languages

MDPI

Article

Forensic Audio and Voice Analysis: TV Series Reinforce False Popular Beliefs

Emmanuel Ferragne [1,*], Anne Guyot Talbot [1], Margaux Cecchini [1], Martine Beugnet [2], Emmanuelle Delanoë-Brun [2], Laurianne Georgeton [3], Christophe Stécoli [3], Jean-François Bonastre [4] and Corinne Fredouille [4]

1 Laboratoire CLILLAC-ARP, UFR d'Études Anglophones, Faculté Sociétés et Humanités, Université Paris Cité, 75013 Paris, France; anne.talbot@u-paris.fr (A.G.T.); margaux.cecchini@etu.u-paris.fr (M.C.)
2 Laboratoire LARCA, UFR d'Études Anglophones, Faculté Sociétés et Humanités, Université Paris Cité, 75013 Paris, France; martine.beugnet@u-paris.fr (M.B.); delanoee@u-paris.fr (E.D.-B.)
3 Laboratoire Central de Criminalistique Numérique, Service National de Police Scientifique (SNPS), 69130 Écully, France
4 Laboratoire Informatique d'Avignon, Avignon Université, 84000 Avignon, France; jean-francois.bonastre@univ-avignon.fr (J.-F.B.); corinne.fredouille@univ-avignon.fr (C.F.)
* Correspondence: emmanuel.ferragne@u-paris.fr

Abstract: People's perception of forensic evidence is greatly influenced by crime TV series. The analysis of the human voice is no exception. However, unlike fingerprints—with which fiction and popular beliefs draw an incorrect parallel—the human voice varies according to many factors, can be altered deliberately, and its potential uniqueness has yet to be proven. Starting with a cursory examination of landmarks in forensic voice analysis that exemplify how the voiceprint fallacy came about and why people think they can recognize people's voices, we then provide a thorough inspection of over 100 excerpts from TV series. Through this analysis, we seek to characterize the narrative and aesthetic processes that fashion our perception of scientific evidence when it comes to identifying somebody based on voice analysis. These processes converge to exaggerate the reliability of forensic voice analysis. We complement our examination with plausibility ratings of a subset of excerpts. We claim that these biased representations have led to a situation where, even today, one of the main challenges faced by forensic voice specialists is to convince trial jurors, judges, lawyers, and police officers that forensic voice comparison can by no means give the sort of straightforward answers that fingerprints or DNA permit.

Keywords: forensic phonetics; speaker identification; voice comparison; TV series

Citation: Ferragne, Emmanuel, Anne Guyot Talbot, Margaux Cecchini, Martine Beugnet, Emmanuelle Delanoë-Brun, Laurianne Georgeton, Christophe Stécoli, Jean-François Bonastre, and Corinne Fredouille. 2024. Forensic Audio and Voice Analysis: TV Series Reinforce False Popular Beliefs. *Languages* 9: 55. https://doi.org/10.3390/languages9020055

Academic Editors: Julien Longhi and Nadia Makouar

Received: 10 December 2023
Revised: 22 January 2024
Accepted: 24 January 2024
Published: 2 February 2024

1. Introduction

1.1. Background and Goals

Although firm quantitative evidence is still sparse (Eatley et al. 2018), it is often assumed that the many popular TV shows revolving around forensic science that emerged in the early 2000s (e.g., the various versions of the *Crime Scene Investigation* (*CSI*) franchise) have triggered what is now known as the "CSI Effect". To give but one example supporting this assumption, Call et al. (2013) surveyed 60 jurors from five malicious wounding juries in the United States. Their findings show that 95% of the jurors watched *CSI*, and 73% considered that the series influenced their verdict. The CSI effect is characterized by at least two phenomena (Eatley et al. 2018): (i) jurors in trials now tend to have unrealistic expectations regarding the presence of scientific evidence, and (ii) when scientific evidence is available, they are more prone to view it as infallible. This effect possibly extends beyond the general public and is thought to affect professionals (Call et al. 2013; Trainum 2019) and even criminals (Baranowski et al. 2018).

One of our central claims in the current article is that forensic voice analysis lends itself particularly well to misconceptions, which are, in turn, promoted by TV crime shows.

What is it that makes forensic voice analysis so different from other biometric evidence like fingerprints or DNA? Firstly, instances of forensic voice analyses have been comparatively sporadic in France: over the last 12 years, *Service National de Police Scientifique* (SNPS: national police forensic department) has performed 233 such analyses, while in August 2022 only, they carried out as many as 10,000 fingerprint and 7000 DNA analyses. Consequently, most courts have little to no experience in voice analysis, and fictional representations may sometimes be their only available reference. Secondly, while most people have never performed fingerprint or DNA analyses, they have constantly relied on their ears to identify people around them. Therefore, we all have strong intuitions based on experience, and we may have implicitly made the wrong generalization that auditory speaker identification is reliable under all circumstances. Thirdly, the analogy with fingerprints, which is constantly reinforced by works of fiction and the media, is deeply rooted in people's minds to the extent that they forget how variable and alterable voices can be.

The goal of this article is to analyze the representation of forensic voice analysis in a set of popular TV procedurals that have been broadcast in France. We focus on English-speaking, mainly US-based, shows since the latter are ubiquitous on French TV. We follow a mixed-methods approach combining qualitative analyses partially inspired by the field of visual studies with quantitative data to support our findings. A small-scale experiment involving plausibility ratings complements the analysis. The originality of our approach is strengthened by the diversity of the authors' backgrounds: there are three phoneticians and two specialists of speech processing who have been involved to varying degrees in forensic voice analysis over the last 5 (for the relative newcomers) to 30 years. Two authors specialize in visual studies, especially cinema and television studies, with one of them focusing on crime TV shows. The two forensic scientists from the audio section of SNPS are currently the only two scientists who perform forensic voice analysis in the police force in France, and they, therefore, provide first-hand experience and knowledge. In the fields of phonetics and voice processing, our areas of specialization cover a broad spectrum that encompasses most (if not all) fields related to forensic audio: perceptual, acoustic, and articulatory phonetics, automatic speech processing and recognition, and speaker identification. The next section briefly reviews historical landmarks that illustrate the two popular beliefs on which the article focuses: human listeners' ability to identify people by their voices and the idea that voices are as unique and unalterable as fingerprints (the voiceprint fallacy). Then, we briefly describe the specific context of this study: forensic voice analysis in France after a more thorough review of the situation at the international level. The rest of the article shows how we collected and annotated our data, and the type of qualitative and quantitative analyses that were performed.

1.2. Some Landmarks in Forensic Voice Analysis

About 90 years ago, Bruno Hauptmann was executed in the US for kidnapping and murdering Charles Lindbergh's two-year-old son (Solan and Tiersma 2003). Three years earlier, as Lindbergh and another man were delivering the ransom to the kidnappers in a cemetery, Lindbergh, who had remained in the car 60 to 100 m away, overheard the words "Hey, doctor. Here, doctor, over here" spoken with a German accent. Over two years later, Lindbergh confessed that it would be very difficult for him to recognize the man by his voice. Then, a district attorney asked Lindbergh if he wanted to see the man who killed his son; Hauptmann was brought in and asked to pronounce "Hey, doctor. Here, doctor, over here". Lindbergh said he recognized the voice he had heard in the cemetery. This case has remained controversial and has triggered a host of research on the many limitations of earwitness identification and voice parade methodology (see, e.g., Humble et al. 2022; McDougall et al. 2016; Yarmey et al. 1994), starting with McGehee (1937).

Thirty years after the controversial outcome of the Lindbergh case following questionable auditory voice identification methods, science came to the rescue: in an article entitled *Voiceprint identification*, Kersta (1962) suggested that one can identify people with a spectrographic analysis of their voice, and explained that very much like "people's finger-

prints, voiceprint identification uses the unique features found in their utterances". The original voiceprint method involved the visual analysis of the overall shape of spectrograms of ten frequent English words. Incidentally, Kersta's spectrograms came in two varieties: the broadband spectrogram, which is today's standard for phonetic description, and a contour plot that had better amplitude resolution than the other method at that time. Quite interestingly, contour plots are evocative of fingerprint ridge patterns, which may have strengthened the analogy between voice and fingerprints when Kersta's research was published.

In 2002, Elodie Kulik was raped and murdered after being involved in what seems to be a deliberate car accident caused by another car. Just after the crash, Elodie Kulik called the emergency services. On the bad-quality 26-s recording of this call, during which she realizes that the people who came to her were not here to help her, at least two male voices can be heard. Once the first man, who had died shortly after the events, was formally identified thanks to DNA evidence, a friend of his, Willy Bardon, was arrested as the potential second man. Bardon was eventually sentenced to 30 years of prison in 2021. One key aspect of this case is that some of Bardon's friends and relatives thought they had recognized his voice on the recording. His nephew says the voice on the recording displays "intonations" that sound like his uncle's. Even Bardon himself concurs: "It's my voice, it sounds like my voice; but I wasn't there." (Guiho 2020).

The Lindbergh and Kulik cases illustrate that while 90 years of research into our ability to auditorily identify someone by their voice have furthered our understanding of the limitations of such approaches, asking people if they can recognize someone's voice has remained a classic of police interviews and trial testimonies. The collective belief that people can definitely identify others by listening to their voice is firmly established, and for good reason: we rely on our ears to identify (or confirm the identity of) people around us on a daily basis.

Although Kersta's claims were soon disproven (Bolt et al. 1969), the voiceprint fallacy has lingered ever since then. For the anecdote, even the French researchers and engineers who worked with the Voice Identification Inc., *Sound Spectrograph*, *model 700* (before the generalization of personal computers for speech analysis) would borrow the English word and call it "le voiceprint". Note that it was only in 2007 that the International Association for Forensic Phonetics & Acoustics passed a resolution banning the use of the voiceprint approach.

1.3. Forensic Voice Analysis and Voice Comparison in the World Today

Forensic voice analysis involves a number of disciplines and tasks (signal processing, acoustic and perceptual phonetics, transcription of what is being said, etc.), and a comprehensive state-of-the-art section is well beyond the scope of the current article (see, e.g., De Jong-Lendle 2022; Hudson et al. 2021; Morrison and Thompson 2017; Watt and Brown 2020). However, some basic knowledge is needed so that readers can more fully appreciate the similarities and divergences between reality and fiction. One particular subfield that has generated much scientific debate and constitutes perhaps the most frequent type of analysis is voice comparison. The aim is to compare (at least) two recordings and determine how likely it is that they were spoken by the same person. There are four different methods, according to Morrison and Thompson (2017): auditory, spectrographic (this is equivalent to the voiceprint technique), acoustic–phonetic, and automatic. These methods are frequently combined, and some of these generic terms can be split into more fine-grained categories (e.g., acoustic analysis with or without statistical modeling).

The various studies surveying international practices over the past 20 years (e.g., Broeders 2001; Gold and French 2011, 2019; Morrison et al. 2016) show great variation between countries and practitioners. Some of these discrepancies stem from differences in national legal frameworks; others are the result of practitioners' habits and preferences. The evolution in recent years shows a general move towards standardization and methods that more closely match the requirements of general science. For example, the need for

reproducible results has led to the increasing use of automatic methods: between 2011 and 2019, the percentage of forensic experts in the world who use automatic speaker recognition systems went from 17% to over 40% (Gold and French 2019). Following the same paradigm shift towards reinforced scientificity, more and more practitioners use the Bayesian statistical framework and likelihood ratios in their reports rather than binary decisions or probabilities. Awareness of potential cognitive bias is yet another sign of the evolution of forensic science towards more controlled scientific methods (Cooper and Meterko 2019; Gold and French 2019).

However, discrepancies remain. One particular source of variability among the 39 laboratories surveyed in Gold and French (2011) is the number of cases: it ranged from 4 to 6000 with a mean of 506 (compare with France in Section 1.4). Another factor affecting variation between countries is the specific legal framework and, in particular, the admissibility of expert testimony. For instance, in the US, Federal Rule of Evidence 702 and the Daubert standard offer explicit criteria to determine admissibility (Morrison and Thompson 2017), while some other countries do not provide such explicit guidelines.

As far as methods are concerned, and in spite of the general trend towards standardized, thoroughly tested procedures, the latest surveys, by Morrison et al. (2016) and Gold and French (2019), illustrate that no unified methodology has emerged yet. For instance, in Morrison et al. (2016), while the auditory acoustic phonetic method was the preferred approach in Europe, all other possible approaches (including auditory-only and spectrographic) were used.

Other discrepancies between countries or forensic laboratories arise from differences in how they express their conclusions. The most frequent conclusion framework in Gold and French (2019) is the verbal likelihood ratio. It is surprising to note that around 5% of those surveyed still use binary conclusions (the criminal and the suspect are the same person or not).

In a word, it would be difficult to summarize what the state of the art in forensic voice analysis and voice comparison is because (i) practitioners' habits and legal frameworks are quite variable and (ii) the field encompasses many disciplines (e.g., phonetics, psychology, denoising, automatic speaker recognition, etc.) with their own performance tests, internal debates, etc. A recurring question that we are asked very often is how reliable automatic systems are. A reasonable answer is that it depends on the particular conditions. The following common sense example will illustrate this. In a white paper (Nuance Communications 2015), a company specialized in automatic voice authentication (e.g., a customer accesses online services provided by their bank by speaking a password that is then compared to a stored recording of this password by the customer) claims that the system can achieve 99% accuracy. The bank stores a finite set of pre-recorded utterances, the customers were cooperative when they recorded them, they probably make every effort to be as "recognizable" as possible each time they utter their password, and to be intelligible (e.g., they probably speak close to the telephone, at a volume that will not cause distortions, they avoid background noises, etc.). Now consider a realistic forensic scenario: the criminal's voice was captured by a microphone placed at a distance, drowned in background noise, and heavily distorted. Then, the forensic specialist organizes an interview with a suspect in order to record as much spoken material as possible so as to collect a reliable "known" sample for voice comparison. But the suspect will not cooperate and only provides one-word replies, or maybe they deliberately change their voice or their accent. If we then factor in that, contrary to the bank example, the number of "customers" is now potentially unlimited, it is easy to understand that accuracy levels drop dramatically. Morrison and Thompson (2017) argue convincingly that the admissibility of any approach in forensic voice analysis depends on "whether it has been empirically tested under conditions reflecting those of the particular case under investigation, and found to be sufficiently valid and reliable". In other words, reliability and performance should be assessed on a case-by-case basis.

1.4. Forensic Voice Comparison in France

What we describe in this article applies to a particular context: French police officers, forensic scientists, and viewers watching crime dramas shot (mainly) in the US. Background information is, therefore, necessary to clarify the context against which our analysis has been performed. At the time of writing this article, among the French police and gendarmerie forces, only the audio section of SNPS—basically the two co-authors whose affiliation is SNPS—perform forensic audio analysis and voice comparison in France. They use the human-supervised automatic approach complemented by the acoustic–phonetic method. No explicit criteria for the admissibility of evidence in court (like Rule 702 in the US) exist in the French system. The conclusions of an automatic voice comparison are expressed as a verbal likelihood ratio with an interpretative 11-point scale with five degrees reflecting the strength of the difference between the two samples, one neutral degree, and five steps indicating the degree of similarity between the two samples.

Through the Association Francophone de la Communication Parlée (AFCP—French Association for Spoken Communication) and its predecessors, French academics specializing in audio voice processing and phonetics passed a motion in 1990 (and again in 1997 and 2002) warning against the weaknesses involved in trying to identify someone by their voice given the current state of knowledge. In addition, the motion insists that academics should not perform forensic voice comparison, and overall, academics have complied with the injunction ever since it was accepted. This is in stark contrast to international practices: Gold and French (2011), in their survey, had 18 voice forensic practitioners affiliated with universities or research institutes out of their 36 respondents (the number dropped to 8 out of 39 in Gold and French 2019).

The history of forensic voice comparison in France over the past 30 years has been rather tumultuous (Boë 2000; Bonastre 2020). Relationships between academics, forensic scientists from law enforcement agencies, and private labs with self-styled audio experts have been, at times, very tense indeed. With the growing number of projects involving audio scientists from SNPS and academic phoneticians and specialists in signal processing over the last few years, collaborations have become fruitful, which is bound to have positive effects on the whole field.

As noted in the Introduction, not only have cases involving speaker identification been rare (compared to fingerprints or DNA) in France but also cases that have attracted national media attention are few and far between. In the last 15 years or so, the following cases can be mentioned: Benalla-Crase, Cahuzac, Chikli, etc. The iconic Chikli case has inspired movies and documentaries (e.g., Ratliff 2022) and may, therefore, have had a great impact on the public. Gilbert Chikli invented the CEO scam: he would call employees of large companies, pretend he was their CEO, and have money transferred for him. He even went so far as to pass himself off as the French Defense Minister for the same purposes.

We contend that the scarcity of cases of speaker identification that have hit the headlines in France, together with the low number of actual cases (relative to fingerprints or DNA), lead to a situation where judges, lawyers, police officers, and the public are not prepared to deal with such cases and may, therefore, rely on fictional representations.

2. Methods

2.1. Database Collection

The Springfield! Springfield! Website[1]—which was only accessible via the Internet Archive when we carried out this research—contains orthographic transcriptions of the dialogues of several thousands of English-speaking TV series and films. We manually explored the available titles, predominantly those of the crime genre, and selected 50 series that have been broadcast in France on non-encrypted, freely available TV channels. Our initial database comprises 5372 episodes, totaling 26,782,309 words.

We then wrote a MATLAB script to extract concordances targeting the word "voice", along with 80 characters of context before and after it. The 3321 occurrences of "voice" and their immediate lexical context were then individually inspected, and those that

seemed to be related to forensic audio analysis, and in particular to the identification of an individual by their voice, were kept. Some of the 160 passages we had thus identified were excluded after viewing them because they turned out to be irrelevant to our goals. Additionally, two shows, *Drop Dead Diva* and *Law and Order: UK*, were not accessible through the platforms we used: Prime Video and FMovies.

In total, 106 scenes from 28 different series were thus identified and extracted as video files. They were viewed and manually annotated based on several criteria:

- The episode, season number, and year of the first broadcast;
- Did the excerpt involve speaker identification?
- What methods were used?
- Was the analysis auditory and/or supported by visualizations of the signal?
- What type of visualizations were used;
- Was the display actually used or just decorative?
- Whether there was an explicit analogy with DNA;
- Whether there was an explicit analogy with fingerprints

Since this annotation stage was manual, we also collected more descriptive data: portions of dialogues that caught our attention, legitimate and erroneous uses of technical terms, and special descriptions of the graphical interfaces of signal processing programs.

The titles of the 28 TV shows with colors representing the number of episodes from each series in the dataset are shown in Figure 1.

Figure 1. TV show title with colors indicating the number of excerpts.

Figure 2 shows the distribution of excerpts according to year. Most of them were produced during the *CSI* era; the last excerpt in the dataset first aired in 2016. The earliest scene was from an episode of *The Prisoner* from 1967: S01E10T21:57. In our notation, "S" introduces the season, "E", the episode, and "T" is followed by the time stamp corresponding to the beginning of the excerpt. Excerpt duration ranges from 15 s to 3 min and 15 s, with a mean of 59 and a standard deviation of 29 s for a total duration of 1 h and 45 min.

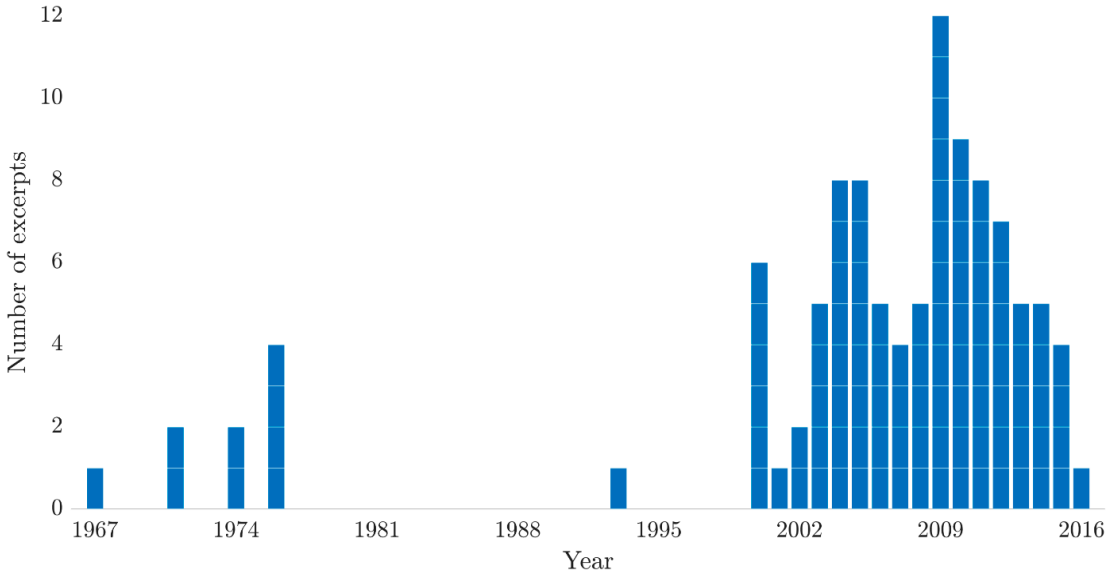

Figure 2. Distribution of excerpts according to year.

The claim that the TV shows we chose to study are omnipresent on French TV is substantiated by Figure 3. The graph shows the number of days when at least one episode of our series was broadcast on French TV between 1 January 2022 and 4 December 2023 (704 days). We wrote a MATLAB script to search a website that stores TV listings[2].

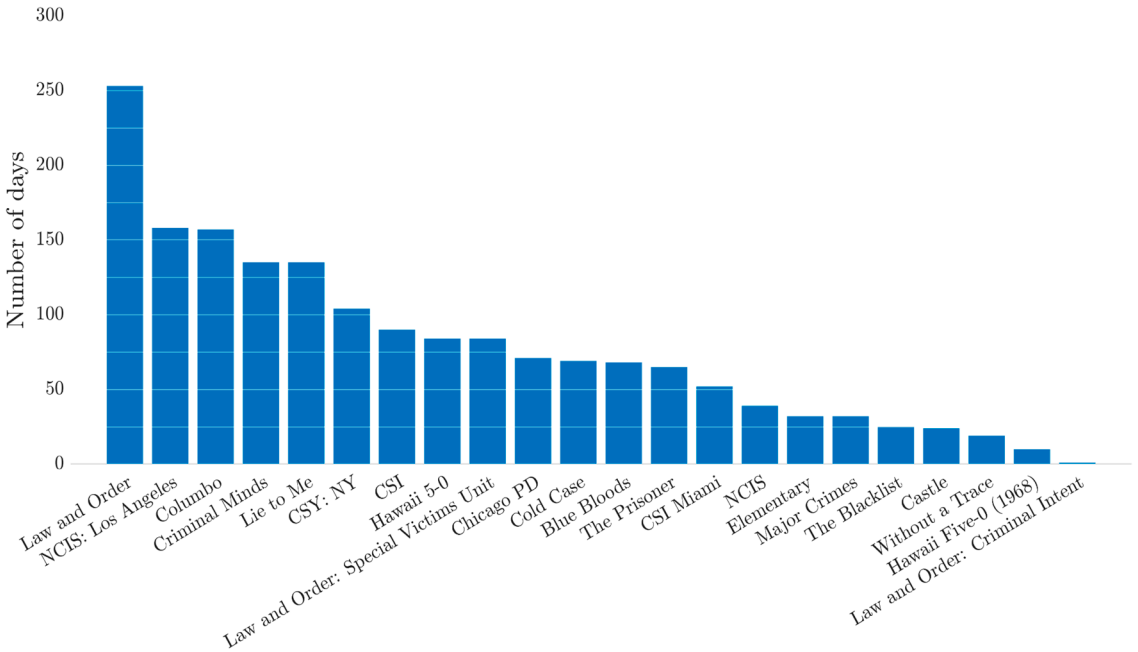

Figure 3. Number of days when at least one episode of the show on the x-axis was broadcast in France between 1 January 2022 and 4 December 2023.

2.2. Plausibility Ratings

The two forensic scientists from SNPS, as well as one speech-processing specialist from our team, rated the plausibility of 41 chosen excerpts that included visualizations of the speech signal. The five-point scale had the following values: 1 = totally unlikely, 2 = rather not likely, 3 = do not know, 4 = rather likely, and 5 = totally plausible. The participants carried out the rating experiment at their own pace, in uncontrolled environments, using a spreadsheet file. They were free to add any comment they deemed relevant.

3. Results

3.1. Database Analysis

Table 1 offers a synoptic view of the main findings resulting from the manual annotation and classification of the video excerpts in our database.

Table 1. Frequency of particular features of interest in our database.

Features of Interest	Number of Excerpts (Out of 106)
Auditory identification of a speaker from an audio recording	22
Comparison of two recordings in order to identify the speaker	42
Audio signal is graphically represented	65
Superimposed waveforms	5
Comparison of two waveforms side by side	4
Decorative waveforms	27
Voice is compared to DNA	0
Voice is compared to fingerprints	3
Analysis of the speaker's accent	7

Voice comparison (here, strictly speaking, the acoustic comparison of two recordings) occurs in 42 excerpts. For 20 of these, the analysis takes place within the excerpt, while the others refer to past or future analyses. Among the different types of methods, we found 22 instances of auditory analyses in which an individual is identified by their voice, with 15 of them occurring within the excerpt. The remaining cases are quite diverse; they involve simply listening to a recording, sometimes employing signal enhancement techniques, filtering, and source separation, particularly to analyze background noise.

An extreme case is depicted in Figure 4: at the top of panel A, the signal comes from one speaker's voice, at the bottom is the voice of another speaker, and between them, the two perfectly identical waveforms are being overlaid in the middle panel of the interface, resulting in a perfect overlap in panel B, with the description: "MATCH 99.675%". All the authors of the current paper agree that this is unrealistic because (i) no matter how hard one may try to produce two identical versions of an utterance, the minute variations in air pressure that are reflected in the waveform are well beyond his or her control, leading to different waveforms, (ii) software would normally output likelihood ratios rather than raw percentages, and (iii) the human eye cannot detect speaker-specific information in such raw representations of the signal as waveforms. In our dataset, the comparison of two identical waveforms occurs on nine occasions: five of them have waveforms displayed side by side, and the remaining four are superimposed. We contend that this deceptive trick is particularly appealing because of its simplicity and its similarity with fingerprint analysis.

We found references to accents or dialects in seven episodes. For example, a voice sample is submitted to "a beta version of Shibboleth, [. . .] an accent identifier", a very aptly denominated fictional software program, in *NCIS: Los Angeles* S02E12T39:50. In *Law and Order: New York* S03E17T29:07 a 911 caller articulates his /t/ in the words *battery* and *city* as plosives rather than the flap consonants that are much more usual in American English. The voice expert in this excerpt regards this as an idiosyncrasy that, taken in conjunction with other similarities, does not constitute a positive ID but a probable match. In *Hawaii Five-0 (1968)* S08E23T18:26, the voice specialist says she "did pick up on a couple of flat A's and a dropped G in the word *tellin*", which leads her to conclude that the caller comes from the South West of the United States. While these examples are not too far-fetched—automatic accent identification in English is very accurate (Zuluaga-Gomez et al. 2023), and auditory accent identification has sometimes been used (e.g., S. Ellis 1994), their forensic value when it comes to identifying a single speaker is limited.

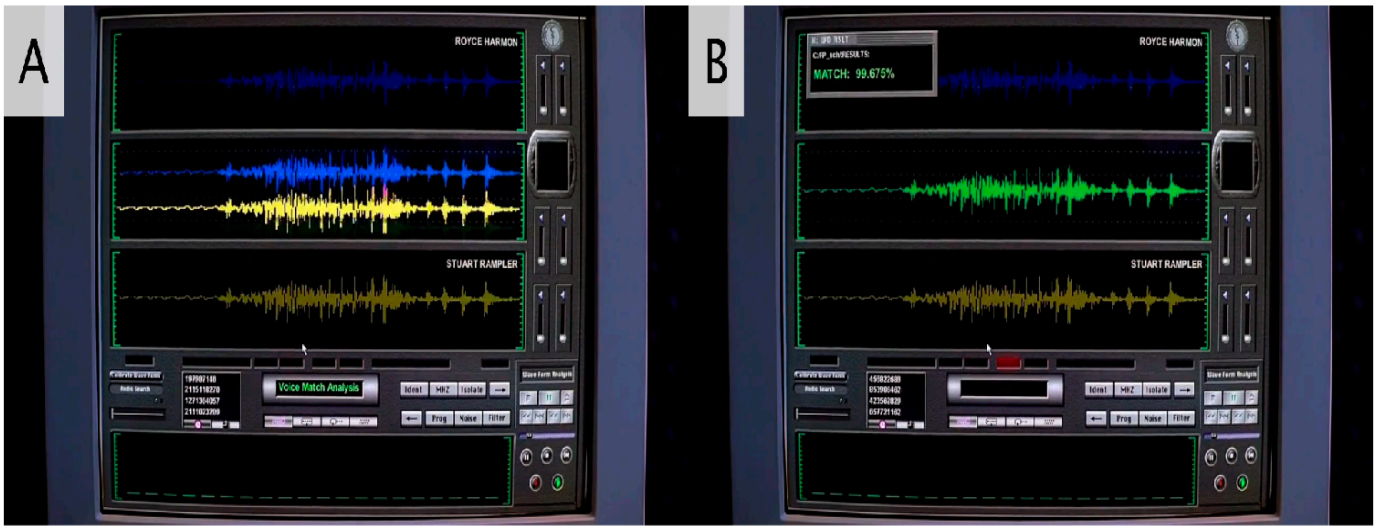

Figure 4. The (unrealistic) classic perfect matching of two waveforms (*CSI*: S01E08T18:39). The identical blue and yellow curves in (**A**) are gradually merging into a green one in (**B**).

Regarding the potential analogy of the voice with biometric data, a possible similarity with DNA is never mentioned in our dataset. An explicit comparison with fingerprints is present in three excerpts. In *The Prisoner* S01E10T21:57, Number Two explains that "Voices are like fingerprints; no two are the same. Even if the voice is disguised, the pattern doesn't change". In *CSI NY* S07E15T24:50, Detective Mac Taylor claims that "voice patterns are as distinct as fingerprints". In *Law and Order: New York* S14E16T29:59, a character says: "I heard your voice on the tape. It's like a fingerprint". These episodes first aired in 1967, 2011, and 2013, respectively, which shows that the use of the term extends well beyond the rejection of the concept by the academic community.

Additionally, there are three excerpts where the term "voiceprint" (with various spellings) is displayed on software graphical user interfaces; two of them are shown in Figure 5. Incidentally, the two screenshots bear a close resemblance to one another because they come from episodes of the same show that came out one year apart. In addition, the superposition of key acoustic cues is arguably reminiscent of a DNA electropherogram. Therefore, although DNA is never mentioned explicitly as an analogy, it is evoked through visual displays.

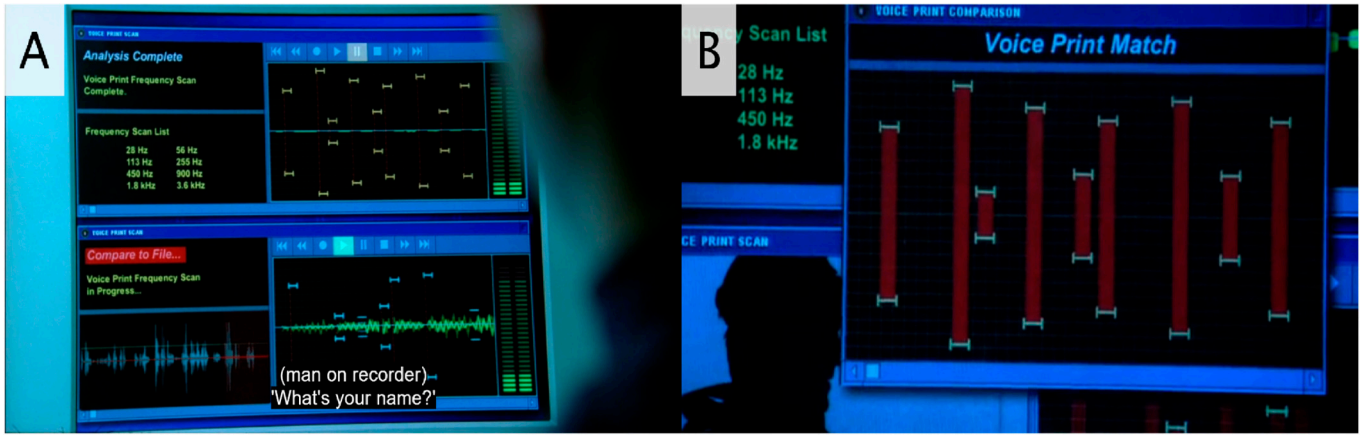

Figure 5. A visual evocative of a DNA electropherogram where "voice print" also appears. ((**A**) *Without a trace* S02E22T05:39; (**B**) *Without a trace* S01E18T29:45).

"Voiceprint" is explicitly mentioned in the dialogues of five excerpts. In *Criminal Minds* S11E09T28:41, Special Agent Jennifer Jareau threatens a woman that voiceprint recognition

will be applied to the recording of a phone call, and the woman immediately (rightly) opposes: "that's not foolproof". This is the only excerpt where a character's awareness of the limitation of the method is clearly stated.

Familiarity with a speaker's voice is sometimes used as a reason supporting speaker identification: in *NCIS* S01E09T16:14, a man claims he recognized, without a doubt, someone's voice on the answering machine because he and the caller have known each other since they "were second lieutenants at the Basic School". Similarly, in *Castle* S02E23T16:26, "even though they wore ski masks, he recognized them by their voices because they grew up together". The familiarity argument is taken a step further in *Criminal Minds* S01E05T15:22 where, as a kidnapper on the phone wants to speak to the twin sister of the woman he has kidnapped, Special Agent Elle Greenaway stands in, but the kidnapper says: "I know her voice therefore I know her sister's. Get off the phone". It is true that the voices of monozygotic twins tend to be more similar than those of genetically unrelated human beings (San Segundo and Künzel 2015). In *Criminal Minds* S03E12T03:29, familiarity with the voice is expected to be a robust predictor of successful identification ("the parents can ID the voice") and again in S04E14T19:24 (from the same show): a reproachful mother complains: "you think I don't know my own daughter's voice?". Our ability to recognize familiar voices has been shaped by evolution, it has reached a high degree of sophistication, and we rely on it to structure the world that surrounds us (Sidtis and Kreiman 2012). There is also compelling evidence that the processing of familiar and unfamiliar voices is distinct (Stevenage 2018). Therefore, in the excerpts we mentioned, scientific evidence supports, to a certain extent, the characters' claim that they can recognize familiar voices. But, as always with forensic voice analysis, this ability comes with a certain error rate that becomes worse as the sample length becomes shorter, its audio quality is more degraded (noisy or over the phone), and it may not be robust to voice disguise.

Visual representations of the audio signal appear in 65 of our excerpts. Fifty-two of these are (or have as their dominant plots) amplitude–time graphs, i.e., waveforms. Five other cases show spectra or spectrograms, and for the remaining eight cases, the display features a combination of graphs and sometimes graphs whose nature is difficult to determine. It appears then that by far the most frequent plot is the waveform, which is, arguably, the least informative type of signal visualization when one is interested in voice and phonetic analysis. A small sample of visual representations of the signal, reflecting, among other things, technological developments in the history of speech analysis on TV, are presented in Figure 6. Panel A shows an oscilloscope that is contemporary with the excerpt (1976); it is used here as a simple visual cue signaling that some audio signal is being played back (very much as modern phone apps would). Panel B shows a software program from 1993, supposedly from the "voice biometrics lab at Georgetown", with an "oscilloscope" (the signal dimension that it displays remains unclear) and a spectrogram. Panel C—which is, incidentally, the third graphical interface with "voiceprint" on it—shows a curious type of speech waveform that is reminiscent of the kind of low-bit-depth signal of electrocardiograms. Panel D displays the state-of-the-art televisual voice analysis gear with touchscreen capabilities and translucent colorful waves that seem to populate the whole room.

The visualizations play various roles. In Figure 7, panel A, the flashy displays in the background are totally independent of the ongoing analysis by the two protagonists. Colorful curves seem to be serving a purely decorative purpose in 27 excerpts. In panel B, what the analyst is looking at on the screen in front of him is duplicated on the huge screens behind him. The waveforms are obviously only intended for the viewers, probably both as a mise en scène ploy to make the passage livelier and as a way to call viewers as witnesses.

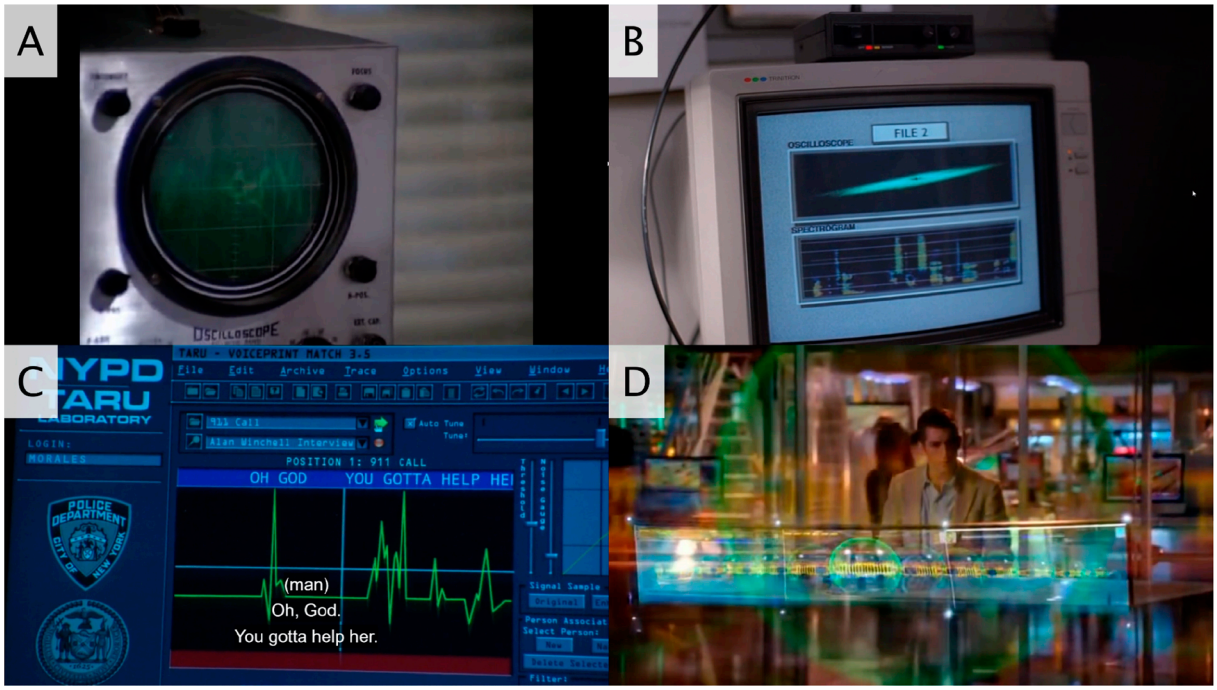

Figure 6. Various types of signal visualizations. (**A**) An oscillogram from *Hawaii Five-0* (1968) S08E23T18:26. (**B**) Oscilloscope and spectrogram from *The X-Files* S01E07T17:38. (**C**) A waveform from *Law and Order: Special Victims Unit* S07E15T03:34. (**D**) Waveforms from *CSI: Miami* S10E08T28:51.

Figure 7. (**A**) *CSI* S15E18T27:40. (**B**) *CSI NY* S05E22T07:15.

3.2. Plausibility Ratings

The distributions of plausibility ratings by rater are shown in Figure 8. The two SNPS forensic scientists in the author list appear as SNPS-1 and SNPS-2, and ACADEMIC is one of the speech-processing specialists from the authors. All distributions are skewed in that they tend to exhibit more low than high scores. A Kruskal–Wallis test shows a difference in median scores among the three raters ($\chi^2 = 6.62$, $p = 0.04$), which, after post hoc comparison, is due to lower scores by SNPS-2 compared to ACADEMIC. Cohen's κ analysis shows that only the judgments of the two SNPS members exhibit significant consistency (SNPS-1~SNPS-2: $\kappa = 0.307$, $p < 0.01$; SNPS-1~ACADEMIC: $\kappa = 0.126$, $p > 0.05$; SNPS-2~ACADEMIC: $\kappa = 0.078$, $p > 0.05$). By-rater mean ratings are ACADEMIC: 2.68, SNPS-2: 1.98, and SNPS-1: 2.44. By-excerpt median scores show that only 8 of them score 4 or 5 (rather likely/totally plausible), while 28 score 2 or 1 (rather not likely/totally unlikely).

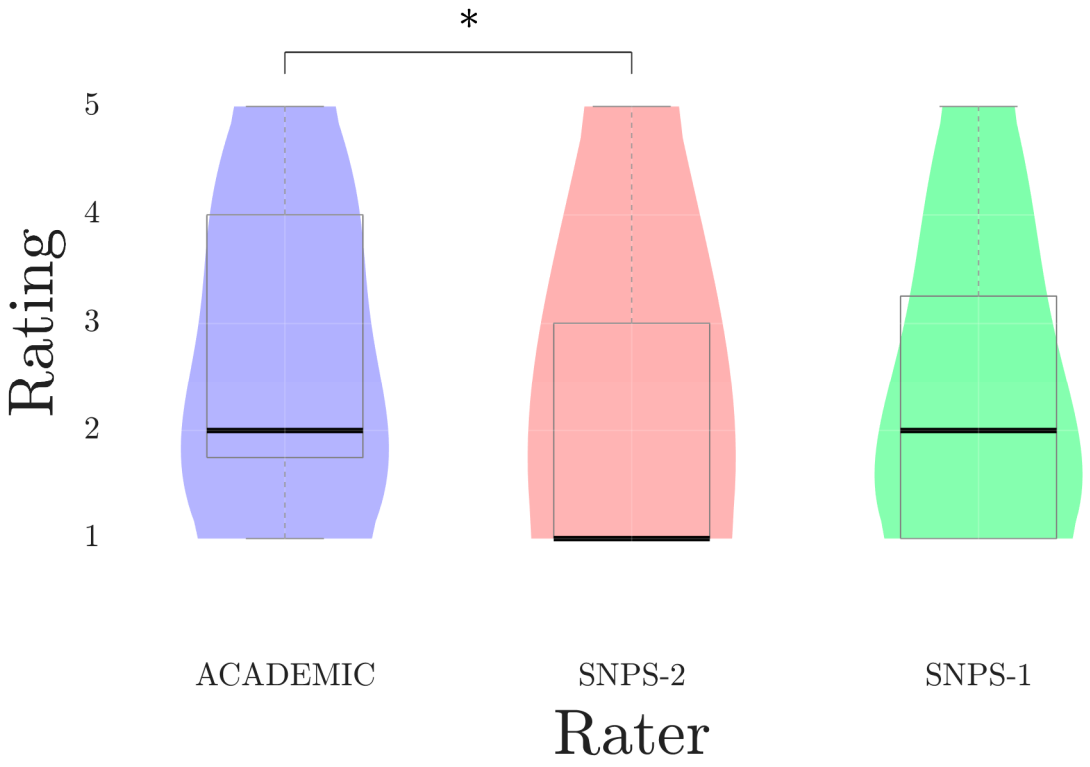

Figure 8. Plausibility ratings for each rater; *: statistically significant at the $p < 0.05$ level.

A quick look at the raters' comments shows that the disagreement between fiction and reality may stem from several reasons. SNPS-1 asserts that the visual displays are hardly ever credible. This discrepancy is particularly marked when voice comparison is concerned, with TV series showing, e.g., flashing and superimposed waveforms while the software used at SNPS outputs tables with likelihood ratios and unspectacular curves. Some differences may be the result of country-dependent rules: SNPS-1 and SNPS-2 remark that the French judicial system would not allow the recording of a person in a police interview without the person being informed prior to the interview. For example, in *Without a Trace* S02E22T31:30, as soon as the character enters the room, she declares, "I'm not sure what I can tell you, but if you think I can help. . .". Her words are instantly recorded, and within the next 10 s, her voice is compared with that of an unknown caller, and the message "Voice Print Match 100%" flashes on the computer screen. The overall feeling shared by all three raters is that the signal processing techniques used in the excerpts are usually too good to be true. Audio enhancement and source separation give much better results than what one would expect in real-life situations. The time dimension is nearly always unrealistic. While real-life forensic scientists spend long hours listening to and transcribing audio recordings, their fictional counterparts complete the job in a split second. This is the case in the aforementioned scene from *Without a Trace*. The scene in *CSI* S06E12T26:05 follows the same pattern when Forensic Scientist Nick Stokes asks another scientist: "if you've got a couple of minutes, I need a voice comparison". Another recurring critique concerns the exaggerated use of signal visualizations, as demonstrated in the previous section.

3.3. More on Aesthetics

In the context of TV shows, where visibility is key, voice analysis, as performed by experts, pertains to the immaterial. For a viewer accustomed to seeing clues, bodies, or bullets, visual representations of audio signals resolve the challenge posed by sound, invisible by nature, in a medium where the image takes precedence (Chion 2003). Non-figurative visualizations belong to a category of images commonly referred to as operational or operative (Hoel 2018). Even when they stem from the "remediation" (Bolter and Grusin 1999) of an older medium (such as analog audio recordings), operative images that are associated with

automated systems are no longer designed for the human eye; they participate in a mode of representation that depends neither on the human scale nor perspective (Hoel 2018).

In police procedural drama shows, the presence of these images emerges from a dual discursive regime that merges belief and scientific expertise, manifesting itself in staging, scientific discourse, and character typification (with a dark room and geeky technician). The operative image, sometimes the main source of light, is the basis for an explanatory exchange, where the description of processes and scientific data allows the expert to demonstrate mastery of specialized terminology.

3.4. Specialized Terminology

One aspect that makes the selected scenes potentially more convincing is the use of technical terms that are, at least partially, correct. In *The X-Files* S01E07T17:38, Special Agent Fox Mulder, after being shown two spectrograms exhibiting a near-perfect match, states, "He may have disguised his voice electronically, but he couldn't alter the formants unique to his own speech patterns". Of course, "formants" is a technical term that refers to frequency bands with high energy in the spectrogram. Now, as to whether one can alter one's formant pattern: clearly, at least the lowest three formants, those that are used to form speech sounds, can easily be altered without the help of technology. And it is reasonable to say that all formants can be altered electronically.

In *NCIS* S01E09T16:14, Forensic Specialist Abby Sciuto explains that "Ma Bell eliminates any frequency that's below 400 Hz and above 3400: it allows for longer distance transmission". To the best of our knowledge, this is accurate, but Abby's credibility quickly evaporates when, seconds later, her computer screen displays perfectly matching waveforms that lead her to conclude that they both come from the same speaker.

In *Law and Order* S11E08T07:30, the forensic specialist claims that "the average grown male has a pitch frequency of 130 Hz; a teenage boy post-puberty is about 140; this one is at 152". When detective Lennie Briscoe objects: "How does that make him a teenager?" the audio specialist replies: "People go up 10 to 15 Hertz when they're screaming". The reference values here do not seem far-fetched, and the audio specialist cautiously mentions at one point that this is just an educated guess. However, these are clearly mean values, and these reference pitch values would be of very limited use in authentic forensic contexts, given the range of within and between-speaker variation. For example, in a study involving 100 male speakers of British English aged 18–25 years old, mean individual pitch in spontaneous conversations ranges from about 85 Hz to about 140 Hz (Hudson et al. 2007).

From these three (and other) examples in our dataset, it appears that the technical jargon is not necessarily used to deceive viewers but is not to be blindly trusted: both accurate and inaccurate technical vocabulary and facts can occur within very short time windows. As experts, the authors can scrutinize techniques and jargon with a critical eye, but for the lay observer, discriminating between veracity and implausibility is challenging. The fictional aspects, exacerbated by an improbable timeline and the ease with which the unfolding of events occurs, nevertheless yield a cohesive construct to most viewers.

3.5. Lab Technicians

While technical terminology is linked to the sanctuary (a dark room illuminated by artificial light sources), the signal analysis expert is often no ordinary individual. A lonely IT genius navigating a parallel digital world of lavish sartorial opulence or Gothic fashion style, a gifted musician who "hears in perfect pitch" (*CSI* S01E08T17:20), characters with technical expertise handle seemingly incomprehensible waveforms, akin to gurus conveying an enigmatic message. Figure 9 shows a sample of forensic analysts who have eccentric behaviors or outfits. In panel A, Sherlock Holmes, who performs voice analysis himself here, is a recovering drug addict. In panel B, Forensic Scientist Abby Sciuto mixes gothic fashion with a formal lab coat. In C, the forensic specialist is a cliché Black American gifted musician with the stereotypical dress code, language, and tendency to flirt; his name says it all: Disco Placid. And in D, Penelope Garcia is yet another nonconformist analyst

who, besides being famous for her colorful clothes and eccentric behavior, is a former hacker who was caught by the FBI and given the choice to either spend the rest of her life in prison or work for them.

Figure 9. (**A**) *Elementary* S02E15T28:57. (**B**) *NCIS* S04E19T17:28. (**C**) *CSI* S01E08T17:20. (**D**) *Criminal Minds* S11E22T6:32.

4. Discussion

Forensic voice analysis in TV series shows great variation both in terms of techniques (acoustic, auditory, automatic or not, etc.) and degree of plausibility. From the grotesque perfect match between two waveforms to more moderate (and realistic) opinions and statements, popular crime procedurals mix science and entertainment. According to Kirby (2017), we live in the golden age of the fusion between the two, and our analysis confirms this: scientific terminology is applied to realistic and unrealistic contexts alike; scientific-looking waves are constrained to perform unscientific tricks, etc. Even in a recent serious podcast entirely devoted to one of us describing his profession as a forensic voice specialist, dramatic music effects and descriptions that are typical of detective books were used to glamourize the story.

Speech visualization in detective series is a recurrent motif. In addition to the static presence of more conventional items (recorders, cassette tapes, USB keys, etc.) and traditional narrative strategies (close-up shots on a silent character whose face reflects an effort of attention), speech signal visualizations substitute a dynamic solution which gives voice analysis a live and dramatically tangible dimension compared to simple listening. We noted that the preferred visual representation of the speech signal is the waveform, i.e., the amplitude-time graph, whereas it is our experience that for the analysis of speech, other visuals, like spectrograms, are much more informative. Intuitively, we feel that waveforms not only convey a "live" dimension thanks to their rapidly changing patterns, but they also favor the analogy with fingerprints. And the latter is made possible because it is easy to overlay two waveforms and let the viewer confirm that a perfect match has been found. The main difference, which is critical here, is that viewers know what fingerprints represent; they are figurative, i.e., they are a faithful representation of the thing they represent. Signal visualizations, on the contrary, are the result of a conventional synesthetic transformation that non-experts may not fully understand. It is, therefore, easy to trick people into believing that waveform matches are as robust as fingerprint matches.

The plausibility ratings we obtained from two forensic scientists and one academic specializing in speech processing show low values, supporting the overall lack of realism. However, the results are just preliminary. A tentative explanation for the lack of agreement between the academic and the two raters from SNPS is that the academic probably assessed plausibility with respect to what speech processing techniques would allow, whereas the SNPS colleagues evaluated plausibility against their actual professional practice. But only a more comprehensive rating scheme, with more raters and a more detailed set of instructions, would allow robust generalizations. In particular, the current ratings assess the plausibility of these excerpts against forensic audio analysis in France and the French judiciary. A panel of international forensic voice specialists would, therefore, be a useful addition to the current study.

The reception of American fictional crime shows by French viewers warrants a few comments. English-speaking TV series are systematically dubbed; it is impossible to quantify how many viewers switch to the original soundtrack, and we, therefore, cannot guarantee that they heard the exact terms we comment on in this article. A comparative study of the French and English versions of the dialogues constitutes an interesting potential follow-up. A remarkable by-product of dubbing is that it confirms the strong influence of US TV series on French audiences. French judges have been annoyed at being often called "Votre Honneur" (a calque from the English "Your Honor" that is very frequent in dubbed versions), and policemen are reportedly irritated when someone they have just arrested wants to make the phone call people in custody in American TV series generally make (Villez 2005).

The discrepancy between the primary intended target—the North American audience—and the secondary, French, viewership may have unexpected effects. *Law and Order*, for example, has a very local flavor: references to real events and criminal cases that took place in NYC are numerous, many actors from other shows have participated, two mayors (Bloomberg and Giuliani) have appeared as themselves, etc. In short, *Law and Order* has become an institution that reflects the local context (Villez 2014). French viewers are bound to miss a number of these allusions and references, resulting in an increased distance between the show and its audience once it has crossed the Atlantic. Le Saulnier (2012) found that the French police officers he surveyed preferred TV crime series whose setting was remote from their own professional setting. Such series allow them to drop their expert judgments and enjoy an action they now regard as plausible since, e.g., they do not specialize in the North American legal system.

As far as the CSI effect is concerned, our study does not go so far as to investigate a potential link between people's viewing habits and their faith in forensic evidence. Our aim was to offer an overview of the various on-screen representations of forensic voice analysis in order to examine what French jurors, police officers, and judges potentially have in mind when evidence based on voice analysis is presented to them. This by no means implies that the overstated efficiency of the techniques shown in TV series actually affects people. In fact, Ribeiro et al. (2019) studied the link between their participants' exposure to forensic science on television and their beliefs about the accuracy of various types of analyses. They did not find evidence that, as the CSI effect predicts, the more you are exposed, the more you trust these techniques. When comparing voice analysis to other techniques, such as DNA, toxicology, or blood pattern analysis (etc.), their participants responded that voice analysis was among the least accurate techniques and those that involve a high proportion of human judgment. Why this is the case is unclear to us at the moment, but perhaps we can assume that various efforts to vulgarize forensic science and voice analysis (Gully et al. 2022; Mauriello 2020; Smith 2023) have come to fruition, and we hope the current article will serve the same function and add to the existing body of knowledge.

Possible extensions include the collection of more recent TV shows since it appears that we are now in the post-CSI televisual age where the deductive (and fallible) reasoning that was typical of the pre-CSI series has been resurrected (Bull 2016). And quite logically, a study of forensic voice analysis in French crime series would be very informative. The

extension to the big screen would also be welcome since the various screen sizes, from the cinema to mobile phones, do not imply the same constraints to captivate the viewers (Beugnet 2022; J. Ellis 2006). Such studies would be all the more useful as Rafter (2007) has noted that contemporary fiction contributes to developing, alongside professional criminology, a "popular criminology".

Many new challenges in forensic speaker identification and voice comparison have emerged. Some of them are the result of recent technological advancements. For example, voice cloning technology, which "impersonates" someone after learning this person's main vocal features from audio samples, yields outputs that are becoming more and more convincing. And beyond sheer fraud detection, these new technological possibilities pose social, ethical, and legal problems linked, among others, to intellectual property (Watt et al. 2020). Other challenges include the impact of the media, social media, and fiction and how we, as scientists and forensic specialists, should disseminate our knowledge on voice analysis and speaker identification. The audio experts from SNPS, among the authors, insist that a sizeable part of their time is devoted to explaining the limitations of their work and that, contrary to what most people think, speaker identification and voice comparison should not be taken for granted. Now that, in recent years in France, audio specialists from SNPS have started collaborating with the academic world, one of the challenges for the near future is to maintain our efforts in this direction. In parallel, training new specialists and ensuring that forensic science complies with the basic rules of general science (transparency, peer review, replicability, etc.) are among our priorities.

5. Conclusions

The inspection of over 100 excerpts from (mostly) American TV series that portray forensic voice analysis showed that, more often than not, fiction exaggerates the possibilities of speech processing and the human ear. We expect that such fictional depictions favor the persistence of the voiceprint fallacy and the false belief that humans can identify people's voices reliably. The constraints inherent in entertainment make the various forensic techniques in TV crime fiction more efficient, more visual, and less time-consuming than their real-life counterparts. Plausibility ratings of our excerpts by two forensic audio specialists and one speech-processing researcher were very low overall. Given the relative scarcity of criminal cases involving speaker identification and voice comparison (at least in France), our default expectation is that the only representation that people (jurors, lawyers, judges, police officers, forensic scientists in other fields) have in mind come from fictional works, and TV series in particular. Here is a tentative description of the average televisual false representation of forensic voice analysis: lay people can infallibly recognize somebody's voice, especially if they are familiar with the speaker. No matter how intelligible the original audio signal is, the speaker's voice can easily be isolated from surrounding noises and enhanced if necessary. Televisual forensic experts, who tend to specialize in an unrealistically wide range of scientific disciplines and whose behavior and outfits are stereotypically eccentric, have suspects record exactly the same words as those on the questioned sample. Then, typically, strictly identical waveforms appear on a computer screen with a very high percentage supporting a "match" between the two voices. Now, back to reality, we will maintain our efforts to disseminate the type of educational content we have analyzed here, and we hope that other forensic voice specialists will use these examples to explain to others in what ways fictional depictions of forensic voice analysis may have biased their expectations. Beyond the study of fictional representations, we are quite confident that the general paradigm shift towards more scientifically validated methods in our field will increase the reliability of forensic voice analysis.

Author Contributions: Conceptualization, E.F. and A.G.T.; methodology, E.F. and A.G.T.; software, E.F.; validation, E.F., A.G.T., M.C., M.B., E.D.-B., L.G., C.S., J.-F.B. and C.F.; formal analysis, E.F., A.G.T., M.C., M.B., E.D.-B., L.G., C.S., J.-F.B. and C.F.; investigation, E.F., A.G.T., M.C., M.B., E.D.-B., L.G., C.S., J.-F.B. and C.F.; resources, E.F., A.G.T., M.C., M.B., E.D.-B., L.G., C.S., J.-F.B. and C.F.; data curation, M.C.; writing—original draft preparation, E.F. and A.G.T.; writing—review and editing, E.F.

and A.G.T.; visualization, E.F.; supervision, E.F.; project administration, E.F.; funding acquisition, E.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Idex Université de Paris, ANR-18-IDEX-0001: VoCSI-Telly-Émergence en Recherche.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data is copyrighted and cannot be shared.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] https://www.springfieldspringfield.co.uk/ (accessed on 23 January 2024).

[2] https://programme-tv.nouvelobs.com/programme-tv/ (accessed on 23 January 2024).

References

Baranowski, Andreas M., Anne Burkhardt, Elisabeth Czernik, and Heiko Hecht. 2018. The CSI-education effect: Do potential criminals benefit from forensic TV series? *International Journal of Law, Crime and Justice* 52: 86–97. [CrossRef]

Beugnet, Martine. 2022. The Gulliver effect: Screen size, scale and frame, from cinema to mobile phones. *New Review of Film and Television Studies* 20: 303–28. [CrossRef]

Boë, Louis-Jean. 2000. Forensic voice identification in France. *Speech Communication* 31: 205–24. [CrossRef]

Bolt, Richard H., Franklin S. Cooper, Edward E. David, Peter B. Denes, James M. Pickett, and Kenneth N. Stevens. 1969. Identification of a Speaker by Speech Spectrograms: How do scientists view its reliability for use as legal evidence? *Science* 166: 338–43. [CrossRef] [PubMed]

Bolter, Jay David, and Richard A. Grusin. 1999. *Remediation: Understanding New Media*. Cambridge, MA: MIT Press.

Bonastre, Jean-François. 2020. 1990–2020: Retours sur 30 ans d'échanges autour de l'identification de voix en milieu judiciaire. In *2e atelier Éthique et TRaitemeNt Automatique des Langues (ETeRNAL)*. Edited by Gilles Adda, Maxime Amblard and Karën Fort. pp. 38–47. Available online: https://aclanthology.org/2020.jeptalnrecital-eternal.5.pdf (accessed on 23 January 2024).

Broeders, Ton. 2001. Forensic Speech and Audio Analysis Forensic Linguistics 1998 to 2001. Paper presented at the 13th INTERPOL Forensic Science Symposium, Lyon, France, October 16–19; pp. 54–84.

Bull, Sofia. 2016. From crime lab to mind palace: Post-CSI forensics in *Sherlock*. *New Review of Film and Television Studies* 14: 324–44. [CrossRef]

Call, Corey, Amy K. Cook, John D. Reitzel, and Robyn D. McDougle. 2013. Seeing is believing: The CSI effect among jurors in malicious wounding cases. *Journal of Social, Behavioral, and Health Sciences* 7: 52–66.

Chion, Michel. 2003. *Un art sonore, le cinéma: Histoire, esthétique, poétique*. Paris: Cahiers du cinéma.

Cooper, Glinda S., and Vanessa Meterko. 2019. Cognitive bias research in forensic science: A systematic review. *Forensic Science International* 297: 35–46. [CrossRef]

De Jong-Lendle, Gea. 2022. Speaker Identification. In *Language as Evidence*. Edited by Victoria Guillén-Nieto and Dieter Stein. New York: Springer International Publishing, pp. 257–319. [CrossRef]

Eatley, Gordon, Harry H. Hueston, and Keith Price. 2018. A Meta-Analysis of the CSI Effect: The Impact of Popular Media on Jurors' Perception of Forensic Evidence. *Politics, Bureaucracy, and Justice* 5: 1–10.

Ellis, John. 2006. *Visible Fictions: Cinema, Television, Video (Nachdr.)*. London: Routledge.

Ellis, Stanley. 1994. The Yorkshire Ripper enquiry: Part I. *Forensic Linguistics* 1: 197–206. [CrossRef]

Gold, Erica, and Peter French. 2011. International Practices in Forensic Speaker Comparison. *International Journal of Speech, Language and the Law* 18: 293–307. [CrossRef]

Gold, Erica, and Peter French. 2019. International practices in forensic speaker comparisons: Second survey. *International Journal of Speech Language and the Law* 26: 1–20. [CrossRef]

Guiho, Mickaël. 2020. Willy Bardon condamné dans l'affaire Kulik: Les jurés expliquent leur décision. France 3 Hauts de France. Available online: https://france3-regions.francetvinfo.fr/hauts-de-france/somme/amiens/willy-bardon-condamne-affaire-kulik-jures-expliquent-leur-decision-1760827.html (accessed on 23 January 2024).

Gully, Amelia, Philip Harrison, Vincent Hughes, Richard Rhodes, and Jessica Wormald. 2022. How Voice Analysis Can Help Solve Crimes. *Frontiers for Young Minds* 10: 702664. [CrossRef]

Hoel, Aud Sissel. 2018. Operative Images. Inroads to a New Paradigm of Media Theory. In *Image—Action—Space*. Edited by Luisa Feiersinger, Kathrin Friedrich and Moritz Queisner. Berlin: De Gruyter, pp. 11–28. [CrossRef]

Hudson, Toby, Gea de Jong, Kirsty McDougall, Philip Harrison, and Francis Nolan. 2007. F0 Statistics for 100 Young Male Speakers of Standard Southern British English. Paper presented at the 16th International Congress of Phonetic Sciences: ICPhS XVI, Saarbrücken, Germany, August 6–10; pp. 1809–12. Available online: https://api.semanticscholar.org/CorpusID:17550455 (accessed on 23 January 2024).

Hudson, Toby, Kirsty McDougall, and Vincent Hughes. 2021. Forensic Phonetics. In *The Cambridge Handbook of Phonetics*, 1st ed. Edited by Rachael-Anne Knight and Jane Setter. Cambridge: Cambridge University Press, pp. 631–56. [CrossRef]
Humble, Denise, Stefan R. Schweinberger, Axel Mayer, Tim L. Jesgarzewsky, Christian Dobel, and Romi Zäske. 2022. The Jena Voice Learning and Memory Test (JVLMT): A standardized tool for assessing the ability to learn and recognize voices. *Behavior Research Methods* 55: 1352–71. [CrossRef]
Kersta, Lawrence G. 1962. Voiceprint Identification. *Nature* 196: 1253–57. [CrossRef]
Kirby, David A. 2017. *The Changing Popular Images of Science*. Edited by Kathleen H. Jamieson, Dan M. Kahan and Dietram A. Scheufele. Oxford: Oxford University Press, vol. 1. [CrossRef]
Le Saulnier, Guillaume. 2012. Ce que la fiction fait aux policiers. *Réception des médias et identités professionnelles: Travailler* 27: 17–36. [CrossRef]
Mauriello, Thomas P. 2020. *Public Speaking for Criminal Justice Professionals: A Manner of Speaking*, 1st ed. Boca Raton: CRC Press.
McDougall, Kirsty, Francis Nolan, and Toby Hudson. 2016. Telephone Transmission and Earwitnesses: Performance on Voice Parades Controlled for Voice Similarity. *Phonetica* 72: 257–72. [CrossRef]
McGehee, Frances. 1937. The reliability of the identification of the human voice. *The Journal of General Psychology* 17: 249–71. [CrossRef]
Morrison, Geoffrey S., and William C. Thompson. 2017. Assessing the admissibility of a new generation of forensic voice comparison testimony. *Columbia Science and Technology Law Review* 18: 326–434.
Morrison, Geoffrey S., Farhan H. Sahito, Gaëlle Jardine, Djordje Djokic, Sophie Clavet, Sabine Berghs, and Caroline Goemans Dorny. 2016. INTERPOL survey of the use of speaker identification by law enforcement agencies. *Forensic Science International* 263: 92–100. [CrossRef] [PubMed]
Nuance Communications. 2015. [White Paper]. The Essential Guide to Voice Biometrics. Available online: https://www.nuance.com/content/dam/nuance/en_us/collateral/enterprise/white-paper/wp-the-essential-guide-to-voice-biometrics-en-us.pdf (accessed on 23 January 2024).
Rafter, Nicole. 2007. Crime, film and criminology: Recent sex-crime movies. *Theoretical Criminology* 11: 403–20. [CrossRef]
Ratliff, Evan. 2022. Persona: The French Decepion [Audio podcast]. Pineapple Street Studios—Wondery. Available online: https://wondery.com/shows/persona/ (accessed on 23 January 2024).
Ribeiro, Gianni, Jason M. Tangen, and Blake M. McKimmie. 2019. Beliefs about error rates and human judgment in forensic science. *Forensic Science International* 297: 138–47. [CrossRef] [PubMed]
San Segundo, Eugenia, and Hermann Künzel. 2015. Automatic speaker recognition of spanish siblings: (Monozygotic and dizygotic) twins and non-twin brothers. *Loquens* 2: e021. [CrossRef]
Sidtis, Diana, and Jody Kreiman. 2012. In the Beginning Was the Familiar Voice: Personally Familiar Voices in the Evolutionary and Contemporary Biology of Communication. *Integrative Psychological and Behavioral Science* 46: 146–59. [CrossRef] [PubMed]
Smith, Peter Andrey. 2023. Can We Identify a Person from Their Voice? Digital Voiceprinting May Not Be Ready for the Courts. *IEEE Spectrum*, April 15.
Solan, Lawrence M., and Peter M. Tiersma. 2003. Hearing Voices: Speaker Identification in Court. *Hastings Law Journal* 54: 373–435.
Stevenage, Sarah V. 2018. Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. *Neuropsychologia* 116: 162–78. [CrossRef] [PubMed]
Trainum, James L. 2019. The CSI effect on cold case investigations. *Forensic Science International* 301: 455–60. [CrossRef] [PubMed]
Villez, Barbara. 2005. *Séries télé, visions de la justice*, 1st ed. Paris: Presses universitaires de France.
Villez, Barbara. 2014. *Law and Order. New York Police Judiciaire. La Justice en Prime Time*. Paris: Presses Universitaires de France. Available online: https://www.cairn.info/law-and-order-new-york-police-judiciaire--9782130594239.htm (accessed on 23 January 2024).
Watt, Dominic, and Georgina Brown. 2020. Forensic phonetics and automatic speaker recognition. In *The Routledge Handbook of Forensic Linguistics*, 2nd ed. Edited by Malcolm Coulthard, Alison May and Rui Sousa-Silva. London: Routledge, pp. 400–15. [CrossRef]
Watt, Dominic, Peter S. Harrison, and Lily Cabot-King. 2020. Who owns your voice? Linguistic and legal perspectives on the relationship between vocal distinctiveness and the rights of the individual speaker. *International Journal of Speech Language and the Law* 26: 137–80. [CrossRef]
Yarmey, A. Daniel, A. Linda Yarmey, and Meagan J. Yarmey. 1994. Face and voice identifications in showups and lineups. *Applied Cognitive Psychology* 8: 453–64. [CrossRef]
Zuluaga-Gomez, Juan, Sara Ahmed, Danielius Visockas, and Cem Subakan. 2023. CommonAccent: Exploring Large Acoustic Pretrained Models for Accent Classification Based on Common Voice. *Interspeech* 2023: 5291–95. [CrossRef]