوأ استنتاج علمي خاطئ حول هذه الظاهرة. الكلمات المفتاحية: اللسانيات الجنائية؛ الصوتيات الجنائية؛ تحديد هوية المتحدث؛ مقارنة الأصوات؛ نسبة الاحتمالية؛ تمويه الصوت؛ تحديد نسبة النص للمؤلف؛ اللهجة الفردية.(Idiolect)

1. Introduction

1.1 Background

The wish to identify a human being by means of their voice is as old as humanity itself, and for most of that time exceeded any means available to satisfy it. Judges, investigators and police forces have been chasing the holy grail of being able to connect or eliminate a suspect from a recording with similar level of certainty as fingerprinting allows (Bolt et al., 1970; Rose, 2002). This is born in that data gap—between police detectives, for instance interested in using an expert witness who just happens to have expertise in linguistics (the scientific study of language) and the absence of data-driven expertise until recently. It employs knowledge of language to provide solutions in legal contexts, particularly two questions common in criminal work: who is the speaker of this recording and who was the author of this text (Coulthard, 2004; Gibbons, 2003). This review is concerned with speaker identification, which sits on the spoken end of that continuum (although the other end runs very close in parallel as shown in Section 6).

Those jobs are more difficult than the public realises. In a case often the samples are not even comparable in relevant aspects, like one an unvoiable attempt at an exclamative and the other a quiet whisper, one could be couched or influenced by varying levels of intoxication/stress they vary length-wise and loudness wise and occasionally lack material to tie comparisons with (Jessen, 2008; Rose, 2002). On top of those slip-ups is the subtlety of speech itself, which can truly never be avoided: no one ever says a given word exactly as someone else. Identification is a hard problem, even for automatic systems and in practice for human examiners on data dated only to October 2023.

1.2 The Problem: The Voiceprint Fallacy and the Status of Voice Evidence

The single, misleading word that accounts for much of the disaster in that area submitted a succinct report to Nature in 1962 announcing vocabulary printing identification (Kersta, 1962) and also the label stuck. It resides in tabloids, on crime shows and in the courts, and it offers a false kind of perfection: that spectrograms are to voice as papillary ridge is to finger — an immutable unique fingerprint that you wear forever. It is not. A recording is not an imprint of a steady state organ, but rather a remanent trace of a fluid, variegate, imitable stance [16]; and the idea that all voices are the same and recognizable with ease has done real damage (Boë 2000) The field has been attempting to fix it ever since.

The scientific community did just that, early and often. The view that spectrographic identification could be applied successfully in courts of law had been called into doubt by a group of scientists bemoaning the paucity of cases where, for example (Bolt et al., 1970); as their findings were also published in the national press, they were followed, first, by a panel confronting the theory and practice /of voice identification /head on (Bolt et al., 1979), then by British phoneticians expressing similar misgivings over the next decades; and from 1990

onward many commentators drawn from within speech community at large voiced similar concerns about deceitful evidence beneath all their philosophical interest in "dead" texts (Boë, 2000). In 2003, a panel of experts put together what can be described as the definitive statement on the matter (Bonastre et al., 2003): no current method can demonstrate that two recordings have one speaker instead of simply having been produced under similar circumstances, and both human or machine require careful control and interpretation when used to identify individual speakers. Voice evidence is not without merit — quite the contrary, of course. In other words, the evidence must be presented as evidence, with the uncertainty surrounding it openly articulated rather than left cloaked in secrecy and, perhaps, ambiguity.

1.3 Aims and Scope

A word on genre, as that will colour the following. This is a conceptual review. It does not offer its own experiments, but rather assembles existing research and sorts through it. Three questions guide the paper. What is the first link between forensic analysis of a voice and written language, both referred to as forensic linguistics? Second, what is a speaker identity in detail, and to what degree does each element hold up against masking and lossy coding? Third, how does one next describe evidential weight in terms fit for courts from that deep structural similarity of samples? Scope — Criminal and forensic speaker verification regarding the task; here we cover forensic phonetics and automatic speaker recognition (as opposed to aiming for completeness over all of language-and-law, with MA as a written corollary).

1.4 Significance

The stakes are not abstract. A better account of the identification of speakers can fill–in (make a case), break (exonere) or help support credible expert testimony perplexing cases but requiring keeping within what science will really evidence given to prove their case, Morrison 2009. The standing of the field also depends on it. With courts growing ever more dependent on linguistic and acoustic testimony, explain with great credibility why spoken language has a closer connection to thought than writing does; how the existence of this field will be secured potentiation by defining limits audibly as they do findings. What this review adds is an attempt to pin together the disparate findings across a field where someone implementing them could quickly glean why one cue trumps another on occasion, or how a very different looking but equally valued match will actually hold far less water.

1.5 Review Approach and Literature Selection

However, this current paper is a narrative conceptual review and does not follow PRISMA protocol. The selection of the literature was based more on conceptual relevance and representativeness than an aim to comprehensively cover the dimensions of this framework. Design was informed by (1) our ornamental theory of forensic phonetics and speaker individuality, (2) empirical literature exploring voice disguise, signal degradation and their impact on recognition as well as by (3) evaluative literature examining typicality and the likelihood ratio. We complement these foundations with recent scientific contributions, powered by deep and self-supervised speaker representations, spoofing and deepfake detection, and authorship attribution in the large language models era making sure that the review is up to date w.r.t. current research landscape.

This review takes on a subset of the more well-measured periodical journals and their associated global conferences from informatics, linguistic recognition and forensic evidence analysis (the International Journal of Speech, Language and the Law; Speech Communication; Science & Justice; and Journal of the Acoustical Society of America), too with an essential distillation of earlier ISCA (Interspeech, Odyssey, Eurospeech), and IEEE (ICASSP) Procedures. The articles selected were mostly from peer-reviewed journals or other well-established, peer-reviewed proceedings and include a broad range extending from landmark mid-twentieth century papers to 2024; when an assertion depends on a clearly defined empirical result, citation is provided if needed to its original report rather than a secondhand account. Thus these two allied traditions have still been considered separate worlds: forensic (human-expert) analysis on the one hand, automatic speaker recognition on the other; they address the same forensic question (what is a conclusion) but necessarily using different tools and with different error profiles. Due to its story-driven, the choice it involves is inescapably nan-subjective and section 9 paper this misconduct.

2. Clarifying the Terrain: Forensic Linguistics and Forensic Phonetics

Forensic linguistics is a catch-all term that encompasses two rather disparate enterprises, and conducting both in tandem creates confusion where none need exist. In another, work occurs in social speech as written and recorded language: authorship, its meaning and threat as are interpreted in something said or wrote (Correa, 2013; Ariani et al., 2014). The second, known generally as forensic phonetics, synthesises within this framework but obtains its data from the speech signal itself: what your acoustics and your phonetics tell you about the speaker (Jessen 2008; Nolan 1997). Other popular topics include INDIVIDUAL CLIENT FACTORSSpeech language identification, which exists mainly in the second, drawing near-constant from the first — since a voice carries not only acoustic utilities but also vocabulary and grammar (even dialect) — as so much an individualizing characteristic as any formant.

The history is short. Speakers had been identified as far back as the early twentieth century, but it took most of a mid-century for the field to gain acceptance. Spectrographic "voiceprint" analysis in the 1960s faced immediate court challenges, while acoustic phonetics advances which studied the physical properties of speech as opposed to detailed visual representations provided more reliable footing (Kersta, 1962; Bolt et al., 1970). Since 1990, machine learning and related statistical methods based on author research provide a largely automatic angle on normal science in which systems learn to identify patterns over massive collectives far beyond the memorization capacity of analysts (Reynolds, 2002; Snyder et al., 2015). More recent systems learn speaker representations directly with deep neural networks, and the current generation increasingly draws on self-supervised models pre-trained on very large unlabelled corpora (Bai & Zhang, 2021; Chen et al., 2022). Alongside these methods runs speaker profiling, which infers a speaker's likely age, sex, or regional background rather than naming them, and which has become a recognised sub-task in its own right (Schilling & Marsters, 2015; Jessen, 2008).

3. The Levels of Speaker Individuality

If a voice is to identify a person, something in it has to be individual. That something is not one thing but many, spread across the levels of language, and the first move of the framework offered later is simply to keep those levels apart. No two people are identical in their vocal make-up, since they differ in anatomy, in physiology, and in the acoustic properties these produce; yet even identical twins, whose anatomy is as close as nature allows, articulate the same sounds in measurably different ways (Nolan & Oh, 1996). Individuality, in other words, is partly given by the body and partly learned or chosen, and evidence of it can be gathered at several levels at once.

3.1 Acoustic-phonetic cues

At the lowest level sit the physical properties of the speech sounds: fundamental frequency and its behaviour over time, the formant frequencies that reflect the size and shape of the vocal tract, spectral detail, and timing. These features are the mainstay of both human and automatic comparison, and in favourable conditions they discriminate speakers well (Rose, 2002; Nolan, 1997). Their weakness is that several of them, fundamental frequency above all, are partly under the speaker's control and so are open to disguise, a point Section 4 takes up.

3.2 Phonological and dialectal cues

Above the raw acoustics lies the speaker's sound system: which vowels and consonants they use, how they realise them, and the regional or social accent these choices reveal. Part of the reason there is such a strong emphasis on accent and dialect in profiling is because they can, even when the identity itself is unknown, weed out most of that population into a speaker (Schilling & Marsters 2015). Since these features are learnt endeavoured and run deep, they are far more difficult to erase convincing than only a pitch shift (though an ingenious impersonator can still send them to their torment).

3.3 Lexico-grammatical, idiolectal, and stylistic cues

Moving up, you find the elements which cross over from phonetic to genuine linguistic features: habitual word choice, grammatical preferences, discourse patterns; even the fillers and formulae a person turns to automatically. Together these form something like an idiolect, the individual ego's version of the a collective tongue (Coulthard, 2004). Such features are valuable for two reasons. They survive some acoustic disguises untouched, since a person who raises their pitch does not thereby change their favourite connective, and they connect the spoken task directly to authorship analysis, where the same habits are read off the page. Automatic systems have tried to exploit exactly this high-level, idiolectal information rather than acoustics alone (Reynolds et al., 2003). Their difficulty is that they need enough material to show through, and short samples rarely provide it.

4. Voice Disguise and Signal Degradation

Whatever individuality a voice holds, a case rarely delivers it cleanly. Two forces stand between the examiner and a fair comparison: deliberate disguise by the speaker, and accidental degradation of the signal. Both attack the cues of Section 3, and, importantly, they do not attack them evenly.

4.1 Pitch disguise

Altering pitch is the commonest way a speaker tries to hide (Zhang, 2012). It can be done by hand, by raising or lowering the voice, or electronically, and electronic methods of disguise are by now well documented in the literature (Farrús, 2018). Pitch disguise degrades automatic recognition, and both raised and lowered pitch alter acoustic properties such as intensity, formant frequencies, speaking rate, and long-term spectral measures, though in different ways (Zhang, 2012; Zhang & Tan, 2008). Speakers themselves vary in how well they can carry a disguise off. A line of work has therefore tried to detect and classify electronic pitch-shifting rather than merely suffer it, using cepstral features and machine classifiers to separate disguised from original speech and to sort the direction of the shift (Wu et al., 2014; Ye et al., 2020), while related methods target spoofed or synthetically manipulated voices (Liang et al., 2017). Human listeners, and not only automatic systems, are affected by electronic disguise as well (Clark & Foulkes, 2007). The motivation is practical: knowing the kind of disguise in play can be used to recover some of the lost recognition accuracy (Farrús, 2018). Beyond deliberate disguise by a human speaker, synthetic manipulation is now a related and fast-growing threat, since text-to-speech and voice-conversion systems can fabricate speech that imitates a target, and dedicated benchmarks have been created to drive the detection of such spoofed and deepfake audio (Yamagishi et al., 2021).

4.2 Speaker sex, age, and disguise type

The damage a disguise does is conditional, not fixed. It depends on the type of disguise, on the speaker's sex and age, and on individual ability (Hautamäki et al., 2017). The extent to which fundamental frequency and formants move under disguise varies from speaker to speaker, and the effect on automatic systems varies with the disguise: an examination of many common disguise types found that system performance rose and fell with the method used, and that some speakers disguise certain ways well while giving themselves away in others (Zhang & Tan, 2008). Voice quality enters the picture as well, and corpora have been built specifically to study how phonation type bears on forensic speaker comparison (San Segundo et al., 2013). Intentional imitation and voice conversion are a secondary threat, as this type of exaggerated convergence raises error rates in automatic recognition, as shown by (Farrús et al., 2010) A detailed study explored distortions in formants, bandwidths, fundamental frequency and speaking rate since these parameters varied differently between male vs female speakers (González Hautamäki et al., 2019) within modal, intended-old and intended-child environments corresponding to differences in error rates. For vocal types including creak, having apparently resulted from stress and many times observed in the so-called real-world recordings (Tavi et al., 2019), phonation constitutes a target of forensic detection attempts.

4.3 Channel, duration, and mismatch

Recording conditions can cancel out comparison even without a disguise. So when a system is trained on one type of phone i.e. Male and evaluated on the other type of phone i.e. Female, there appears a serious performance drop (as these two phones are mismatching in training and evaluation data). Thus, training could be done under congruent conditions (the gender of the speaker and type of disguise being identical) resulting in accurate recognitions to much greater degree than else have been if this scenario were based upon voice alone (Künzel et al., 2004). A potential solution (Prasad & Prasad, 2019) is concurrent multiple-style training — purposely interleaving styles prior to the input so a rudimentary system can work well when examples are genuine. Wrapped this in with the twins studies of course. They point to the way that features can vary in the absence of degrees in anatomy and other details, and they maintain that judgements which carefully attend to just those relevant features can classify speakers who are extremely similar on the relevant dimension, even in some of these restrictive serious cases (Nolan & Oh, 1996).

5. From Similarity to Evidence: Typicality and the Likelihood Ratio

The key idea in the libertad moderna and arguably the most overlooked in the courtroom. It is not sufficient to only show how similar two samples are. Similarity must be balanced against typicality, that is, how frequent the common features are in a relevant population (Rose 2002). An agreement on a nascent feature points to one speaker much more decisively than does an accord on an ordinary one, and a measure that only identifies what matches—without any scale of how typical those matches are—is silent as to the magnitude of its agreement. This is what the likelihood-ratio framework provides. Instead of declaring an identification, the examiner estimates the probability of the observed similarity on two competing hypotheses, that the samples come from the same speaker and that they come from different speakers, and reports the ratio between them (Champod & Meuwly, 2000; Aitken & Taroni, 2004). A ratio well above one supports same-speaker; well below one supports different-speaker; near one, the evidence is neutral. Automatic systems can help estimate these probabilities against reference populations, and Bayesian interpretation of their scores has become the accepted way to fold them into an evidential statement (Drygajlo et al., 2003). The shift from categorical identification to calibrated likelihood ratios is the paradigm change that has reshaped responsible forensic voice comparison over the last two decades (Morrison, 2009). A subsequent expert consensus has gone further, holding that a forensic-voice-comparison system should be empirically validated under conditions reflecting those of the case at hand before its likelihood ratios are relied on (Morrison et al., 2021). Its virtue is honesty: it states the strength of the evidence and its uncertainty, rather than borrowing the false certainty of the fingerprint that the voiceprint metaphor smuggled in.

6. Authorship Attribution: The Written Parallel

The spoken task has a written twin, and the two illuminate each other. Authorship analysis asks who, among a set of candidates, most likely produced a questioned text, and it has become an established strand of forensic linguistic practice (Coulthard, 2004). The premise is the one met already in Section 3: individuals do not use a language in quite the same way, so a text carries traces of its author's idiolect, spelling, syntax, vocabulary, and habitual patterns (Ariani et al., 2014; Coulthard, 2004). The forensic linguist compares known writings of the candidates with the questioned text and assesses how well each fits (Correa, 2013).

Method here has moved in the same direction as in voice work, toward measurement and away from impression. Analysts draw on content, structural, syntactic, and lexical features, and on stylometric markers such as function-word frequencies, vocabulary richness, and character or part-of-speech n-grams, several of which have proved reliable for distinguishing authors (Grant, 2007; McMenamin, 2002), while the analysis of deliberately disguised writing forms a further strand (Gupta, 2019). Patterns and data sets about greater than human scale that only machines can extract from texts that none, or few, laborers could read (Koppel et al., 2009). Two ports over from voice experience and one decorum caution away. Software, at least on its own, is abhorrently parsimonious in providing any reasoning for decisions, so trusting such software rubs against the demand that an examiner must possess sufficient justification of rational belief to support their decision; recently one has called instead for (Durán et al., 2024) showing that those outputs which are even authoritative regarding their embedded systems could be trusted rather than spread--justifiably--following a

similar evidential logic: Of what little worth might common property enjoy but whatever value inheres in presence alone and not through rarity? The latest twist here is common to both of them, e.g., the period of superfluous prose writing by large language models wrote WPM who mimics such an astonishing degree in somebody else s style that a challenge to quiescent-symbiosis argument between human and machine authorship Analysis has been performed (Huang et al., 2024). Strands of spoken and written projects combine to displace a recycling problem across two media in one discipline.

7. An Integrative Framework: Levels–Robustness–Evaluation (LRE)

Now we can tie all these threads of the review together into one line of thought. This is certainly not my likelihood-ratio paradigm, it is that of the researchers in Section 5. What I am contributing to that is a thread of how it hangs together with the cue levels raised in Section 3, and with what was found in robustness that follows this — so those three can be read off as one judgement. I have coined a new acronym for this relatively simple framework approach to LLRE that stands for Levels–Robustness–Evaluation (LRE) It maintains that evidence of a voice comparison must be based on these three facets, and all in relative time-order for its evidentiary effect.

Levels (L): the linguistic level of an individuating cue; from acoustic-phonetic through phonological, dialectal to lexico-grammatical and idiolectal (Section 3)

Robustness (R): the degree to which that cue survives detection when listeners are best at counter-detection, erosion of signal quality, limits of safe exposure durations and ordinary levels of inter-speaker variation (Section 4).

Evidence (E): resemblance has been seen as do the observed sample occurs in a population that is of practical interest to this research/and more relevantly what we call the so-called likelihood ratio (from Section 5).

Table 1 gives what each one looks at, the input it takes, and its output in order to make these dimensions actionable rather than descriptive.

Dimension	What it examines	Inputs	Output
Levels (L)	Which linguistic levels the two samples agree on: acoustic-phonetic, phonological and dialectal, and lexico-grammatical or idiolectal	The two samples; a feature inventory for each level.	A profile of agreements and disagreements across levels, with their mutual independence noted
Robustness (R)	How far each agreeing cue withstands the disguise, channel, duration, and within-speaker variation present in the case	The case conditions (disguise type, signal quality, sample length); prior findings on cue vulnerability.	A robustness rating per cue, flagging those a speaker could control or the channel could distort.
Evaluation (E)	How typical the surviving agreements are in a case-relevant population, expressed as a likelihood ratio.	A relevant reference population; a comparison method validated under case-like conditions.	A calibrated likelihood ratio with stated uncertainty, valid only under those conditions.

Table 1. Operationalising the three dimensions of LRE.

A decision sequence. It does not apply in parallel, rather sequentially along multiple dimensions. For an examiner the first step is to identify every single common cue and its locus in the shared environment. Next up is robustness: Evaluating the cue in the context of the case — and eliminating those cues that could be affected by the speaker, or manipulated by how the audio was recorded. Only the cues that survive this test are carried forward to evaluation, where the typicality of the agreement is estimated in a case-relevant

population and expressed as a calibrated likelihood ratio. Finally, the examiner combines across levels, giving weight to convergent agreements on independent, robust, and rare features and withholding it from isolated agreements on controllable or common ones.

A gating rule. Robustness gates similarity, and not the other way round. A cue that fails on robustness does not regain evidential weight through similarity alone: a close match on a feature the speaker could fake, or the channel could distort, is weak evidence however close it looks. Similarity is a necessary input to evaluation, never a substitute for robustness or for typicality.

Read this way, LRE yields five propositions, each open to test.

P1. Individuating information is distributed across linguistic levels, and no single level is sufficient. Evidential strength grows as similar findings converge across levels that are, as far as possible, independent of one another.

P2. The forensic weight of a cue is not similarity alone but similarity with respect to how common it is in the population: something reporting similarity without typicality has no evidential meaning.

P3. Robustness to disguise and degradation, at each level. Conspicuous signs, like fundamental frequency, that a speaker can actively control are the first to go, while those associated with adult idiolectal and dialectal features and some involuntary source traits in general outlast disguises that overcome superficial acoustics.

P4. Comparison, alongside almost everything else in this world of ours, is fixed — only conditional. As a function of mask used, speaker sex and age, as well as whether matching was achieved (you should evaluate voice identity on an individual case-by-case basis instead of making inferences across multiple identity cases).

P5. Fair results are contingent upon being offered with qualifications not as absolute accounts but rather merely likelihood ratios that properly quantify uncertainty and the active mitigation of speaker voiceprint fallacy and examiner bias.

Failure conditions. The architecture also notes when it produces no actionable inference. They arise if a non-population member would potentially be classified as part of it; when the questioned sample is too short or degraded to measure its cues reliably;–when the agreeing cues are dependent, making their convergence spurious; or when the competing method has not been validated in conditions similar to those employed in the case (Morrison et al., 2021). Since the framework cannot fail without guidance, conditions are named by design.

LRE is not an empirically validated model itself but rather a decision architecture as well as a source of testable propositions. It lies in the expedient of transforming a very general question (does this voice match) into as ordered an array of answerable questions, on what levels do the samples agree, how reliable are those agreements under the particular conditions of this case represented by sampling and taxonomy; and how representative are the surviving properties in the population that matters. Features that fall under a relatively sure agreement (even of weak strength) on which C of the trait is also general, but for practical purposes are easy to control from one particular level, were coded as trivial features in contrast to LRE; weak cross independent levels gave rise to features with high compensatory effects and ones that cannot be easily concealed or where a restricted range of population numbers present the feature being counted as important rather than viewed as unimportant. Yet it is section 9 that elaborates the mechanisms by which the propositions could be tested.

8. Implications for Forensic Practice and Admissibility

Several practical lessons follow. This first one is about the way we report conclusions. Courts have been asking for a yes/no response forever, and the voiceprint metaphor pushed experts to do so—and it is rare that the evidence justifies that answer. To begin, uncertainty in a LR is both a more honest and defensible position (in legal systems that admit scientific evidence using standards of testability and error rate) (Morrison, 2009; Bonastre et al., 2003). The second concerns standardisation. Tumour markersNew potential biomarkersIt is not enough that no single current procedure has been universally accepted, which

could easily lead to variability and disputes with respect to admissibility; consensus on approved methods, reference populations and reporting formats would legitimize the field (Jessen 2008).

The third lesson: The examiner, Some sections of the exam, namely grasping sonic detail, will be subject to interpretation with a biased lens. Here, as in any scientific forensic discipline, are the mere defences of blind procedures, documented reasons for decision and division between analysis and knowledge of case equally relevant. Fourth — it is about limits but that arguably is more advisory in nature. When an expert exaggerates the power of voice evidence, she causes a wrongful conviction; when an expert diminishes it, she lets the guilty go free — and because voice evidence doth convict and doth exonerate. The corrective is not caution in the absence of evidence but calibration — erring on the side of making claims proportionately to the evidence (and neither larger nor smaller).

9. Limitations and Directions for Research

This has limits that are common to this type of research. As it was a narrative conceptual review, it has a more interpretive than exhaustive sourcing; different evidence could emerge with systematic search that would change the emphasis. In the rarity of LRE, it's said for and not shown (LRE itself has argued — so proof pending). The framework also simplifies. They are not sharply defined, the cues are correlated, and "robustness" and "typicality" must be computed quantities that all too often fail to come up roses. This all begs the question, maybe they should slow down the framework and test it?

You can play with it, and much of what you need to make the device exists in the fields already. This allows for the likelihood-ratio computation of acoustic, phonological and/or lexical features in both databases to investigate how cues from independent levels are combined (P1). An alternative testing approach for P3 and P4 consists of gradually degrading and concealing a controlled material, allowing to trace the level-level and across-speaker-sex-and-age cue stabilization and collapse (González Hautamäki et al. 2019; Farrús 2018). P2 and P5 reference the simulations of the calibration studies that measure how well categories probabilistic judgments follow ground truth for realistic case conditions (Morrison, 2009). This work would move the field from merely describing what cues convey information to measuring how much, under which conditions, and associated with how much uncertainty — just the kind of input necessary for a court.

10. Conclusion

In this review, we tried to make a description rather than survey what really makes forensic linguistics shine for speaker identification. Real, but conditional contribution. It retains information on personhood and that diffusion is auditory & phonetic (voice) & verbal (style), which human or automated methods can retrieval. But the individual-specificity is neither as unique as a fingerprint, nor stable to mockery and impaired capture, and its usefulness in any given application depends on both how reliable the cues are and how ubiquitous they are in the target population.

The LRE framework brings the judgements together, and the propositions apply pressure to the field providing something against which to test. Crucially, the alarm bell rung by the discipline (and often misread) since voiceprint got abused has to do less with weak measurement than with weak evaluation, confident conclusions — not supported by the data. Every report, every courtroom must insist on evaluation that spells out the strength of the evidence and just as clearly its limitations.

References

Aitken, C. G. G., & Taroni, F. (2004). Statistics and the evaluation of evidence for forensic scientists (2nd ed.). Wiley.

Ariani, M. G., Sajedi, F., & Sajedi, M. (2014). Forensic linguistics: A brief overview of the key elements. Procedia – Social and Behavioral Sciences, 158, 222–225. https://doi.org/10.1016/j.sbspro.2014.12.078

Bai, Z., & Zhang, X.-L. (2021). Speaker recognition based on deep learning: An overview. Neural Networks, 140, 65–99. https://doi.org/10.1016/j.neunet.2021.03.004

Boë, L.-J. (2000). Forensic voice identification in France. Speech Communication, 31(2–3), 205–224. https://doi.org/10.1016/S0167-6393(99)00079-5

Bolt, R. H., Cooper, F. S., David, E. E., Denes, P. B., Pickett, J. M., & Stevens, K. N. (1970). Speaker identification by speech spectrograms: A scientists' view of its reliability for legal purposes. The Journal of the Acoustical Society of America, 47(2B), 597–612. https://doi.org/10.1121/1.1911935

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.

Bonastre, J.-F., Bimbot, F., Boë, L.-J., Campbell, J. P., Reynolds, D. A., & Magrin-Chagnolleau, I. (2003). Person authentication by voice: A need for caution. In Proceedings of Eurospeech 2003 (pp. 33–36). ISCA.

Champod, C., & Meuwly, D. (2000). The inference of identity in forensic speaker recognition. Speech Communication, 31(2–3), 193–203. https://doi.org/10.1016/S0167-6393(99)00078-3

Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., Li, J., Kanda, N., Yoshioka, T., Xiao, X., Wu, J., Zhou, L., Ren, S., Qian, Y., Qian, Y., Wu, J., Zeng, M., Yu, X., & Wei, F. (2022). WavLM: Large-scale self-supervised pre-training for full stack speech processing. IEEE Journal of Selected Topics in Signal Processing, 16(6), 1505–1518. https://doi.org/10.1109/JSTSP.2022.3188113

Clark, J., & Foulkes, P. (2007). Identification of voices in electronically disguised speech. International Journal of Speech, Language and the Law, 14(2), 195–221. https://doi.org/10.1558/ijsll.v14i2.195

Correa, M. (2013). Forensic linguistics: An overview of the intersection and interaction of language and law. Studies About Languages, 23, 5–13. https://doi.org/10.5755/j01.sal.0.23.4939

Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. Applied Linguistics, 25(4), 431–447. https://doi.org/10.1093/applin/25.4.431

Drygajlo, A., Meuwly, D., & Alexander, A. (2003). Statistical methods and Bayesian interpretation of evidence in forensic automatic speaker recognition. In Proceedings of Eurospeech 2003 (pp. 689–692). ISCA.

Durán, J. M., van der Vloed, D., Ruifrok, A., & Ypma, R. J. F. (2024). From understanding to justifying: Computational reliabilism for AI-based forensic evidence evaluation. Forensic Science International: Synergy, 9, Article 100554. https://doi.org/10.1016/j.fsisyn.2024.100554

Farrús, M. (2018). Voice disguise in automatic speaker recognition. ACM Computing Surveys, 51(4), Article 68, 1–22. https://doi.org/10.1145/3195832

Farrús, M., Wagner, M., Erro, D., & Hernando, J. (2010). Automatic speaker recognition as a measurement of voice imitation and conversion. International Journal of Speech, Language and the Law, 17(1), 119–142. https://doi.org/10.1558/ijsll.v17i1.119

Gibbons, J. (2003). Forensic linguistics: An introduction to language in the justice system. Blackwell.

González Hautamäki, R., Hautamäki, V., & Kinnunen, T. (2019). On the limits of automatic speaker verification: Explaining degraded recognizer scores through acoustic changes resulting from voice disguise. The Journal of the Acoustical Society of America, 146(1), 693–704. https://doi.org/10.1121/1.5119240

Grant, T. (2007). Quantifying evidence in forensic authorship analysis. International Journal of Speech, Language and the Law, 14(1), 1–25. https://doi.org/10.1558/ijsll.v14i1.1

Gupta, R. (2019). Forensic analysis of characteristic features of disguised writing in fixing authorship. IP International Journal of Forensic Medicine and Toxicological Sciences, 4(1), 22–27. https://doi.org/10.18231/j.ijfmts.2019.006

Hautamäki, R. G., Sahidullah, M., Hautamäki, V., & Kinnunen, T. (2017). Acoustical and perceptual study of voice disguise by age modification in speaker verification. Speech Communication, 95, 1–15. https://doi.org/10.1016/j.specom.2017.10.002

Huang, B., Chen, C., & Shu, K. (2024). Authorship attribution in the era of large language models: Problems, methodologies, and challenges. ACM SIGKDD Explorations Newsletter, 26(2), 21–43. https://doi.org/10.1145/3715073.3715076

Jessen, M. (2008). Forensic phonetics. Language and Linguistics Compass, 2(4), 671–711. https://doi.org/10.1111/j.1749-818X.2008.00066.x

Kersta, L. G. (1962). Voiceprint identification. Nature, 196(4861), 1253–1257. https://doi.org/10.1038/1961253a0

Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. Journal of the American Society for Information Science and Technology, 60(1), 9–26. https://doi.org/10.1002/asi.20961

Künzel, H. J., Gonzalez-Rodriguez, J., & Ortega-García, J. (2004). Effect of voice disguise on the performance of a forensic automatic speaker recognition system. In Proceedings of ODYSSEY 2004: The Speaker and Language Recognition Workshop (pp. 153–156). ISCA.

Liang, H., Lin, X., Zhang, Q., & Kang, X. (2017). Recognition of spoofed voice using convolutional neural networks. In 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp. 293–297). IEEE. https://doi.org/10.1109/GlobalSIP.2017.8308651

McMenamin, G. R. (2002). Forensic linguistics: Advances in forensic stylistics. CRC Press.

Morrison, G. S. (2009). Forensic voice comparison and the paradigm shift. Science & Justice, 49(4), 298–308. https://doi.org/10.1016/j.scijus.2009.09.002

Morrison, G. S., Enzinger, E., Hughes, V., Jessen, M., Meuwly, D., Neumann, C., Planting, S., Thompson, W. C., van der Vloed, D., Ypma, R. J. F., & Zhang, C. (2021). Consensus on validation of forensic voice comparison. Science & Justice, 61(3), 299–309. https://doi.org/10.1016/j.scijus.2021.02.002

Nolan, F. (1997). Speaker recognition and forensic phonetics. In W. J. Hardcastle & J. Laver (Eds.), The handbook of phonetic sciences (pp. 744–767). Blackwell.

Nolan, F., & Oh, T. (1996). Identical twins, different voices. Forensic Linguistics (International Journal of Speech, Language and the Law), 3(1), 39–49. https://doi.org/10.1558/ijsll.v3i1.39

Prasad, S., & Prasad, R. (2019). Fusion multistyle training for speaker identification of disguised speech. Wireless Personal Communications, 104(3), 895–905. https://doi.org/10.1007/s11277-018-6058-x

Reynolds, D. A. (2002). An overview of automatic speaker recognition technology. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (Vol. 4, pp. 4072–4075). IEEE. https://doi.org/10.1109/ICASSP.2002.5745552

Reynolds, D. A., Andrews, W. D., Campbell, J. P., Navrátil, J., Peskin, B., Adami, A., Jin, Q., Klusáček, D., Abramson, J., Mihaescu, R., Godfrey, J., Jones, D., & Xiang, B. (2003). The SuperSID project: Exploiting high-level information for high-accuracy speaker recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (Vol. 4, pp. 784–787). IEEE. https://doi.org/10.1109/ICASSP.2003.1202924

Rose, P. (2002). Forensic speaker identification. Taylor & Francis.

San Segundo, E., Alves, H., & Trinidad, M. F. (2013). CIVIL corpus: Voice quality for speaker forensic comparison. Procedia – Social and Behavioral Sciences, 95, 587–593. https://doi.org/10.1016/j.sbspro.2013.10.686

Schilling, N., & Marsters, A. (2015). Unmasking identity: Speaker profiling for forensic linguistic purposes. Annual Review of Applied Linguistics, 35, 195–214. https://doi.org/10.1017/S0267190514000282

Snyder, D., Garcia-Romero, D., & Povey, D. (2015). Time delay deep neural network-based universal background models for speaker recognition. In 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU) (pp. 92–97). IEEE. https://doi.org/10.1109/ASRU.2015.7404779

Tavi, L., Alumäe, T., & Werner, S. (2019). Recognition of creaky voice from emergency calls. In Proceedings of Interspeech 2019 (pp. 1990–1994). ISCA. https://doi.org/10.21437/Interspeech.2019-1253

Villalba, J., Chen, N., Snyder, D., Garcia-Romero, D., McCree, A., Sell, G., Borgstrom, J., García-Perera, L. P., Richardson, F., Dehak, R., Torres-Carrasquillo, P. A., & Dehak, N. (2020). State-of-the-art speaker recognition with neural network embeddings in NIST SRE18 and Speakers in the Wild evaluations. Computer Speech & Language, 60, Article 101026. https://doi.org/10.1016/j.csl.2019.101026

Wu, H., Wang, Y., & Huang, J. (2014). Identification of electronic disguised voices. IEEE Transactions on Information Forensics and Security, 9(3), 489–500. https://doi.org/10.1109/TIFS.2014.2301912

Yamagishi, J., Wang, X., Todisco, M., Sahidullah, M., Patino, J., Nautsch, A., Liu, X., Lee, K. A., Kinnunen, T., Evans, N., & Delgado, H. (2021). ASVspoof 2021: Accelerating progress in spoofed and deepfake speech detection. In Proceedings of the ASVspoof 2021 Workshop (pp. 47–54). ISCA. https://doi.org/10.21437/ASVspoof.2021-8

Ye, Y., Lao, L., Yan, D., & Wang, R. (2020). Identification of weakly pitch-shifted voice based on convolutional neural network. International Journal of Digital Multimedia Broadcasting, 2020, Article 8927031, 1–10. https://doi.org/10.1155/2020/8927031

Zhang, C. (2012). Acoustic analysis of disguised voices with raised and lowered pitch. In 2012 8th International Symposium on Chinese Spoken Language Processing (pp. 353–357). IEEE. https://doi.org/10.1109/ISCSLP.2012.6423510

Zhang, C., & Tan, T. (2008). Voice disguise and automatic speaker recognition. Forensic Science International, 175(2–3), 118–122. https://doi.org/10.1016/j.forsciint.2007.05.019