



HAL
open science

Représentations de l'expertise vocale dans les séries policières : quand la fiction s'invite dans les enquêtes et au tribunal

Emmanuel Ferragne, Anne Guyot-Talbot, Margaux Cecchini, Martine Beugnet, Emmanuelle Delanoë-Brun, Laurianne Georgeton, Christophe Stécoli, Jean François Bonastre, Corinne Fredouille

► To cite this version:

Emmanuel Ferragne, Anne Guyot-Talbot, Margaux Cecchini, Martine Beugnet, Emmanuelle Delanoë-Brun, et al.. Représentations de l'expertise vocale dans les séries policières : quand la fiction s'invite dans les enquêtes et au tribunal. Journées d'Etudes sur la Parole, Jun 2022, Noirmoutier, France. hal-03834291

HAL Id: hal-03834291

<https://hal.science/hal-03834291>

Submitted on 29 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Représentations de l'expertise vocale dans les séries policières : quand la fiction s'invite dans les enquêtes et au tribunal

Emmanuel Ferragne¹ Anne Guyot-Talbot¹ Margaux Cecchini¹
Martine Beugnet² Emmanuelle Delanoë-Brun² Laurianne Georgeton³
Christophe Stécoli³ Jean-François Bonastre⁴ Corinne Fredouille⁴

(1) CLILLAC-ARP, Université de Paris, 8 place Paul Ricœur, 75 013 Paris, France

(2) LARCA (UMR 8225), Université de Paris-CNRS, 8 place Paul Ricœur, 75 013 Paris, France

(3) Service National de Police Scientifique, 31 avenue Franklin Roosevelt, 69 130 Écully, France

(4) LIA, Université d'Avignon, 339 chemin des Meinajaries, 84 000 Avignon, France

emmanuel.ferragne@u-paris.fr

RÉSUMÉ

À travers le prisme pluridisciplinaire de la phonétique, du traitement du signal et des études visuelles, et sous l'œil des experts de la police scientifique, nous avons analysé plus d'une centaine de scènes de séries américaines impliquant une expertise du signal audio et, en particulier, de la voix. Nous tentons d'identifier des schémas récurrents qui contribuent à entretenir certains mythes liés à l'identification d'un individu par sa voix. Nous évaluons le degré de plausibilité de certains extraits et proposons les prémisses d'une esthétique de l'expertise vocale. Nous souhaitons ainsi mieux comprendre comment le grand public se représente les applications criminalistiques des sciences de la parole et espérons, en confrontant la fiction à la réalité, faire passer un message pédagogique à destination non seulement des téléspectateurs, mais également des professionnels de la police et de la justice.

ABSTRACT

Voice analysis in TV crime dramas: criminal investigations and trials through the lens of fiction.

This project is based on a multidisciplinary approach, at the intersection of phonetics, signal processing and visual studies, and benefits from the input of forensic scientists. We have examined more than a hundred scenes from American series that involve the analysis of an audio signal and, more specifically, of voice recordings. We are trying to identify recurrent patterns that contribute to maintaining a number of myths related to the identification of an individual by their voice. We evaluate the degree of plausibility of some of the excerpts and sketch out an aesthetics of forensic voice analysis. The aim is to better understand how the general public perceives forensic applications of speech sciences, and hopefully, by confronting fiction with reality, to contribute to a more accurate and pedagogical account aimed not only at the viewers, but also at the police and justice professionals.

MOTS-CLÉS : comparaison de voix, criminalistique, séries policières.

KEYWORDS: voice comparison, forensics, TV crime dramas.

1 Introduction

1.1 Séries policières et *CSI Effect*

Le déferlement de séries télévisées mettant en scène des experts en criminalistique – e.g. *CSI (Crime Scene Investigation)* : *Les Experts* – depuis les années 2000 a conduit à l'émergence du *CSI Effect*. Cet effet se traduit par au moins deux phénomènes (Eatley *et al.*, 2018) : les jurés lors de procès auraient désormais des attentes démesurées concernant la présence de preuves scientifiques ; et quand les preuves scientifiques sont disponibles, ils auraient tendance à les considérer comme infaillibles. Plusieurs études évoquent un impact des séries policières non seulement sur le grand public, mais également sur la pratique des professionnels (Call *et al.*, 2013; Trainum, 2019), ainsi que celle des criminels (Baranowski *et al.*, 2018). Par exemple, l'étude de Call *et al.* (2013), qui s'appuie sur un questionnaire distribué à 60 jurés ayant siégé dans des affaires de coups et blessures volontaires aux États-Unis, rapporte que 95 % d'entre eux regardaient *CSI* et 73 % ont considéré que la série a influencé leur verdict. Trainum (2019) recense plusieurs comportements qu'il attribue à l'influence des séries : cela va de l'adoption par les enquêteurs de la terminologie entendue dans la dernière série à la mode, à la bonne résolution de conserver des preuves dans des conditions idéales au cas où la science soit un jour en mesure de les exploiter, en passant par une fâcheuse tendance à communiquer, comme leurs « collègues » télévisuels, des certitudes inébranlables lors de procès. Le *CSI Effect* confirme que ce type de fiction contribue à développer, en parallèle de la criminologie professionnelle, une « criminologie populaire » dont l'impact sociétal est à prendre très au sérieux (Rafter, 2007).

1.2 Comparaison de voix dans le domaine criminalistique en France

Ce dernier point est particulièrement saillant dès lors qu'on aborde la question de la comparaison de voix en criminalistique. En effet, définir si deux enregistrements émanent d'une même personne a souvent été présenté à l'écran comme un exercice aussi fiable que comparer des empreintes papillaires ou des traces ADN. Le mythe de l'empreinte vocale (Campbell *et al.*, 2009) a ainsi été renforcé par les séries policières alors que, simultanément, des sociétés savantes (dont l'Association Francophone de la Communication Parlée) appelaient les scientifiques à ne pas s'improviser experts (Bonastre, 2020). Les collègues de la section audio du Service National de Police Scientifique (SNPS) traitent, quant à eux, en moyenne 45 affaires par an donnant lieu à un rapport ; 22 d'entre elles concernent la comparaison de voix. Ils sont amenés à intervenir entre 2 et 4 fois par an en cour d'assise. Outre la comparaison de voix, leur travail comprend des techniques telles que l'amélioration acoustique (rendre un signal plus audible), la transcription orthographique d'enregistrements, ou encore l'authentification (détection de traces de montage). Ces activités se font en parallèle d'un investissement pédagogique marqué auprès des autres professionnels de la police et de la justice, et d'une implication forte dans des projets de recherche avec des universitaires.

1.3 Objectifs

Nous esquissons ici l'analyse de plus d'une centaine d'extraits de séries américaines diffusées en France ces cinquante dernières années présentant des scènes dans lesquelles une expertise du signal audio, et en particulier de la voix, est menée. Nous commençons par un bref aperçu de la diversité des situations rencontrées à l'écran, pour ensuite les soumettre à des jugements de plausibilité. Puis,

nous focalisant sur les représentations visuelles du signal, nous proposons une ébauche d'esthétique de l'expertise vocale à l'écran, dans l'optique d'identifier des mécanismes récurrents et de décrypter les divers procédés de mise en scène qui viennent s'interposer entre la réalité scientifique et les téléspectateurs.

2 Méthodologie de constitution de la base VoCSI-Telly

Le site internet *Springfield! Springfield!*¹ – seulement accessible via *Internet Archive* pendant nos recherches – contient les transcriptions orthographiques de plusieurs milliers de séries et de films anglophones. Nous avons exploré manuellement les titres disponibles, très majoritairement ceux du genre policier, et avons retenu 50 séries dont nous savons qu'elles ont été ou sont actuellement diffusées en France sur les chaînes non cryptées de la TNT (ou, avant, sur leur équivalent analogique). Nous avons ensuite conçu un programme pour extraire des concordances avec pour cible le mot *voice*, et 80 caractères de contexte avant et après. Notre base de données initiale comporte 5 372 épisodes pour un total de 26 782 309 mots².

Les 3 321 occurrences de *voice* en contexte ainsi obtenues ont ensuite été inspectées une à une, et celles qui paraissaient en lien avec l'expertise audio, et en particulier l'identification d'un individu par sa voix, ont été repérées. Une partie des 160 passages initialement identifiés ont été écartés après visionnage car jugés non pertinents. Et deux séries, *Drop Dead Diva* et *Londres, police judiciaire* n'ont pas été accessibles à travers les plateformes que nous avons utilisées : Prime Video et FMovies. Au total, 106 scènes ainsi identifiées (voir le détail dans la Figure 1), émanant de 28 séries différentes, ont été extraites sous forme de fichiers vidéos au moyen du logiciel Xbox Game Bar sous Windows 10. Elles ont été visionnées et manuellement annotées en fonction de plusieurs critères, dont seulement une partie est présentée ici.

3 Résultats

3.1 Généralités

Notre échantillon offre un large éventail de contextes dans lesquels le signal audio, et en particulier la voix, sont analysés. Il s'agit de comparaison de voix par le biais de méthodes automatiques dans 42 extraits. Parmi ceux-ci, 20 présentent l'analyse pendant l'extrait ; les autres font référence à une analyse passée ou à venir. Un cas extrême est représenté dans la Figure 3, où on voit en haut de l'image A, le signal émanant de la voix d'un locuteur, tout en bas, la voix d'un autre locuteur, et au milieu les deux formes d'ondes parfaitement identiques en train d'être superposées dans le panneau du milieu de l'interface, aboutissant à une superposition parfaite dans l'image B, avec l'indication : « MATCH 99.675 % ». Parmi les autres types d'analyses, on dénombre 21 expertises auditives pendant lesquelles un individu est identifié par sa voix, 14 d'entre elles se déroulant dans l'extrait. Les cas restants sont assez variés ; il peut s'agir de simple écoute d'un enregistrement, impliquant parfois des techniques d'amélioration du signal, de filtrage et de séparation de sources, notamment pour analyser le bruit de fond ou encore, pour l'anecdote, identifier un Boeing 767 à partir du « pitch »

1. https://www.springfieldspringfield.co.uk/tv/_show/_episode/_scripts.php

2. Estimation ne prenant en compte que 48 séries en raison de problèmes liés à *Internet Archive*.

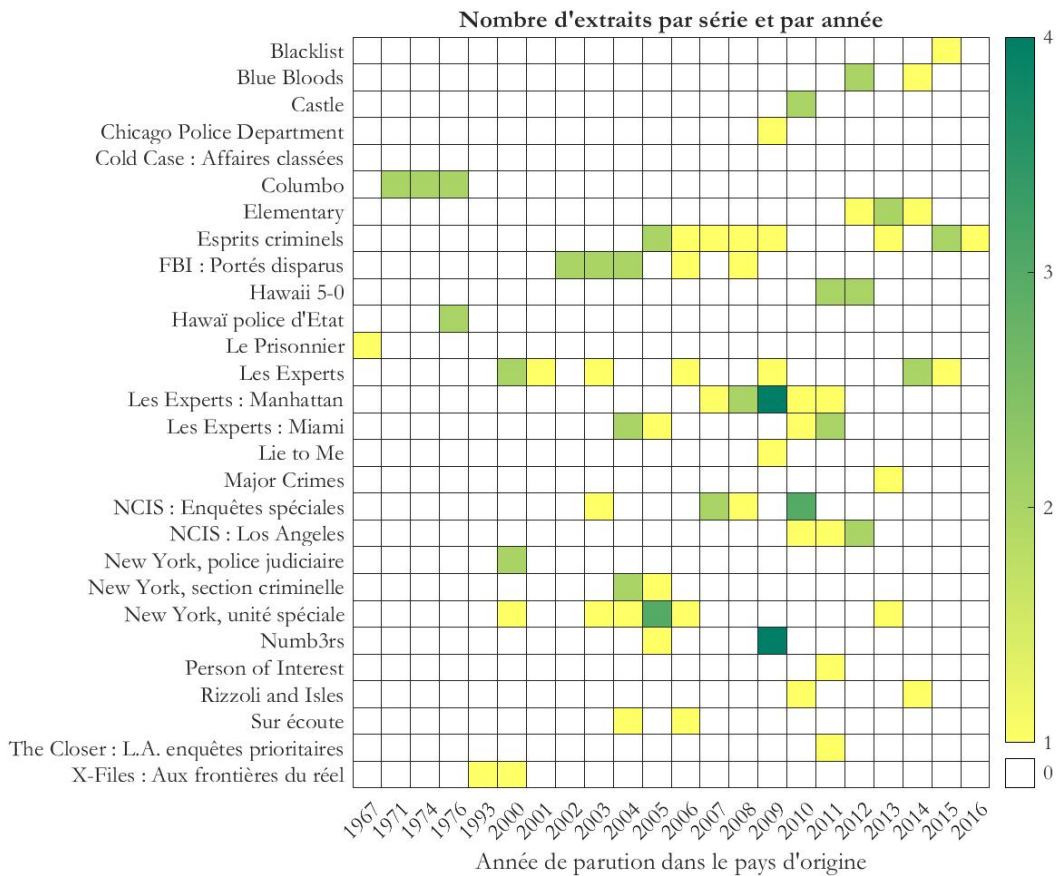


FIGURE 1 – Répartition des 106 extraits en fonction de la série.

de son moteur (*Les Experts : Manhattan* S04E03T17:18³). On trouve des allusions aux accents ou aux dialectes dans 7 épisodes. En particulier, un échantillon de voix est soumis à « a beta version of Shabboleth, [...] an accent identifier », logiciel fictif (à notre connaissance) mais néanmoins bien nommé (*NCIS : Los Angeles* S02E12T39:50). Concernant la potentielle analogie de la voix avec des empreintes biométriques, il n'est jamais fait mention d'une éventuelle similitude avec l'ADN (voir néanmoins Section 3.3). Une comparaison explicite avec les empreintes digitales est présente dans 3 extraits. Il y a en outre 4 extraits où le terme « voiceprint » est affiché sur le logiciel, et 6 extraits où ce même terme est utilisé dans les dialogues.

3.2 Réalisme des expertises dans les séries

Parmi les auteurs, les deux collègues du SNPS et une spécialiste de traitement de la parole (nommée « INFO » dans la Figure 2) ont été invités à noter la plausibilité des 41 extraits qui présentaient une visualisation du signal en utilisant une échelle allant de 1 (totalement improbable) à 5 (tout à fait plausible), 3 correspondant à la valeur « ne sais pas ». Les résultats sont représentés dans la Figure 2, où la taille des disques est proportionnelle au nombre d'épisodes. Un test de Kruskal-Wallis permet de détecter une différence de score médian entre les 3 évaluateurs ($\chi^2 = 6,62$ $p = 0,04$), qui, après comparaison post-hoc, est due à des scores globalement plus faibles chez SNPS-2 par rapport à INFO. L'analyse des κ de Cohen montre que seuls les jugements des deux membres du SNPS présentent une cohérence significative, ce qui traduit peut-être deux manières différentes d'envisager la plausibilité : plus générale pour INFO, et plus en rapport avec leur quotidien pour les collègues du SNPS (ce que confirment nos discussions). Dans l'attente d'une analyse plus détaillée – que les contraintes de longueur nous interdisent ici – on peut néanmoins affirmer en calculant un score médian par extrait, que seuls 8 d'entre eux obtiennent 4 ou 5 (assez/tout à fait plausible), et 28 obtiennent 2 ou 1 (peu probable ; totalement improbable). Il est typiquement reproché à certaines méthodes présentées à l'écran d'être plus efficaces que dans la réalité. Une autre critique récurrente concerne l'usage exagéré de visualisations du signal, qui font l'objet de la prochaine section.

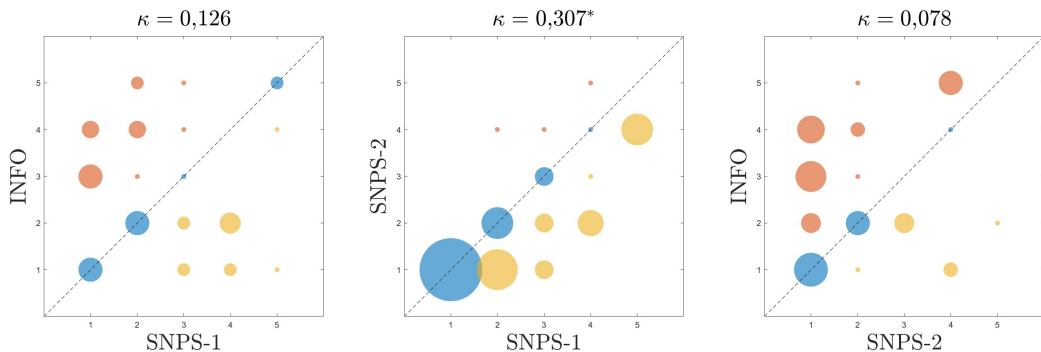


FIGURE 2 – Jugements de plausibilité sur une échelle de 1 (totalement improbable) à 5 (tout à fait plausible). Taille du disque proportionnelle au nombre d'extraits. Bleu : accord. Jaune : score plus élevé pour l'évaluateur en x. Rouge : score plus élevé pour l'évaluateur en y.

3. Par convention : S : saison ; E : épisode ; T : repère temporel au format minutes:secondes.

3.3 Représentations visuelles du signal

Des représentations visuelles du signal audio apparaissent dans 66 de nos extraits ; la représentation exclusive ou nettement dominante dans 55 de ces extraits est du type amplitude-temps, qu'on appelle ici oscillogramme ou forme d'onde. Dans 5 autres cas, on a affaire à des spectres ou des spectrogrammes ; et pour les 6 cas restants, l'affichage présente une combinaison de représentations et parfois des graphiques dont la nature est difficile à déterminer. Le rôle de ces différentes visualisations est très varié, comme l'illustrent les cas emblématiques de la Figure 3. Dans l'image C, les visualisations en arrière-plan sont purement décoratives car indépendantes de l'analyse en cours. En D, le signal, à nouveau dos au technicien qui l'analyse, correspond cette fois-ci à ce qu'il écoute via son casque et voit, selon toute vraisemblance, sur les (trois) écrans devant lui ; ces oscillogrammes figurant à l'arrière-plan sont donc exclusivement destinés aux téléspectateurs. En E, on nous montre un cas typique de pistes multiples résultant d'un algorithme de séparation de sources. L'image F présente une superposition d'indices acoustiques clés évoquant un électrophorégramme d'ADN ; ce type d'illustration se retrouve 3 fois dans nos extraits, tirés de différents épisodes de la même série.

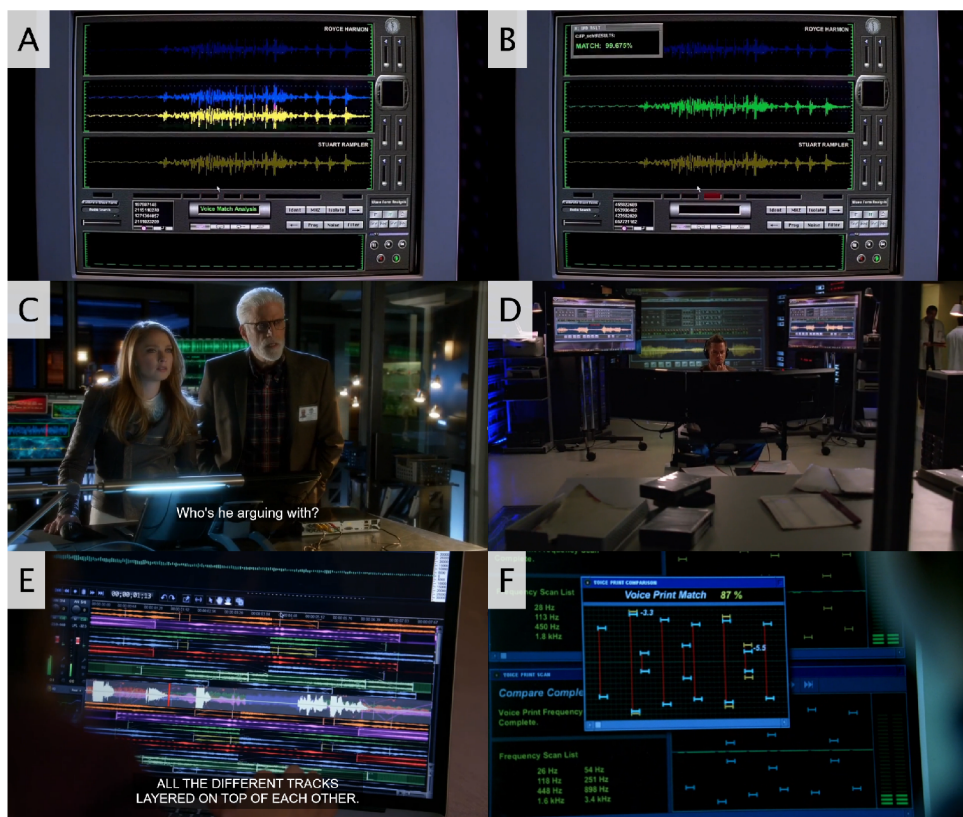


FIGURE 3 – Exemples emblématiques de représentations visuelles du signal. **A** et **B** : *Les Experts* S01E08T18:39 ; **C** : *Les Experts* S15E18T27:55 ; **D** : *Les Experts : Manhattan* S05E22T07:18 ; **E** : *Esprits criminels* S09E10T23:17 ; **F** : *FBI : Portés disparus* S02E22T06:18.

4 Esthétique de l'expertise

L'expertise vocale et sonore sert un certain nombre de visées dans les échantillons relevés. Identifier une victime ou un coupable et les localiser, confondre un témoin récalcitrant, aboutir à des aveux, etc. Elle s'insère également dans le processus judiciaire, pour valider, ou invalider, un témoignage durant un procès. Les types d'expertise vont de l'identification empirique, à la démonstration de la toute-puissance scientifique par la reconstruction d'une conversation enregistrée à partir d'empreintes sonores lacunaires. Depuis les années 2000, la visualisation des ondes sonores et vocales sur écran offre une caution scientifique à un travail de reconnaissance vocale montré comme délicat et complexe : il relève de l'immatériel, de l'évanescent, pour un spectateur habitué à voir l'indice, le corps, la balle... Paroles et sons, en revanche, apparaissent comme des traces fugaces, qui même enregistrées, sont susceptibles de corruption, volontaire ou non. Moyennant un certain nombre de variations – plusieurs écrans où s'affichent simultanément des courbes de forme et de couleur variées, des oscillogrammes couplés à des spectrogrammes, des légendes explicatives ou encore une carte de la ville avec localisation de l'enregistrement – présente dans plus de 60 % des 106 extraits étudiés, cette visualisation de la voix, dans les séries d'investigation, est un topos du genre.

Elle permet de résoudre cette difficulté que pose le son, et plus précisément la voix, dans le contexte de fictions où la visibilité est clef. Productions audiovisuelles issues d'un médium où l'image prime (Chion, 2003), les séries d'investigations se fondent, narrativement et visuellement, sur l'établissement d'une vérité basée sur la preuve. Dès lors, ce qui est en jeu, et donc prioritairement mis en scène, c'est le déroulement de l'expertise où les indices, collectés et préservés, deviennent preuves grâce au processus d'analyses menées par des spécialistes (Esquenazi, 2007). À l'évidence, parce qu'il n'est pas directement perceptible par l'œil, le son se prête mal à un traitement, par l'image, des procédés d'observation et de détection qui contribuent à cette logique narrative de révélation. Comment rendre visible l'invisible pour le soumettre à l'examen criminalistique dont le spectateur sera le témoin ? À la présence statique des accessoires plus conventionnels (enregistreurs, bandes magnétiques, cassettes et clefs USB...), et aux stratégies narratives traditionnelles (gros plan sur un personnage silencieux dont le visage reflète un effort d'attention), les oscillogrammes, spectrogrammes et autres modes de représentation graphique substituent une solution dynamique : les courbes sont animées, leur flux et leurs variations donnent à l'analyse de la voix une dimension *live* et dramatique plus tangible que la simple écoute. Par leurs qualités graphiques – mouvement, luminosité, couleur, recadrage (effets de cadres dans le cadre) – elles contribuent aussi à complexifier et à esthétiser les environnements professionnels : les laboratoires se muent en temples technologiques illuminés par des néons, où règne la transparence et la technicité, où l'uniforme est parfois de mise.

Visualisations non-figuratives, produites par un ordinateur, les oscillogrammes et spectrogrammes appartiennent à une catégorie d'images que l'on désigne couramment comme « opératives » ou opérationnelles (Hoel, 2018). Même lorsqu'elles procèdent de la « remédiation » (Bolter & Grusin, 1999) d'un médium plus ancien (un enregistrement audio analogique par exemple), les images opératives qui sont « associées à des systèmes automatiques » ne sont plus conçues « pour l'œil humain » ; elles participent d'un régime de représentation qui ne dépend ni de l'échelle humaine, ni de la perspective (Hoel, 2018). Quoiqu'instrumentales (produites à des fins utilitaires), ces images ont pour caractéristiques d'être à la fois visibles, matérialisées, mais non décodables pour le commun des mortels sans la médiation d'un expert : elles entrent dans le domaine de la preuve tangible, mais au titre d'indices uniquement exploitables par des techniciens hautement spécialisés dont la parole apparaît dès lors incontestable. Dans les séries policières, la présence de ces images ressortit donc d'un régime discursif duel, qui fusionne croyance et expertise scientifique, et qui se manifeste à la

fois dans la mise en scène, le discours scientifique, et le typage des personnages.

Les bureaux et laboratoires des experts en criminalistique ressemblent souvent à des sanctuaires : lieux clos, sombres, dénués de fenêtres, accessibles à une équipe restreinte de policiers qui se regroupe autour de l'écran contrôlé par un.e expert.e. Principale source de lumière, l'image opérative est à la fois le centre d'attention et la base d'un échange explicatif, où la description de procédés et de données scientifiques permet à l'expert de démontrer sa maîtrise de la terminologie spécialisée. Ainsi la dimension « magique » du lieu et du processus d'élucidation trouve un contrepoint dans un jargon technique destiné à confirmer la base objective et scientifique du propos. On ne parlera donc pas de sons graves, aigus, et de qualité d'enregistrement, mais de fréquences, décibels, spectre et Hertz. On filtre, on passe un enregistrement à « l'épurateur audio » (« audio scrubber ») et on observe les variations de la « rampe d'amplification » (« amplification ramp ») afin de parvenir à des données les plus factuelles possibles.

Manipulable, malléable, le son est devenu un matériau, a gagné en épaisseur visuelle et narrative. Ce faisant, son exploitation s'affranchit de l'empirisme douteux pour devenir un domaine d'experts, médiateurs obligés auxquels est déléguée la tâche de déchiffrer la manifestation d'une vérité qui échappe au béotien. Figure d'une science magique alliant le technologique à la religiosité, le jargon au sanctuaire, le technicien déroge souvent à la normalité : le modèle du geek, génie informatique évoluant dans un monde parallèle à la société des citoyens lambdas, s'impose, avec son cortège de stéréotypes : il/elle est souvent jeune, atypique dans ses apparences et comportements, gourou d'une science incontestable parce qu'elle nous est étrangère en même temps qu'on nous la montre.

5 Conclusion

Cet article avait pour objectif d'esquisser une réflexion sur la représentation des technologies de la parole, et en particulier, de la comparaison de voix, dans les séries américaines. Partant du postulat que ces séries à grande diffusion offrent une vision potentiellement déformée de l'expertise vocale, nous avons analysé plus de 100 extraits afin de prendre la juste mesure de toutes les situations rencontrées. Nous les avons ensuite confrontés aux possibilités réelles du traitement du signal et à la réalité du métier d'expert en criminalistique. Nos jugements de plausibilité varient d'un extrait à l'autre, et la concordance entre évaluateurs reste modeste. En nous concentrant sur la modalité visuelle, nous avons notamment mis en évidence une prépondérance, parfois décorative, de la représentation amplitude-temps, dont la pertinence est souvent surestimée. Notre réflexion nous a conduit à esquisser une esthétique de l'expertise dans laquelle, par exemple, l'analogie avec un rituel à la gloire de la science toute-puissante émerge naturellement. Faute de place, nous sommes contraints à un survol d'éléments qui seront développés ultérieurement, mais il paraissait opportun de partager d'ores et déjà quelques points clés tant la représentation de la science (y compris des études sur la parole) dans les séries, et par extension, dans les médias, comporte des enjeux sociétaux essentiels.

Remerciements

Le projet VoCSI-Telly – *Voice and Crime Scene Investigators on Telly* – est financé dans le cadre de l'AAP Émergence en recherche de l'Idex Université de Paris, ANR-18-IDEX-0001.

Références

- BARANOWSKI A. M., BURKHARDT A., CZERNIK E. & HECHT H. (2018). The CSI-education effect: Do potential criminals benefit from forensic TV series? *International Journal of Law, Crime and Justice*, **52**, 86–97.
- BOLTER J. D. & GRUSIN R. A. (1999). *Remediation: Understanding New Media*. Cambridge, Mass: MIT Press.
- BONASTRE J.-F. (2020). 1990-2020 : Retours sur 30 ans d'échanges autour de l'identification de voix en milieu judiciaire. In G. ADDA, M. AMBLARD & K. FORT, Eds., *2^e Atelier Éthique et TRaitemEnt Automatique Des Langues (ETeRNAL)*, p. 38–47, Nancy, France.
- CALL C., COOK, K. A., REITZEL, D. J. & MCDOUGLE, D. R. (2013). Seeing is believing: The CSI effect among jurors in malicious wounding cases. *Journal of Social, Behavioral, and Health Sciences*, **7**(1), 52–66.
- CAMPBELL J., SHEN W., CAMPBELL W., SCHWARTZ R., BONASTRE J.-F. & MATROUF D. (2009). Forensic speaker recognition. *IEEE Signal Processing Magazine*, **26**(2), 95–103.
- CHION M. (2003). *Un art sonore, le cinéma: histoire, esthétique, poétique*. Essais. Paris: "Cahiers du cinéma".
- EATLEY G., HUESTON, H. H. & PRICE K. (2018). A Meta-Analysis of the CSI Effect: The Impact of PopularMedia on Jurors' Perception of Forensic Evidence. *Politics, Bureaucracy, and Justice*, **5**(2), 1–10.
- ESQUENAZI J.-P. (2007). Le crime en série(s). Essai de sociologie du mal américain. *Cinémas*, **16**(2-3), 240–258.
- HOEL A. S. (2018). Operative Images. Inroads to a New Paradigm of Media Theory. In L. FEIERSINGER, K. FRIEDRICH & M. QUEISNER, Eds., *Image – Action – Space*, p. 11–28. De Gruyter.
- RAFTER N. (2007). Crime, film and criminology: Recent sex-crime movies. *Theoretical Criminology*, **11**(3), 403–420.
- TRAINUM J. L. (2019). The CSI effect on cold case investigations. *Forensic Science International*, **301**, 455–460.