Judges and prosecutors tend to trust blindly forensic evidence, but this is often misguided. In this piece of investigative journalism, EuroScientist looks at the case of speech forensics. In this field, charlatanism, the lack of regulations and controversies within the scientific community sometimes act together to the detriment of justice. Further validations of the methods used in speech forensics—which are much less reliable than DNA profiling or fingerprint testing—have yet to be established. Until then, experts warn, caution is in order.

Speech forensics: when Hollywood seldom mirrors real-life court cases

Forensic speech science lacks regulation and consensus among experts, leaving the door open to charlatans

Ever heard of the CSI effect? The mistaken expectation by judges, lawyers and prosecutors that forensics provides a superior kind of proof, as portrayed in the TV shows CSI, Bones, or Criminal minds. Such fictionalised accounts sometimes feature speech forensics. Together with other forensic disciplines—including hair and footprint analysis and forensic entomology—these are much less well-established than DNA and fingerprint tests. Yet, they are frequently used. Although there is a dearth of official data on their popularity, speech forensics methods are used in what experts estimate to be a few hundreds forensic reports per year throughout Europe.
Here, *EuroScientist* investigates what happens when shaky forensic methods are employed to provide evidence in courts, sometimes leading to charlatanism. Their use has previously led to severe *miscarriages of justice*. This article points to deep scientific controversies in the field and to the need for greater validation and regulations of speech forensics methods. Speech forensics could also be the test case needed to impulse greater rigour in the application of wider forensics methods in judicial courts.

**Huge gap between fiction and reality**

In the US thriller *Clear and Present Danger*, a forensic expert is presented with an eight syllable-long recorded sound bite. That is enough for him to conclude that the voice belongs to "a Cuban, aged 89, educated in the United States . . . Eastern United States." Then he uploads the sample in a supercomputer. Within seconds, the machine compares it with a recording of a suspect, providing a 65% probability of a match.

This is a quintessential CSI effect-sequence: it contains several mistaken assumptions on the power of an expert’s ear, the way the computer should be used and the kind of evidence speech forensics can actually provide. Typically, speech forensics reports rely on speakers’ comparisons and profiling, voice line-ups, transcriptions, and recordings authentications. They all constitute a set of challenging tasks, often performed on noisy samples of just a few seconds, available in a low-quality recording of a telephone conversation.

In real-world trials, speech forensics have sometimes led to misleading results that clearly depart from the idealised Hollywood account. To date, there are more than 20 known cases of this kind—more than a dozen took place in Europe. For example, in 2011, Óscar Sánchez, who was then working in a carwash in Spain, received a 14 years sentence in Italy. Expert Roberto Porto recognised his voice in the wire tapped calls of a drug trafficker. Sánchez spent two years in jail before being cleared. It took a few other experts to point out that the trafficker spoke a Latin American variant of Spanish.

Another example is the case of a Dutch expert who in 2012 identified a speaker in a threatening anonymous voicemail by merely comparing it to a video in which the person was acting dressed as a pastor. In 2006, expert Sameh Rahman processed with his own software the recording of an allegedly racist attack in Germany. The program identified just two voices in the recording, while everyone else in the court could clearly hear three.

**Lack of scientific consensus**

The main challenge in speech forensics is that a voice is "not as stable and distinctive as a fingerprint," says Juana Gil, a researcher in forensic linguistics at the Spanish Superior Council for Scientific Research (CSIC) in Madrid. "When you are discussing with your partner, your voice is completely different than when you are making jokes with your baby," she says. Besides, all educated smokers of 40 in a given city perhaps have similar voice features, she points out. This is why many experts prefer to talk about compatibility rather than identification.
A lot of the work in the field focuses on identifying speech similarities. However, the field is plagued by open scientific controversies. There are two opposing schools of thoughts, according to Gil. Linguists support the supervised use of software, but rely heavily on human interpretation. Audio engineers, on the contrary, give more importance to automated systems. (See Box below).

**Credibility issue**

Beyond expert debates, in some cases the field of speech forensics can suffer from low credibility due to the use of primitive methods, that have already been scientifically discredited. For instance, according to a new Interpol survey of respondents from international law enforcement agencies, about to be published, speech forensics routinely rely on voice recognition techniques based on the spectrographic methods. These are variants of the voiceprint technique discredited in the past due to its lack of reliability. But other kinds of spectrographic techniques are still widely used among all the software-based methods.

A trickle of papers has, in the past few years, called for caution using such primitive approaches. A 1999 petition by the French Acoustic Society (SFA) even went as far as asking to refrain from using speech forensics altogether, on the basis that such methods were not fully validated. This happened after a 30 year old basque independence militant, called Jérôme Prieto was wrongfully jailed for 10 months in 1998. His conviction resulted from a controversial police report identifying Prieto's voice from a phone call claiming the responsibility for the attack.

This is unlikely to be an isolated case. The Interpol survey has revealed that 15 out of 22 respondents from European law enforcement agencies use questionable voice recognition techniques. For example, in several instances agencies relied on the auditory method, akin to "critical hearing" voice samples. The trouble is that "simple hearing can be deceptive," according to Helen Fraser, an Australian forensic phonetic consultant based in Sydney.

No better proof are mistakes in transcription. Fraser, for instance, published in 2014 an experiment where she used recorded and transcribed evidence from a 2008 murder case in Australia. She found background knowledge of a case can dramatically increase listeners' acceptance of a police transcript, even when the transcript is manifestly inaccurate. "It's a phenomenon called priming: once you have heard something, it's difficult to unhear it," says Fraser.

**Waiting for a paradigm shift**

Expectations that experts rely on deeper levels of speech analysis are also increasingly coming to the fore. Indeed, it is no longer enough to find that the suspect's voice has features very similar to that of a criminal. Instead, any piece of forensic evidence has to be placed into its wider context by taking into account how typical that evidence may be.

This is a paradigm shift sweeping the entire forensics field. Take the following example, says Geoffrey Stewart Morrison, one of the advocates of the new paradigm, who is co-authors of the forthcoming Interpol study: "A size 43 shoe is found at the crime scene and the suspects wears size 43 shoes. In another case, a size 48 is found and the suspect wears size 48." The key is in how typical one piece of evidence is compared to another. "In the second case the evidence against the suspect
is stronger than in the first one, because size 48 is much rarer than size 43," adds Stewart Morrison, who is an adjunct associate professor of linguistics at the University of Alberta, Canada, and a forensic consultant.

To encompass the notion related to how typical a given piece of evidence may be, experts can rely on available statistical analysis tools, called likelihood ratio, which encapsulate both speech similarity and typicality. As recently as June 2015, the European Network of Forensic Sciences Institutes (ENFSI) published a [Guideline for Evaluative Reporting in Forensic Science](#) recommending the use of such tool. However, 10 of the 22 respondents of the Interpol's survey never followed this recommendation.

**Unregulated disciplines**

To establish reliable speech forensics evidence is challenging. The lack of scientific consensus over methods in the field is compounded by a dearth of regulations at all levels.

This affects, in particular, the selection of experts themselves. For example, "there is no law that establishes the requirements for a linguistic expert, unlike a medical expert who must have specific training," explains Jordi Cicres, a researcher in forensic linguistics at the University of Girona, Spain.

In parallel, the nascent European regulations are not yet fully comprehensive when it comes to establishing standards of the evidence used in court. It is only since 2011 that the European Commission approved the creation of a European Forensic Sciences Area due to be established by 2020 ([EFSA2020](#)). “The idea was to extend DNA and fingerprints safeguards to all forensic fields, speech included”, says Paweł Rybicki, former director of the Polish Central Forensic Laboratory of the Police ([CLKP](#)), who drafted the Commission’s conclusions. But not much followed, as EFSA2020 remains “under evaluation” according to the Commission.

Regulation would clearly help to stir charlatans away from courts. It could also restrict the judicial application to the most established parts of what remains an immature science. Meanwhile, regulatory pressure could also contribute to drive further scientific validation of the main speech forensics methods in use today. Lessons learned from this investigation show that the real CSI remains a cumbersome enterprise, far from the shiny television representation.

**Michele Catanzaro** is an Italian science journalist based in Barcelona, Spain.

**Astrid Viciano** is a German science journalist based in Paris, France.

**Philipp Hummel** is a German science journalist based in Berlin, Germany.

**Elisabetta Tola** is an Italian science journalist based in Bologna, Italy.

---

**Scientific divide on voice recognition techniques**

The most popular technique in forensic speech science, according to a [2011 survey](#) and the forthcoming Interpol study, is the acoustic-phonetic method. In that approach, especially preferred
by linguists, experts hear voice samples, select specific fragments—vowels, for example—and measure acoustic parameters with a dedicated software.

Another increasingly popular approach - the favourite of engineers - is called Automated Speaker Recognition (ASR). It involves computer-based extraction of features from speech signals and calculation of parameters like the cepstral coefficients, which are considered as related to the unique properties of each vocal tract. "What we do [with ASR] is 100% different from what linguists do," says Antonio Moreno, vice-director of Agnitio, a company producing Batvox, the most popular ASR software according to the Interpol’s survey. He adds: "Our system is much more precise, reproducible under equal conditions and with a measurable precision."

However, Juana Gil, a researcher in forensic linguistics at the Spanish Superior Council for Scientific Research (CSIC) points out that she has “compared voices that the machine considered compatible [and has] found an awful lot of differences: from accent to a disphony, because the speaker had a cold.” Other experts point out that users should understand how ASR works before employing it as a black box. “These products are like a plane: you can buy it in a day but you will not learn to fly in three weeks,” points out Didier Meuwly, of the Netherlands Forensic Institute. Agnitio offers a 3 years course to use Batvox, but its director Emilio Martínez admits that only about 20% of the users complete it.

Photo credit: Gianluca Battista.

This article was developed with support by Journalismfund.eu