

Des évaluations des systèmes de vérification du locuteur à la mise en cause des expertises vocales en identification juridique

Louis-Jean Boë¹, Frédéric Bimbot², Jean-François Bonastre³, Pierre Dupont⁴

¹ Institut de la Communication Parlée, INPG-Université Stendhal, CNRS, Grenoble

² Institut de Recherche en Informatique et Systèmes Aléatoires, INRIA, Rennes

³ Laboratoire Informatique d'Avignon, Université d'Avignon et des Pays de Vaucluse

⁴ Département de Mathématiques, Université Jean Monnet, Saint Étienne

Dans certaines affaires pénales l'enregistrement d'une voix réalisé à partir d'un appel téléphonique est le seul indice dont disposent les enquêteurs. Il existe donc une demande très pressante et parfaitement justifiée de la part de la police judiciaire et des magistrats. Peut-on utiliser ces enregistrements pour orienter l'enquête, pour établir la culpabilité d'un suspect ou faire la preuve de son innocence ? On comprend tout l'intérêt que pourraient offrir des techniques fiables d'identification du locuteur. Mais la voix ne présente pas des caractéristiques individuelles aussi stables et fiables que celles des empreintes digitales (Galton, 1892), utilisées par la police depuis près d'un siècle, aussi difficilement réfutables que celles des empreintes génétiques découvertes beaucoup plus récemment (Jeffreys & al., 1985). Dans l'état actuel des connaissances, il n'existe pas de procédures permettant d'avancer avec certitude qu'une personne est – ou n'est pas⁵ – l'auteur d'un appel téléphonique. Bien que certains magistrats restent très réservés sur la possibilité de retenir la voix comme un élément de preuve, des expertises vocales sont quand même effectuées dans le cadre de procédures pénales. L'analyse de la parole n'étant pas une spécialité répertoriée auprès de tribunaux, certains juges d'instruction requièrent des experts en *acoustique et vibrations* ou même des personnes supposées posséder les compétences nécessaires.

Tous les éléments réunis au cours de l'enquête préliminaire et de la phase d'instruction, tous ceux qui apparaissent lors du procès peuvent être pris en compte par le tribunal qui a toute liberté d'en apprécier la force probante pour décider ou non de la culpabilité d'un prévenu. Une expertise vocale établissant « avec certitude » qu'une personne est ou non l'auteur d'un appel anonyme peut donc jouer un rôle décisif tout au long de la procédure pénale. Un suspect peut être mis en garde à vue, puis en détention provisoire pendant plusieurs mois à la suite d'une telle expertise, un prévenu peut être condamné après avoir été identifié par sa voix.

¹ Ancien Président du Groupe de la Communication Parlée, actuellement rapporteur du Groupe Francophone de la Communication Parlée (GFCP) pour les questions d'identification juridique de la voix.

² Président du comité du GFCP.

³ Vice-président du comité du GFCP.

⁴ Membre du comité du GFCP.

⁵ Bien que cela ne semble pas évident la difficulté du problème est pourtant la même.

L'expertise intervient dans des domaines si différents⁶ qu'il est assez difficile d'établir une réglementation parfaitement adaptée à tous les domaines techniques auxquels il peut être fait appel et « qu'au lieu de la réglementation uniforme que le Code de procédure pénale a prévue, sans doute eût-il été préférable d'organiser des procédures distinctes » (Stefani & al., 1996, p. 598). Comment informer au mieux les instances judiciaires du réel danger que présente la prise en compte de rapports d'expertise qui ne présentent pas de validité scientifique ? Quels arguments scientifiques fournir pour les en convaincre ? Quelles propositions avancer pour que la preuve de la compétence soit un préalable à la réquisition d'un expert en identification vocale ? Quelles preuves de compétence peuvent être considérées comme satisfaisantes d'un point de vue scientifique ? La Justice doit-elle cesser de faire appel à des expertises vocales tant que leur validité n'aura pas été établie ?

Depuis une dizaine d'années, la communauté parole française reste très sensibilisée au problème de l'identification de la voix, compte tenu des enjeux scientifiques et de la gravité des conséquences pénales de telles expertises (Boë, 1998). Dès 1990, les membres du comité du Groupe de la Communication Parlée⁷ (GCP) du Groupement des Acousticiens de Langue Française (GALF) ont élaboré et adopté une position, après consultation de tous les laboratoires consacrant tout ou partie de leurs travaux à la recherche sur la parole. Cette position a été réitérée en 1997 : l'identification d'un locuteur reste encore un problème non résolu, les méthodes utilisées jusqu'à maintenant ne sont pas fiables. Par souci déontologique, il conviendrait que tout spécialiste démontre sa compétence en identification du locuteur avant d'accepter de procéder à une quelconque expertise vocale dans le cadre d'une affaire de justice. À la suite d'un cas judiciaire récent les spécialistes de parole viennent de demander que les expertises vocales ne soient plus utilisées par la Justice tant qu'elles n'auront pas été validées scientifiquement.

Les étapes de la procédure pénale

En France, cette procédure⁸ est actuellement réglementée par le *Code de procédure pénale* (CPP). Il nous semble important d'en rappeler brièvement les étapes, les niveaux auxquels peuvent intervenir les expertises vocales et le poids que les juges peuvent leur accorder (figure 1).

Figure 1. Les étapes de la procédure pénale.

L'enquête préliminaire

L'enquête est l'investigation effectuée pour rechercher les auteurs d'une infraction et les conditions dans lesquelles elle a été commise. Classiquement, c'est au juge d'instruction qu'il

⁶ Il existe quatre sections en domaine d'expertise subdivisées en spécialités : (1) Agriculture – Estimations foncières (14 spécialités), (2) Batiments et travaux publics (16 spécialités, dont acoustique des salles et isolation phonique), (3) Industrie – Comptabilité et disciplines connexes, autres spécialités (35 spécialités), (4) Professions de santé (36 spécialités).

⁷ Actuellement le GFCP, groupe spécialisé de la Société Française d'Acoustique (SFA, président J. Legros). Le GFCP regroupe, en France, la grande majorité des chercheurs en parole ; il organise les *Journées d'Étude sur la Parole* (22^e JEP ont eu lieu depuis 1970) qui attirent environ 150 participants. Le GFCP est aussi affilié à l'*European Speech Communication Association* (ESCA).

⁸ Il a été fixé par la loi du 31 décembre 1957, plusieurs lois ont complété ce code surtout celles du 4 janvier et du 24 août 1993. Nous laisserons ici de côté la procédure civile, qui concerne les différends entre particuliers, pour ne considérer que la procédure pénale, c'est-à-dire celle que le législateur a prévue pour défendre efficacement la société ce qui exige une répression rapide et efficace des infractions.

appartient de procéder sur des faits qui relèvent apparemment d'un délit, mais la loi a donné à la police judiciaire le pouvoir de mener des enquêtes en cas de flagrance ou non, spontanément ou à la demande du procureur de la République. Lorsque la police judiciaire a terminé son enquête préliminaire, elle en transmet les résultats au procureur de la République qui prend une décision : non-lieu ou poursuite de l'action en justice.

La procédure d'instruction

En matière répressive, toutes les affaires ne sont pas portées directement devant la juridiction de jugement. Certaines d'entre elles, plus complexes ou plus graves, sont d'abord soumises à une juridiction d'instruction. Cette étape est facultative pour les délits, obligatoire pour les crimes. Au cours de cette phase de procédure, le juge d'instruction met en œuvre les moyens permettant de réunir tous les éléments nécessaires à la manifestation de la vérité, afin qu'à l'issue du procès, le jugement puisse être rendu en toute connaissance de cause. Actuellement l'instruction est conduite selon les règles de la procédure *inquisitoire*⁹. Elle est écrite, tous les actes de l'instruction et toutes les précisions auxquelles elle donne lieu sont consignés dans un *dossier*. Le juge instruisant à charge et à décharge, l'instruction n'est, en principe, *pas contradictoire*. Il faut souligner que ce caractère a été beaucoup atténué par la loi du 4 janvier 1993 qui s'efforce de placer la personne poursuivie et la partie civile sur un pied de quasi-égalité avec le Ministère public.

Le premier acte du juge d'instruction consiste à faire connaître à la personne poursuivie qu'elle est *mise en examen* et quels sont les faits qui justifient une telle décision. S'il lui est reproché d'avoir commis un crime ou un délit, puni d'au moins 2 ans d'emprisonnement (1 an en cas de flagrant délit), il peut être placé en *détention provisoire* ; une telle décision est motivée par les nécessités de l'instruction ou la gravité du trouble causé à l'ordre public. La personne mise en examen et la partie civile sont tenues au courant du déroulement de la procédure par l'intermédiaire de leur conseil, à qui le dossier est communicable à tout moment, mais l'avocat ne peut le confier à son client. Les témoins ne sont pas mis au courant de leurs dépositions respectives et l'instruction doit rester secrète à l'égard du public. Seule exception, mais elle est d'importance : il est permis, depuis le 1er janvier 1997, aux parties ainsi qu'à leurs avocats – normalement tenus au secret professionnel et au secret de l'instruction – de communiquer à des tiers les copies des rapports d'expertise « pour le besoin de la défense »¹⁰.

Le juge d'instruction procède à la recherche des preuves et il en apprécie toute l'importance. Pour les établir il peut utiliser des écrits, des témoignages, des aveux, il peut faire procéder à des perquisitions, des saisies, etc.

Une fois réunies, les preuves sont appréciées par les juridictions d'instruction qui, selon les charges relevées, décident ou non de l'envoi de l'intéressé devant la juridiction de jugement. La *Chambre d'accusation* est la formation de la Cour d'appel qui statue sur les décisions du juge d'instruction susceptibles d'appel. Elle examine obligatoirement l'instruction de toutes les affaires criminelles avant qu'elles ne soient renvoyées, le cas échéant, devant la Cour d'assises (*mise en accusation*).

La juridiction de jugement

Cette phase de la marche du procès suit normalement les phases d'enquête préliminaire et d'instruction qui ont but de permettre au juge pénal de se prononcer sur la culpabilité de la personne poursuivie et, cette culpabilité reconnue, de décider du moyen adéquat de répression. Les juridictions de droit commun ne sont pas les mêmes selon les infractions : le

⁹ « Elle tire son nom de la formalité initiale qui domine tout le déroulement ultérieur du procès et pèse sur sa solution, l'enquête (*inquisitio*) » Stefani et al., 1996, p. 62-63.

¹⁰ Article 114, alinéa 6, du CPP issu de la loi du 30 décembre 1996 (JO du 1er janvier 1997).

Tribunal de police pour les contraventions, le *Tribunal correctionnel*¹¹ pour les délits, la *Cour d'appel* pour les affaires qui ont déjà été soumises précédemment à un tribunal, la *Cour d'assises* pour les crimes.

À la différence de la procédure d'instruction qui, nous l'avons vu, est secrète et non contradictoire, la procédure de jugement est publique, orale et contradictoire, c'est-à-dire qu'elle fonctionne selon un principe d'égalité entre les parties obligeant à soumettre tout élément et pièces à critique de la ou des autres parties.

Alors que la loi régit la recherche et l'administration des preuves, elle laisse entière liberté d'appréciation au juge. En droit civil, c'est la loi qui détermine les modes de preuves, leur admissibilité et leur valeur probante. En droit pénal, tous les modes de preuve sont admis : écrits, témoignages, aveux, perquisitions, saisies, présomptions de faits, indices, pourvu qu'ils aient été recherchés et produits dans certaines formes et suivant certaines règles, et qu'ils aient été apportés au cours des débats et contradictoirement discutés¹². La prise de la décision finale est régie par le principe fondamental de *l'intime conviction* qui peut être tirée aussi bien des éléments réunis au cours de l'enquête préliminaire, que de ceux qui ont été établis lors de l'instruction, que de ceux qui sont apparus au cours des débats. Le juge apprécie en toute liberté la valeur des preuves qui lui sont soumises. Il se décide d'après sa conscience ; il condamne, acquitte (Cour d'assises) ou relaxe (autres juridictions) suivant qu'il est ou non convaincu de la culpabilité. Dans le cas d'un crime, le juge n'est pas obligé de donner des justifications du poids qu'il attache à chacune des preuves dont il dispose¹³, alors que, dans le cas d'un délit, il doit motiver son jugement.

La voix et son expertise

La voix comme élément de preuve

Pour établir l'identité des suspects, les enquêteurs peuvent utiliser comme éléments de preuve directs ou indirects : des empreintes digitales, des traces de sang, de sperme, des cheveux, des empreintes de pas, des photos, des manuscrits, des textes dactylographiés ou imprimés, les projectiles d'une arme répertoriée... mais aussi des enregistrements réalisés avec l'autorisation du procureur de la République¹⁴ par interception de communications téléphoniques.

Le juge d'instruction a le pouvoir d'ordonner une écoute téléphonique, à condition que la peine encourue soit supérieure ou égale à deux ans d'emprisonnement. La durée ne peut excéder 4 mois (sauf renouvellement) et un procès verbal doit être établi pour chaque enregistrement¹⁵. Il peut faire intercepter et enregistrer des communications par téléphone ou par télécopieur. Peuvent être surveillées, la personne qui a été mise en examen, mais aussi des tiers, la partie civile et même un avocat (avec information du bâtonnier par le magistrat instructeur). Les enregistrements effectués sont placés sous scellés¹⁶. Ces scellés peuvent être ouverts en cas de demande de l'intéressé qui contesterait les paroles qui lui seraient attribuées.

¹¹ Afin de lutter contre le terrorisme et les atteintes à la sûreté de l'État, une loi du 9 septembre 1986 prévoit des dérogations aux règles de compétences, en permettant au procureur de la République de solliciter le dessaisissement de la juridiction normalement compétente, au profit des juridictions de Paris (art. 706-18 du CPP).

¹² Art. 427 du CPP.

¹³ Art. 353, 427 et 536 du CPP.

¹⁴ Pour la jurisprudence, on peut consulter Stefani & al., 1996, p. 36-37.

¹⁵ Loi du 10 juillet 1991, articles 100 à 100-7.

¹⁶ Art. 100-4, alinéa 2 du CPP.

La police peut procéder dans ses laboratoires à une comparaison de la voix anonyme et de celle d'un suspect. Le rapport d'analyse peut être versé au dossier de l'enquête préliminaire.

L'appel aux experts

- Au cours de l'enquête préliminaire : le procureur de la République ou l'officier de police judiciaire, sont autorisés¹⁷ de manière expresse à recourir à toute personne qualifiée¹⁸ pour les examens techniques ou scientifiques qui ne peuvent être différés. La police et la gendarmerie possèdent leurs propres laboratoires (cf. le Laboratoire Central de la Police, Paris 15^e ; l'Institut de Recherche Criminelle de Gendarmerie, Rosny-sous-Bois). Elles peuvent conduire un grand nombre des analyses et expertises dans de très nombreux domaines, grâce à un ensemble de techniques : balistique (projectiles), documents manuscrits ou imprimés (machines à écrire, imprimantes, toner), empreintes digitales, véhicules, photographie, électronique et informatique, chimie, biologie, explosifs, toxicologie, anthropologie, entomologie (insectes nécrophages), médecine légale, odontologie (identification reconstructive et détermination des morsures), analyse de la voix, etc.

En ce qui concerne les analyses de l'ADN, la justice française a commencé par faire appel à des laboratoires anglais. En 1985, Pierre Joxe, le Ministre de l'Intérieur, a inscrit la police technique et scientifique parmi les priorités de son ministère et cinq laboratoires ont été équipés dans ce domaine.

- Au cours de l'instruction : de la même manière, lorsque des compétences techniques peuvent amener à l'établissement de preuves, le juge d'instruction peut demander à être éclairé sur certains points par un spécialiste. Ce recours peut être fait aussi à la demande de la partie civile. Le code de procédure pénale a fixé la réglementation de l'expertise¹⁹. Les experts auxquels le juge d'instruction fait appel sont choisis en principe sur une liste officielle d'experts auprès des tribunaux. Il peut s'agir de personnes physiques ou de personnes morales (dans ce cas c'est leur représentant légal qui désigne le nom de la personne qui en son nom effectuera l'expertise). Le juge d'instruction choisit les experts dont il a besoin, en fonction de la confiance qu'il leur accorde, soit sur la liste nationale, soit sur l'une quelconque des listes de Cour d'appel²⁰. Le juge peut même faire appel à une personne ne figurant sur aucune de ces listes d'experts, mais il doit préciser les raisons de son choix dans une décision motivée. Le juge commet l'expert par voie d'ordonnance et précise la mission qui lui est confiée et qui ne peut avoir pour objet que l'examen de questions techniques. Depuis 1985²¹, le juge d'instruction désigne un expert chargé de procéder à l'expertise et si les circonstances le justifient, il peut en désigner plusieurs. Si la partie civile sollicite une autre expertise, le juge d'instruction peut la juger inutile et il rend dans un délai d'un mois une ordonnance motivée ; si, au contraire, il la considère opportune, il nomme le(s) expert(s).

Les experts doivent prêter serment « d'apporter leur concours à la justice en leur honneur et conscience ». Ils procèdent à leur mission sous le contrôle du juge d'instruction. En principe, les experts ne peuvent pas interroger la personne mise en examen²². Les experts sont tenus d'accomplir personnellement les opérations qui leur ont été confiées ; ce ne sont pas des substituts du juge d'instruction, ils doivent s'efforcer de l'éclairer le mieux possible sur tous

¹⁷ Loi du 30 décembre 1985, art. 77-1 du CPP, alinéas 1 et 2.

¹⁸ Art. 60 du CPP.

¹⁹ Art. 156 à 169 du CPP.

²⁰ Chaque Cour d'appel dresse la liste des experts exerçant dans son ressort. Il existe en outre une liste nationale établie par le bureau de la Cour de cassation.

²¹ Loi du 30 décembre 1985, art. 24

²² Sauf à titre exceptionnel, art. 164, al. 2 du CPP.

les points qui leur ont été définis. La rémunération des experts est prévue par le code civil²³. La responsabilité de l'expert n'est engagée qu'en cas de dol ou de faute grossière²⁴.

Les experts doivent déposer leur rapport dans le délai fixé par le juge. À la suite de quoi le juge d'instruction doit convoquer les parties intéressées et leur(s) avocat(s) pour leur donner connaissance des conclusions des experts. Dans un délai fixé, les parties peuvent formuler leurs observations et, si besoin, demander une contre-expertise.

Comme l'analyse de la parole n'existe pas en tant que spécialité auprès des tribunaux, les Juges font appel à les experts qualifiés en *acoustique et vibrations* ou à des personnes supposées posséder les compétences nécessaires.

Au cours du procès : les experts exposent oralement leurs conclusions, mais il leur est permis de consulter leur rapport et ses annexes²⁵ :

« Après leur exposé les experts peuvent assister aux débats, leur présence est d'autant plus utile qu'un témoin peut contredire à l'audience les conclusions de l'expertise, apporter à la procédure technique des indications nouvelles. Le Président demande aux experts, au ministère public et aux parties de présenter leurs observations sur cet incident. La juridiction de jugement déclare soit qu'il sera passé outre aux débats, soit que l'affaire sera renvoyée à une date ultérieure et, dans ce cas, elle prescrit toute mesure qu'elle juge utile pour compléter l'expertise contestée²⁶. » (Stefani & al., 1996, p. 697).

Les empreintes vocales : une notion fallacieuse

La reconnaissance auditive des voix

L'homme possède des capacités à reconnaître des voix familières (pour des revues de question voir Hecker, 1971 ; Briker & Pruzanski, 1976 ; Van Lancker, Kreiman, Emmorey, 1985a). Certaines lésions cérébrales attestées dans l'hémisphère droit pariétal entraînent la perte de cette compétence, c'est la phonagnosie (Van Lancker, Cummins, Kreiman, Dobkin, 1988), de la même manière que certains accidents cérébraux peuvent conduire à l'incapacité de la reconnaissance des visages familiers (la prosopagnosie). Mais il serait évidemment trop rapide de conclure que chaque voix possède des caractéristiques qui la rendent facilement reconnaissable et unique parmi toutes les voix possibles. En effet la reconnaissance auditive des voix familières est loin d'atteindre des taux proches de 100%. Par exemple avec 29 voix connues, les 13 sujets testés par Ladefoged & Ladefoged (1980) n'ont atteint que 31% d'identification correcte sur le mot *hello*, 66% sur une phrase entière et 83% sur 30 secondes de parole. Pour 2 secondes de voix familières, les 94 sujets testés par Van Lancker et al (1985) n'ont atteint que 26.6% de réussite.

Il existe donc dans la voix des caractéristiques qui permettent d'identifier le locuteur, mais les scores de reconnaissance relativement moyens tendent à montrer que les caractéristiques individuelles de la voix des locuteurs ne sont pas très fiables et/ou que les capacités auditives de reconnaissance de l'identité d'un locuteur ne sont pas très performantes. Comme nous les verrons, les logiciels de reconnaissance de la voix n'atteignent pas non plus des scores très élevés, ce qui tend à montrer que les caractéristiques individuelles de la voix ne sont pas très fiables.

²³ Articles R. 106 à R. 120-1. À titre indicatif environ 30.000 F pour une identification vocale, avec une demande de provision de 17.000 F.

²⁴ Cf. Code civil.

²⁵ Art. 168 du CPP.

²⁶ Art. 169 du CPP.

Il faut savoir aussi que la preuve n'a pas été faite que les phonéticiens, et plus généralement les spécialistes de parole, sont meilleurs que des sujets naïfs dans des tâches d'identification auditive. Pendant longtemps, en Grande-Bretagne, les preuves pour établir l'identité d'un locuteur ont été avancées à partir d'un jugement auditif. Dans le cadre d'un procès, tenu en Écosse, en novembre 1967, un homme accusé d'être l'auteur de fausses alertes d'incendies, a été déclaré coupable à la suite d'une expertise effectuée par un phonéticien requis par la Cour. En fait celui-ci, tout en présentant des sonagrammes²⁷, s'était essentiellement basé sur ses capacités auditives. Cette affaire largement médiatisée par la Presse (*The Times*, 25 novembre 1967) a fait réagir la majorité des phonéticiens. Après avoir constitué un groupe de réflexion pour prendre position sur les expertises vocales, ils ont fait part de leurs sérieuses réserves auprès du *Home Secretary*. À la suite de quoi, les Cours de justice de la Grande-Bretagne n'ont plus fait procéder à ce type d'expertise (Nolan, 1983, p. 16). En 1980, les phonéticiens britanniques ont adopté, lors du *Colloquium of British Academic Phoneticians*, par 30 voix contre 12, une motion précisant que :

« Phoneticians should not consider themselves expert in speaker identification until they have demonstrated themselves to be so » (Nolan, 1983, p. 17).

Les résultats du test conçu par Marion Shirt, à l'Université de Leeds, dans le cadre de cette discussion, donnent un ordre de grandeur des capacités de phonéticiens par rapport à des sujets naïfs pour une tâche d'identification auditive (Nolan 1983). À partir d'enregistrements de très bonne qualité, associant des paires d'échantillons de 5 secondes de parole de locuteurs masculins, les sujets soumis au test devaient indiquer si les deux échantillons provenaient ou non d'un même locuteur. Les phonéticiens étaient dotés de facilités d'écoute (magnétophone à boucle) et disposaient de tout le temps qu'ils souhaitaient, alors que les auditeurs naïfs devaient répondre en un temps très limité. Les deux catégories de sujets ont obtenu pratiquement les mêmes taux d'identification : 53% pour les phonéticiens et 46% pour les naïfs. Il s'agit donc de performances très peu fiables : les sujets se sont trompés une fois sur deux et avec des taux de réussite très dispersés ; pour les phonéticiens : 38% au minimum et 76% au maximum. Plus récemment, en 1998, dans le cadre d'évaluations de systèmes de reconnaissance du locuteur, conduites par le NIST (*National Institute of Standards and Technology*) les évaluations des systèmes les plus performants ont été comparées à des tests perceptifs menés avec des auditeurs. Dans la tâche en question les systèmes présentaient des taux d'erreur (de l'ordre de 15%) moins importants que ceux des humains (de l'ordre de 25%) mais qui ne permettent pas d'envisager une expertise juridique. Nous y reviendrons par la suite (Doddingon, 1998).

Avec le *portrait parlé*, les enquêteurs de la police disposent pour le visage d'une méthode de description qui comporte à la fois des particularités morphologiques (front, nez, oreille, bouche) et des paramètres chromatiques (yeux, cheveux, barbe). À partir de montages photographiques, les méthodes d'identification par *portrait robot* permettent de raviver le souvenir, d'en améliorer la précision et de réaliser un document qui, dans certains cas, peut conduire à l'arrestation du coupable. Avec la voix, nous sommes encore loin de disposer d'une procédure de synthèse qui permettrait, à l'aide d'un jeu de traits, de reconstituer la voix évoquée par un témoin. Des tests ont montré par ailleurs que la description de la voix n'est plus fiable si elle est effectuée par un témoin auditif plus de 24 heures après son écoute (Yarmey, 1991).

Une dénomination erronée

²⁷ Il s'agit d'une analyse spectrographique pseudo 3D (temps, fréquence, amplitude) obtenue grâce à un appareil dénommé sonagraphe.

En 1962, L.G. Kersta, dans un article paru dans *Nature* et intitulé *Voiceprint identification*, a introduit un terme qui fait encore florès dans les quotidiens, les feuillets policiers télévisés et les films policiers ou d'espionnage. Cette terminologie métaphorique erronée *d'empreinte vocale (Voiceprint)* donne à croire — et pas uniquement au grand public — qu'une représentation graphique de la voix (en l'occurrence il s'agissait d'une analyse spectrale sous forme de sonagramme) est tout aussi fiable que la structure des crêtes papillaires des pulpes des doigts ou que les empreintes génétiques, qu'elle peut donc permettre une identification fiable de son auteur.

Et l'on sait bien maintenant qu'il n'en est rien : ni Kersta, ni aucun spécialiste de parole n'a pu apporter la preuve de la fiabilité de l'identification d'un locuteur à partir de l'analyse de sonagrammes. À la suite d'expertises judiciaires controversées, le *Technical Committee on Speech Communication* de l'*Acoustical Society of America* a chargé une équipe de scientifiques de renom de faire un rapport. Ces personnalités avaient alors conclu :

« Today's consensus suggests that speaker identification by voice pattern is subject to error at a high, and yet undetermined, rate. Courts determinations may also depend on the apparent validity of exhibits brought in evidence. Spectrographic evidence may often display features that are overwhelmingly influenced by the words spoken rather than by the speaker's identity. Judge and jury may therefore be misled in understanding the evidence and in assessing an expert's testimony. » (Bolt, Cooper, David, Denes, Pickett, Stevens, 1970, p. 602).

L'usage de l'expression *empreintes vocales* constitue un véritable détournement terminologique et pourtant il perdure, même dans des ouvrages de référence²⁸. D'une façon générale, un enregistrement de parole n'est, *ni un prélèvement direct* opéré sur le corps d'un individu (permettant l'analyse des empreintes génétiques), *ni une trace laissée sur une surface au contact d'une partie de son corps* (empreintes digitales), *ni la trace d'un mouvement d'une partie du corps* (écriture), il ne s'agit que de la capture *indirecte* de mouvements articulatoires complexes faisant intervenir les cordes vocales, la langue, le voile du palais, la mâchoire et les lèvres (figure 2).

Figure 2. Les empreintes digitales sont établies à partir de l'analyse prélèvements corporels, les empreintes digitales présentent une image d'une partie du corps, l'écriture est une trace du mouvement de la main, mais le signal de parole ne reflète que les mouvements de l'air provoqués par le déplacement des organes de l'articulation.

Les mouvements des organes de la parole engendrent des variations de pression acoustique instantanée qui peuvent être captées par un transducteur et transformées en variations de tension électrique. Comme tous les gestes de l'homme, les gestes de parole ne sont pas reproductibles au cours du temps. Les paramètres utilisés pour décrire la parole montrent bien leur dépendance avec la vitesse d'articulation, l'intensité et la hauteur de la voix, l'état psychologique du locuteur et les conditions de stress. Comme on le sait bien, la reconnaissance automatique de la parole est directement confrontée à cette variabilité intralocuteur intrinsèquement liée au processus de production. De plus, il faut évidemment tenir compte des paramètres de transmission et d'enregistrement, de la possibilité d'une superposition de plusieurs voix ou de bruit. Dans le cas d'un enregistrement effectué à partir

²⁸ Ainsi dans *l'Encyclopædia Universalis* (édition de 1996) on peut lire à *Criminalistique*, un paragraphe concernant sur la parole et son identification : “ La production de la parole est un processus complexe dont le son qui en résulte est rapidement évolutif (de l'ordre de la milliseconde). [...] De nature vibratoire, les sons se caractérisent par leur fréquence (aiguë, grave, médium) [...]. Les sons qui forment la voix ne sont pas purs, mais contaminés par les harmoniques ou résidus d'impulsion. [...] Il s'agit là de chercher à reconnaître ou à vérifier par son *empreinte vocale* l'identité du sujet qui parle ” (souligné par nous).

d'une communication vont intervenir les caractéristiques du microphone, celles de la ligne téléphonique et enfin celles du magnétophone. À cela viennent évidemment s'ajouter les possibilités d'imitation et de déguisement de la voix, l'usage possible de toute une gamme de techniques de déformation allant du simple égaliseur de spectre (filtrage), au vocodeur et aux techniques de transformation de la voix par *morphing* (Genoud, 1998). Le message peut avoir été enregistré par son auteur dans des conditions inaccessibles aux enquêteurs : par exemple, pour détourner les soupçons, il peut avoir été fabriqué par le coupable à partir de la voix d'un autre locuteur. Certains téléphones digitaux, permettent de modifier la hauteur et le timbre de la voix avec plusieurs dizaines de variantes. À moins de connaître les caractéristiques du traitement mis ainsi en œuvre, il est impossible d'identifier le locuteur et même de déterminer son sexe (Al Bader, 1992).

Toutes ces difficultés sont connues des spécialistes de criminalistique. Certains, comme l'auteur²⁹ du récent *Que sais-je en Criminalistique* mettent en cause l'utilisation de la voix pour une identification juridique :

« Citons une anecdote qui quantifiera l'obstacle. Certaines zones à l'intérieur du domaine militaire voient leur accès réglementé par des "serrures" à reconnaissance vocale. La personne qui sollicite l'entrée doit prononcer [...] une phrase convenue [...]. Il se trouve que même dans ce cas de figure où le sujet fait en sorte d'être reconnu, s'appliquant pour y parvenir à prononcer, selon un volume et un débit constant, des mots déjà mémorisés par la machine, cette dernière commet un nombre considérable d'erreurs. Cela laisse à imaginer les chances actuelles du policier ou du magistrat lorsque le délinquant masque sa voix, change son accent ou sa prononciation. » (Fombonne, 1996, p. 78).

En France : des repères historiques et déontologiques

Une demande du Ministère de l'Intérieur

Le 16 octobre 1989, le Ministère de l'Intérieur (Direction des transmissions et de l'informatique) a lancé un appel à candidature pour :

« Étude, mise au point et présentation de moyens permettant une identification de locuteur par des méthodes de comparaison à partir d'enregistrements magnétiques. L'administration [fournissant] des échantillons représentatifs des besoins. Le titulaire [devant] mettre au point un système présentant des taux de reconnaissance aussi élevés que possible, le présenter en fonctionnement et réaliser l'ensemble des tests correspondant aux échantillons ».

L'annonce a paru dans le Bulletin Officiel des annonces des marchés publics. Elle a été transmise au Président du Groupe Communication Parlée et largement distribuée dans les laboratoires de recherche sur la parole.

Aucun laboratoire spécialisé dans le domaine de la parole (universitaire ou du CNRS) n'a, à notre connaissance, répondu à cet appel. L'Institut de la Communication Parlée de Grenoble a tenu à faire savoir au Ministère de l'Intérieur qu'il n'était pas possible de pouvoir procéder de manière fiable à l'identification d'un locuteur et qu'il serait très opportun de mettre au point des procédures rigoureuses d'évaluation et de commencer à constituer des bases de données.

Le Laboratoire de Micro Analyse des Surfaces de l'École Nationale Supérieure de Mécanique et de Microtechnique de Besançon (ENSMM) a été retenu par le Ministère de l'Intérieur qui avait apprécié ses compétences *en expertise balistique*. Ce contrat a permis à la

²⁹ Le capitaine J. Fombonne a fait partie de l'Institut de Recherche Criminelle de la Gendarmerie de Rosny-sous-Bois.

Société Micro Surface, créée par des enseignants et chercheurs de l'ENSMM, de développer le logiciel REVAO (Reconnaissance Vocale Assistée par Ordinateur).

Spécialisés dans l'analyse des surfaces, les contractants ont commencé par utiliser des traitements efficaces dans ce domaine : l'analyse fractale et multifractale. Mais les résultats obtenus se sont révélés bien décevants. Ils se sont alors tournés vers l'analyse spectrale à long terme, bien connue des spécialistes de parole depuis 1961 (cf. Boë, Pascal, Raymond, 1984 : Tarnoczy, Fant, 1961 ; Ramishvili, 1966 ; Mangold, 1969 ; Furui, Saito, 1971). Le spectre à long terme est en effet peu discriminant pour caractériser les locuteurs et il est évidemment très sensible à toutes les modifications spectrales introduites tout au long de la chaîne de transmission et d'enregistrement.

Les travaux effectués dans le cadre du contrat avec le Ministère de l'Intérieur n'ont donné lieu à *aucune publication* que ce soit dans une revue nationale ou internationale et n'ont donc pas fait la preuve d'une possibilité d'identification de la voix d'un locuteur. Le logiciel REVAO n'a pas été commercialisé³⁰. Très récemment une thèse a été soutenue sur la recherche d'invariants et de paramètres caractéristiques dans les signaux de parole, au sein du Laboratoire de Microanalyse des surfaces, par un des concepteurs du logiciel REVAO. Celle-ci n'a toujours pas permis de conclure à la possibilité d'identification :

« Il reste à préciser, en préalable, s'il existe des invariants phonatoires réellement pertinents et surtout si leur nombre est suffisant pour être descriptif, avec certitude, d'un locuteur. En d'autres termes, la reconnaissance d'un locuteur par comparaison à partir d'une base de données vocales ne comportant même qu'une centaine de voix répertoriées en ambiance sonore non perturbée, n'est pas certaine » (César, 1996, p. 135-136).

Une motion toujours d'actualité

En France, la communauté parole (les enseignants-chercheurs, chercheurs, ingénieurs qui travaillent dans le domaine de la communication parlée), ne dispose pas d'une structure de type *ordre* comme les médecins, les avocats, les architectes... Elle n'a même pas adopté une simple *charte* comme les journalistes³¹ en ont éprouvé la nécessité. Cependant, compte tenu de l'importance de ce problème scientifique et de la gravité de ses conséquences sociales et pénales, dès qu'ils ont eu connaissance de l'appel à candidature du Ministère de le bureau du Groupe de la Communication Parlée a élaboré et adopté en 1990 une position ferme, après consultation de l'ensemble des laboratoires concernés. Cette motion, qui a été reconduite intégralement et à l'unanimité en 1997, précise pour l'essentiel (cf. annexe) :

- que l'identification d'un locuteur est un problème non résolu et que les méthodes utilisées jusque-là ne sont pas fiables ;
- qu'il convient, par souci déontologique, que tout spécialiste se présentant comme expert en identification, fasse la preuve de ses compétences avant de procéder à une quelconque expertise.

Loin de vouloir évacuer les problèmes que posent l'analyse de la voix dans une enquête de police ou une procédure pénale les membres du comité se proposent de :

- contribuer à toute recherche, constitution de bases de données, expérimentation... qui puisse faire avancer la connaissance des caractéristiques individuelles de la voix,
- participer à l'élaboration de protocoles, à des évaluations d'experts, de logiciels ou de matériels qui se présenteraient pour résoudre les problèmes d'identification juridique.

³⁰ En 1993, la société Micro Surface a été mise en liquidation judiciaire.

³¹ Il s'agit d'un texte élaboré en 1918 par le Syndicat National des Journalistes (SNJ) et révisé en 1939. Il est reconnu par l'ensemble des organisations syndicales.

À la suite de cette motion les membres de la communauté parole, à notre connaissance, n'ont plus procédé à des expertises vocales. Actuellement il semblerait que seuls des experts de la police et de la gendarmerie continuent à être requis par certains juges d'instruction..

En France, le GFPC est un cadre bien adapté à la réflexion qui doit être poursuivie, aux initiatives qui doivent être lancées et appuyées, comme celles qui ont été menées au sein du Programme de Recherche Coordonnée Communication Homme-Machine (cf. le *Groupe Identification du locuteur*), lors du *Workshop ESCA on Automatic Speaker Recognition Identification Verification* (1994) et avec le congrès international *Speech Recognition and its Commercial and Forensic Applications RLA2C* (1998). Il n'est pas question de prescrire, mais seulement d'élaborer des positions scientifiques et déontologiques, de les faire connaître, de les conseiller à tous les spécialistes de parole et de les mettre à jour avec les avancées de la recherche. En demandant que les experts en identification vocale fassent la preuve de leurs compétences, le GFPC adopte ainsi une position comparable à celle qui prévaut dans la pratique de la médecine. En effet, l'article 39 du Code de déontologie médicale³² stipule que :

« Les médecins ne peuvent proposer aux malades ou à leur entourage comme salutaire ou sans danger un remède ou un procédé illusoire ou insuffisamment éprouvé. Toute pratique de charlatanisme est interdite ».

Des échanges entre la police, la magistrature et la communauté parole

Le Ministère de l'Intérieur a eu connaissance de la motion élaborée par la communauté parole et le Laboratoire Central de la Préfecture de Police de Paris (LCCP) a organisé le 22 novembre 1992, dans ses locaux³³, une *Journée d'échange sur la comparaison de locuteurs*³⁴. Elle a réuni 33 participants de la police française (LCCP, police judiciaire) et belge (police judiciaire), de l'École nationale de la magistrature, deux experts judiciaires, deux représentants de l'ENSMB et de la Société Micro Surface de Besançon³⁵ (en charge de l'étude correspondant à l'appel d'offre) ; la communauté parole était largement représentée et plusieurs membres du bureau du GCP étaient présents³⁶.

D'entrée, le représentant de la Magistrature³⁷ a fait état de toutes les réticences que ce type d'expertise suscitait :

- l'absence de cette spécialité dans la liste des domaines d'expertise,
- l'absence de fiabilité des méthodes,
- l'absence de certitude des conclusions.

Il a souligné les questions que soulèvent l'étude comparée :

- de la *pièce en question*, c'est-à-dire la voix anonyme enregistrée dans de mauvaises conditions techniques et dans une situation d'émotion, de peur qui transforme la voix,
- et de la *pièce de comparaison*, constituée par la lecture du texte correspondant à l'appel anonyme enregistré dans de bien meilleures conditions techniques, mais avec plus ou moins de coopération par le suspect qui peut alors modifier sa voix.

Plusieurs exposés scientifiques ont abordé les problèmes techniques de l'identification et la motion du GCP a été lue et commentée. Des enseignants de l'ENST³⁸, spécialistes en

³² Décret n° 95-1000 du 6 septembre 1995.

³³ 41, Rue de Dantzig à Paris.

³⁴ Un enregistrement de la rencontre a été diffusé par la suite à tous les participants qui en avaient fait la demande.

³⁵ Mlle Dalloul Wehbi et M. Dominique César.

³⁶ Régine André-Obrecht (chargée de recherches au CNRS), Henri Méloni (le président du GCP, professeur à l'Université d'Avignon), Claude Montacié (Maître de Conférence à l'Université Pierre et Marie Curie, Paris) et Louis-Jean Boë (ICP, Grenoble).

³⁷ Jean-Louis Croizier de l'École Nationale de la Magistrature de Bordeaux.

³⁸ Par exemple Eric Moulines, Maître de Conférences.

traitement du signal, ont longuement argumenté sur l'impossibilité d'aboutir à une identification fiable. Ils ont attiré l'attention sur le fait, qu'il était possible de modifier une voix avec un simple dispositif électronique. Des spécialistes de parole, qui avaient été consultés dans des affaires de justice³⁹, ont fait état des mêmes difficultés intrinsèques de l'identification. Aucun des participants n'a soutenu qu'il était possible d'avancer avec certitude si deux voix enregistrées provenaient ou non de la même personne. Un membre de la Société Micro Surface, alors en charge du contrat du Ministère de l'Intérieur⁴⁰ a précisé qu'il partageait cet avis et a annoncé que son laboratoire était encore à la recherche de nouveaux invariants pour pouvoir identifier un locuteur.

³⁹ Par exemple Jean-Sylvain Liénard, Directeur de Recherches au CNRS.

⁴⁰ Mlle D. Wehbi.

La justice partagée : deux affaires en question

Nous évoquerons ici deux affaires pénales : dans la première, la justice a refusé de faire appel à l'expertise vocale qui lui est apparue manquer de fiabilité, alors que, dans la seconde, elle lui a accordé un poids déterminant.

Une technique « révolutionnaire » récusée

Le contrat avec le Ministère de l'Intérieur, a permis à la Société Micro Surface de développer le logiciel REVAO (Reconnaissance Vocale Assistée par Ordinateur) :

« L'invention du logiciel a mobilisé toute une équipe : Mlle Dalloul Wehbi, docteur ès sciences physiques et gérante de "Micro Surface", [...] M. Claude Roques-Carmes, Directeur de laboratoire [...] Leur prototype révolutionnaire est capable de "décortiquer" tous les signaux d'une voix : "Notre but est de trouver un paramètre invariant, caractéristique de chaque individu", commente Mlle Wehbi. "On espère trouver une empreinte vocale comme il existe une empreinte digitale". » (L'Est Républicain, « *L'empreinte vocale* » pour identifier le corbeau , 19 septembre 1992, Jean-Pierre Mulot).

Maître Joël Lagrange apprend l'existence de ce logiciel « révolutionnaire » qui permet d'identifier les empreintes vocales. C'est l'un des trois défenseurs d'Albert et Monique Villemin, grands parents paternels du petit Grégory, assassiné le 16 octobre 1984, parties civiles dans le procès de leur belle-fille Christine, la mère de la victime. Le 18 septembre 1992, c'est à dire un mois avant la rencontre organisée par le LCPP à Paris, l'AFP communique :

« Me Lagrange a déclaré à l'AFP [...] qu'un micro-ordinateur venait d'être mis au point, qui reconstitue le spectre vocal de quiconque par des images en couleurs. Selon lui la machine permet de dire avec une certitude absolue si deux voix sont distinctes et d'établir avec un degré de certitude de 95% si deux voix sont celles d'une seule et unique personne. La partie civile entend ainsi faire comparer les voix des protagonistes de l'affaire avec celle des enregistrements des appels téléphoniques du "corbeau", afin de les disculper ou de démasquer ce correspondant anonyme qui avait annoncé la mort du petit Grégory et le lieu où se trouvait son cadavre. » (Agence AFP, 18 septembre 1992).

La demande de M^e J. Lagrange, soumise à la Chambre d'accusation de Dijon, fin 1992, est repoussée par les magistrats. L'année suivante, lors du procès de Mme Villemin, le Président O. Ruyssen considère que la Cour doit se faire une idée de la fiabilité de la « découverte » avant de requérir ou non une expertise. Il demande les témoignages de C. Roques-Carmes, professeur à l'ENSMM conseiller scientifique de la Société Micro Surface, et de la gérante, Mlle D. Wehbi qui demande à être requise comme expert en identification du locuteur :

« Malheureusement les deux spécialistes ne sont pas d'accord. Pour M. Claude Roques-Carmes : "Il n'existe pas encore d'empreintes vocales au sens où l'on parle aujourd'hui d'empreintes digitales ou d'empreintes génétiques. On y arrivera peut-être un jour mais, pour l'instant on n'en est pas là. Il n'existe pas de conviction à 100% sauf si le document fourni est récent et d'une qualité exceptionnelle". Ce n'est évidemment pas le cas des cassettes enregistrées il y a dix ans par les victimes du corbeau sur des vieux magnétophones prêtés par la gendarmerie de Corcieux. Madame Dalloul Wehbi, par contre défend [...] ce qu'elle considère comme "son" invention et qu'elle se prétend la seule à savoir faire fonctionner correctement [...] il lui faudrait entendre les cassettes avant de se prononcer. On réécoute [l'enregistrement]. Le moins que l'on puisse dire est que ce n'est pas de la haute fidélité. Pourtant Mme Wehbi s'en contenterait : "À mon avis, je peux garantir à cent pour cent le résultat si on me fournit des voix de comparaison". [...] À l'issue de l'audition le Président Ruyssen a posé clairement la question : "On vous fournit une cassette des voix du corbeau, des cassettes de voix de comparaison, pouvez-vous nous dire à 100% qui est la voix du corbeau ?". Chacun des deux spécialistes est resté sur ses positions : "Non a répondu M. Roque-Carmes, on ne peut assurer la fiabilité de la méthode il y a trop de bruit de fond", tandis que Mme Wehbi, qui n'est ni phonéticienne, ni acousticienne, mais

qui est entrée dans cette voix avec un regard nouveau, prétend le contraire, à condition répète-t-elle que les outils de référence soient fiables. » (Le Bien Public, *Le corbeau sortira-t-il de l'ordinateur ?* 18 novembre 1993).

À la suite de ces témoignages contradictoires, le Président de la Cour d'assises n'a pas accepté de faire procéder à des expertises vocales. Au moment où se tenait le procès Villemin, la société Micro Surface était déjà en liquidation judiciaire et son matériel avait été mis sous séquestre.

Des expertises lourdes de conséquences

La seconde affaire est beaucoup plus récente :

« Jérôme Prieto est basque [...] le 7 septembre 1996, un inconnu utilise une cabine publique proche du domicile de Prieto pour revendiquer la destruction de plusieurs véhicules appartenant à des policiers. La communication est écoutée. Proximité géographique et profil politique aidant, Prieto est surveillé à son tour. Un rapport d'expertise policier [effectué par Mlle dalloul Wehbi] établit alors formellement le lien entre la voix de Prieto [en français] et celle capturée par l'écoute [en basque]. » (Libération, *Prieto un basque qui défend sa voix*, Sorj Chalendon, 6 novembre 1997).

À la fin de l'année 1997, le président du comité du GFCP a été contacté par l'avocate⁴¹ de J. Prieto, mis en examen pour faits de terrorisme et placé en détention provisoire à la suite d'une expertise d'enregistrements réalisés à la suite d'écoutes téléphoniques dans le cadre d'une affaire pénale.

Le comité du GFCP a chargé un membre de la communauté parole de cette question. Celui-ci a constitué un dossier circonstancié sur la position du GFCP par rapport aux expertises vocales, sur l'historique de la motion du 7 septembre 1990, sur le contenu de la rencontre nationale organisée par le Laboratoire Central de la Police. Il a fait état des réserves que lui inspiraient précisément les expertises menées au cours de cette affaire (comparaison d'appels téléphoniques en basque et en français, utilisation de signaux temporels et d'analyses spectrales très sensibles aux conditions d'enregistrement). Lors de sa réunion du 10 octobre 1997, ce rapport a été approuvé à l'unanimité par le comité du GFCP qui a considéré d'autre part que la motion adoptée le 7 septembre 1990 restait tout à fait d'actualité ; elle a donc été reconduite intégralement et le comité a tenu à réaffirmer qu'il n'existait pas à ce jour de méthode scientifique permettant de conclure de manière formelle à l'identité de deux voix⁴².

La motion et le rapport ont été transmis au défenseur de J. Prieto. À la suite de l'intervention du GFCP, le juge d'instruction⁴³ chargé de l'affaire, a convoqué l'auteur du rapport. Après ce témoignage, le prévenu a été remis en liberté sous contrôle judiciaire. Il a passé près de 10 mois en détention provisoire. Lors de son procès le président du GFCP et l'auteur du rapport ont été entendus comme témoins : ils ont rappelé la position du comité et ses attendus scientifiques, à savoir l'impossibilité actuelle de faire des expertises vocales. Le tribunal a condamné J. Prieto à 3 ans de prison, suivant la demande du Procureur qui s'est refusé à mettre en doute la validité scientifique des expertises. Son avocat a fait appel.

Une pétition pour l'arrêt des expertises vocales

⁴¹ Maître Yolanda Molina-Ugarte, Bayonne.

⁴² Ces résolutions ont été adoptées à l'unanimité des présents ; les absents, consultés dans les jours suivants, les ont tous approuvés.

⁴³ Madame le juge Laurence Levert.

À la suite de ce jugement, avec l'appui de la Société Française d'Acoustique⁴⁴, le GFCP a lancé une pétition⁴⁵ nationale pour :

- demander « que les expertises vocales ne soient plus utilisées par la Justice tant qu'elles n'auront pas été validées scientifiquement »,
- souligner « la gravité des conséquences d'une décision de Justice fondée sur ce type d'expertise, qui pourrait constituer un précédent pour d'autres décisions »,

tout en souhaitant « la définition et la mise en place de procédures permettant d'évaluer au préalable la fiabilité de toute technique d'identification vocale qui pourrait être proposée pour réaliser des expertises » (Annexe 2).

Une lettre a été adressée au Madame le Garde des Sceaux pour « attirer [son] attention sur les problèmes que soulève l'utilisation des expertises d'identification vocale devant les juridictions pénales [...] qui ne paraît pas, en l'état actuel des connaissances scientifiques, concourir à la manifestation de la vérité ».

Le jugement en appel

Enfin la Cour d'appel de Paris, rappelant les expertises établies au cours de l'instruction, les témoignages du rapporteur du GFCP et son président, ayant entendu le Président de la Société Française d'Acoustique confirmer le peu de fiabilité des expertises vocales, a considéré que « l'identification vocale de Jérôme Prieto [...] ne constitue qu'un indice de culpabilité qui ne saurait, en l'absence d'autres éléments de l'enquête ou de l'instruction le corroborant, constituer une charge suffisante à l'encontre du prévenu auquel le doute doit bénéficier. [...] La cour, en conséquence, infirmant partiellement le jugement déféré, déclarera Jérôme Prieto non coupable des faits qui lui sont reprochés et le relaxera des fins de poursuite ».

Un cadre théorique et des données

Un cadre probabiliste

En France, comment une expertise vocale est-elle menée dans le cadre d'une procédure pénale ? Au départ, la police dispose d'un enregistrement d'une voix anonyme : la *pièce de question*. Il va être comparée avec l'enregistrement de la voix d'un suspect, la *pièce de comparaison*. Ce deuxième enregistrement peut être obtenu à partir d'écoutes téléphoniques, ou à partir de lecture du texte retranscrivant l'appel anonyme que la police ou le Juge d'Instruction demandent au suspect d'enregistrer.

À partir de ces deux *pièces*, est-il possible de répondre à la question suivante : les deux enregistrements proviennent-ils ou non de la même voix ? Le suspect et l'auteur de l'appel anonyme sont-ils, ou ne sont-ils pas, une seule et même personne ? Les experts requis par la justice ont le choix entre 7 réponses : identité formelle, identité probable, identité possible, exclusion possible, exclusion probable, exclusion formelle, conclusion impossible.

Pour essayer de répondre à cette question, plaçons-nous dans le cadre bayésien qui permet de formaliser le problème ; ce cadre n'est pas le seul, mais il est bien adapté à l'établissement des preuves dans le cadre juridique (Champod & Meuwly, 1998) et à l'identification vocale⁴⁶. Soient (figure 3) :

⁴⁴ Dont Claude Legros est l'actuel Président.

⁴⁵ <http://www.lapetition.com>

⁴⁶ Au cours d'une table Ronde organisée à l'occasion du Congrès Speaker Recognition and its Commercial and Forensic Applications, George Doddington a défendu le cadre bayésien pour la reconnaissance du locuteur.

- Ω l'ensemble des locuteurs susceptibles de produire l'ensemble des messages X
- ω_x l'inconnu que l'on cherche à identifier et qui est l'auteur du message x_q (la pièce de question)
- ω_s le suspect à qui l'on a fait produire le message x_c (la pièce de comparaison).

Figure 3. Schéma de l'ensemble des locuteurs et des messages possibles.

Mais, peut-on évaluer la probabilité pour que les messages x_c et x_q aient été ou non produits par le même locuteur (le suspect) ? Pour formuler ce problème à partir d'un modèle statistique, il faut définir, comme fonction d'un ensemble de paramètres π (les plus discriminants possibles), une mesure de probabilité $P_\pi(\omega_j | x_i)$ quantifiant la probabilité qu'il s'agisse du locuteur ω_j , un élément de l'ensemble Ω des locuteurs possibles, étant donné que l'on observe le message x_i . Cette mesure est appelée probabilité a posteriori. Par application de la formule de Bayes, cette probabilité $P_\pi(\omega_j | x_i)$ peut se réécrire comme suit :

$$P_\pi(\omega_j | x_i) = \frac{P_\pi(x_i | \omega_j) P_\pi(\omega_j)}{P_\pi(x_i)}$$

Avec :

$P_\pi(x_i | \omega_j)$ la probabilité que le message x_i soit produit, étant donné qu'il s'agit du locuteur ω_j . Cette probabilité est appelée vraisemblance (du message x_i relative au locuteur ω_j). Elle dépend en particulier des paramètres p_i qui sont sensés caractériser (dans un sens probabiliste) chaque locuteur.

$P_\pi(\omega_j)$ la probabilité a priori que le locuteur j ait produit un message, probabilité tenant compte à la fois des paramètres acoustiques, mais aussi des informations établies au cours de l'enquête ; une impossibilité matérielle (par exemple un alibi) peut réduire cette probabilité à zéro, inversement une preuve établie par ailleurs (par exemple, découverte chez le suspect d'un enregistrement correspondant à la pièce de question) peut accroître sérieusement sa valeur. On peut discuter de l'opportunité pour l'expert d'avoir accès à des éléments du dossier, ce qui ne nous semble pas souhaitable (pour une discussion voir Champod & Meuwly, 1998).

$P_\pi(x_i)$ la probabilité que x_i ait été produit par un locuteur quelconque de l'ensemble des locuteurs possibles de Ω .

En appliquant la théorie de la décision statistique (cf. par exemple Duda & Hart, 1973, Nadler & Smith, 1993), le locuteur ω_j le plus probable d'avoir produit la pièce de question est celui qui maximise la probabilité a posteriori $P_\pi(\omega_j | x_i)$. Comme le facteur $P_\pi(x_i)$ n'intervient pas dans la maximisation (sur j), il n'est pas nécessaire d'en tenir compte.

Cette procédure de décision statistique, encore appelée règle de décision bayésienne est *optimale* dans le sens qu'elle minimise la probabilité de l'erreur moyenne de classification (Nadler, Smith, 1993). En d'autres termes, si l'on applique cette règle de décision de façon répétitive sur un ensemble de locuteurs à identifier, elle nous assure, *qu'en moyenne*, l'erreur de classification est minimale. Il est important de savoir que l'*optimalité* de cette règle de décision est indépendante de la forme des distributions (Nadler, Smith, p. 347). En revanche elle nécessite que ces distributions soient connues. Dans l'état actuel des connaissances, aucune liste fiable et générale de paramètres discriminants n'a été dressée pour le français, pas plus que pour n'importe quel autre groupe linguistique. On ne peut pas non plus avancer que $P_\pi(x_i | \omega_j)$ ait été établie. Il n'est donc pas possible d'estimer avec quelle probabilité deux

énoncés ont été produits ou n'ont pas été produits par le même locuteur et a fortiori de conclure à une identité, ou à une exclusion formelle.

Comme il n'est pas possible d'établir P_π ni à partir d'un modèle théorique, ni à partir d'une analyse exhaustive menée sur tout l'ensemble des locuteurs possibles, peut-on au moins procéder à un apprentissage à partir d'un sous ensemble de Ω ? (figure 4).

Figure 4. Un sous-ensemble de locuteurs et de messages retenu pour un apprentissage.

Supposons que l'on dispose d'une base de données qui comprend les messages de locuteurs dont on connaît l'identité. Il est possible de formuler un modèle statistique de P_π et, à partir de méthodes bien connues, d'évaluer la fiabilité de cette relation sur le sous-ensemble d'apprentissage. Avant de tenter de l'appliquer au locuteur ω_x – l'inconnu que l'on cherche à identifier – et au suspect ω_x qui ne font pas partie de la base d'apprentissage, on peut l'appliquer à des locuteurs $\{\omega\}$ et des messages $\{x\}$ mis de côté pour l'apprentissage. Tout ce protocole bien établi pour l'évaluation des performances des systèmes de reconnaissance devrait être mené sur de vastes bases qu'il reste à constituer.

Des évaluations comme éléments de comparaison

À la suite des campagnes d'évaluation des performances des systèmes de reconnaissance automatique de la parole, de la qualité et de l'intelligibilité de la synthèse, ont été mis en place, depuis plusieurs années, des protocoles d'évaluation de la reconnaissance du locuteur. Ces protocoles permettent de quantifier les performances d'un système ayant pour but d'indiquer si une séquence de parole a bien été produite par un *locuteur cible* (*target speaker*) déjà enregistré et répertorié dans une phase d'apprentissage, ou s'il s'agit d'un imposteur. C'est donc une procédure comparable dans une certaine mesure, à une tâche d'identification, mais plus facile : un locuteur qui souhaite être reconnu n'a aucun intérêt à déguiser sa voix ; par contre un imposteur peut imiter la voix d'une autre personne.

Les enregistrements de parole sont recueillis de manière à représenter un échantillonnage statistique représentatif des conditions d'évaluation. Les principaux paramètres du protocole renvoient à deux types de conditions.

- Les paramètres modélisables :
 - le nombre de locuteurs cible et le nombre d'imposteurs ;
 - la dépendance au texte : soit le système sait ce que le locuteur est supposé prononcer, c'est-à-dire une séquence convenue (mot de passe) ou une séquence qu'il lui est demandé de prononcer (*text prompted*) et il a été entraîné sur tous ces items possibles, soit il n'a pas de connaissance a priori sur la séquence prononcée par le locuteur (*text independent*) ; évidemment, dans ce dernier cas, les performances obtenues sont nettement moins élevées ;
 - les conditions d'apprentissage : en particulier l'intervalle de temps qui sépare l'entraînement de la vérification ; le nombre de sessions d'enregistrement ;
 - certaines conditions d'enregistrement et de transmission du signal : le rapport signal sur bruit, le type de bruit ambiant, les caractéristiques du microphone (dont on sait maintenant que les différences sont cruciales dans les scores d'évaluation) ; la nature de la transmission : analogique ou numérique, la seconde améliorant évidemment les résultats.
- Les paramètres non maîtrisés durant l'expérience (parfois simulés, mais sans garantie de réalisme) :

- l'état émotionnel du locuteur au moment des sessions d'apprentissage et de test ; sa capacité à maintenir une certaine constance dans les caractéristiques de sa voix, c'est-à-dire sa variabilité intrinsèque ;
- la coopération des locuteurs : les locuteurs sont toujours coopératifs ou tout au moins ne perturbent pas le système ; il n'est d'ailleurs pas possible de modéliser, hors contexte, des locuteurs non coopératifs ;
- les capacités d'imposture des locuteurs : dans la quasi totalité des expériences, les locuteurs ne connaissent pas les autres locuteurs, ils ne peuvent donc pas les imiter ; leur capacité d'imitation (instinctive ou apprise) ne peut être ni simulée, ni quantifiée.

Les systèmes sont évalués à partir de deux pourcentages d'erreur : le pourcentage de fausses acceptations (*false detection probability*) et le pourcentage de faux rejets (*miss probability*) calculés à partir du nombre de locuteurs cible, du nombre de fois qu'ils n'ont pas été reconnus alors qu'ils auraient dû l'être, du nombre de tentatives d'imposture et du nombre de fois que des imposteurs ont été reconnus alors qu'ils n'auraient pas dû l'être.

Si l'on se réfère aux évaluations coordonnées, en 1997 et 1998, par le NIST (*National Institute of Standards and Technology*), (cf. par exemple : Doddington, 1998 ; Przybocki, Martin, 1998) menées avec :

- 400 locuteurs cibles et 250 imposteurs,
- trois conditions d'apprentissage : une seule session d'entraînement de 2 minutes issues d'une même conversation, un seul ou deux combinés téléphoniques,
- deux conditions de test : 3, 10 et 30 secondes pour la durée de l'enregistrement du test, avec ou non le même combiné téléphonique utilisé pour l'apprentissage.

Neuf laboratoires internationaux⁴⁷ ont participé au protocole du NIST, pour un total de 17 systèmes. Les résultats ont été présentés sous la forme d'un graphique DET (*Detection Error Trade-off*), avec en abscisses et ordonnées le logarithme des pourcentages des deux types d'erreur (figure 5). Il en résulte que l'on peut diminuer le risque de fausse acceptation, mais, dans ce cas, le nombre de faux rejet s'élève. Un des points de cette courbe peut être retenu, il correspond à un taux d'égale erreur pour les deux catégories. Avec deux combinés téléphoniques différents, les erreurs sont de l'ordre de 20%, ce qui est loin d'être suffisant comme fiabilité dans une affaire pénale.

Figure 5. Graphique DET (*Detection Error Trade-off*) montrant les performances d'un système utilisant des combinés téléphoniques identiques ou différents entre l'apprentissage et la reconnaissance, pour 30 secondes de signal test, en mode *text independent* (d'après Przybocki, Martin, 1998 ; Doddington, 1998).

Au niveau de la vérification, les machines sont-elles actuellement meilleures que les humains dont on sait bien que les performances ne sont pas exploitables dans un cadre judiciaire ? Voici l'avis d'un spécialiste mondial de la reconnaissance du locuteur, George Doddington :

« Another stumbling block for automatic speaker recognition is the discrepancies between human and machine performance. Historically, humans have outperformed machines by a wide margin. However, humans and machine give different results and seem to operate quite differently. For example humans are more robust than machines to distortions and noise. On the other hand, machines have demonstrated greater ability than humans to distinguish the voices of identical twins. Perhaps we are approaching the point where machines overtake humans in speaker recognition. Figure [5] shows the results of a recent comparison of human versus machine performances, made during the 1998 NIST evaluation of text-independent speaker recognition. While neither humans nor machine did

⁴⁷ A2RT (Hollande), BBN/GTE (USA), Dragon (USA), ENST (France), EPFL (Suisse), IDIAP (Suisse), IRSA (France), LIA (France), LIP6 (France), MIT-LL (USA), OGI (USA), SRI (USA).

a stellar job, the difference was not great – about a factor of 2 in error rate. » (Doddington, 1998, p. 66).

En vérification du locuteur, les machines ne font donc pas, actuellement, bien mieux que les humains dont les faibles performances sont bien connues. Encore faut-il préciser qu'il s'agit de résultats qui ont été établis avec des locuteurs qui ne cherchaient pas à déguiser leur voix et qu'aucun bruit n'était superposé aux enregistrements (figure 6).

Figure 6. Tracés DET (*Detection Error Trade-off*) permettant la comparaison entre les performances d'un système de reconnaissance du locuteur et des auditeurs pour la même tâche (3 secondes de parole, en mode *text independent*).

Il serait donc souhaitable que ceux qui se présentent, dans le cadre d'une procédure pénale, comme experts en identification vocale soient évalués, à partir de protocoles de ce type et sur des bases de données importantes ; mais, à notre connaissance, cela n'a pas été encore fait.

Conclusion et propositions

Avec des enregistrements effectués dans les meilleures conditions techniques, correspondant à un même texte, est-il possible d'affirmer scientifiquement que deux voix proviennent ou non, d'un même locuteur ? La réponse est actuellement non. Bien sûr il est possible de calculer des distances chiffrant quantitativement des ressemblances ou des différences pour un paramètre ou un jeu de paramètres donnés, d'évaluer les probabilités que deux voix proviennent d'un même locuteur, mais il n'existe aucun consensus scientifique sur le choix des paramètres permettant de calculer ces distances, pas plus que des degrés de confiance sur ces probabilités.

Même s'il existait une réponse positive à cette question, rien ne dit qu'elle resterait valable pour des conditions d'enregistrement très différentes tant au point de vue technique (microphone, réseau téléphonique, caractéristiques des enregistreurs) qu'au point de vue psychologique. Il n'existe pas non de bases de données correspondant à des enregistrements simulant les conditions dans lesquelles sont produites *les pièces de question*, alors que la police dispose de bases de données de centaines de milliers d'empreintes digitales.

Le terme d'*empreintes vocales* fréquemment utilisé dans les médias donne à croire que la voix présente des caractéristiques tout aussi fiables que celles des empreintes digitales ou génétiques. Il n'en est rien. Un enregistrement de parole n'est pas comparable à une image d'une partie des doigts, ni à un prélèvement opéré sur le corps d'une personne. L'enregistrement ne reflète que les mouvements de l'air provoqués par le déplacement des organes de la phonation qui, comme tous les gestes humains, sont non déterministes, peu reproductibles et très facilement modifiables. Pour ces raisons, la fiabilité des techniques actuelles d'identification de la voix n'est pas comparable à celle des empreintes digitales ou génétiques et ne permet pas d'aboutir à des conclusions formelles.

À partir d'enregistrements qui ont été modifiés (analogiquement ou numériquement) en vue de rendre difficile l'identification plus difficile, rien ne dit qu'il soit possible de restituer le timbre originel d'une voix et, a fortiori, de l'identifier.

Les soi-disant experts qui interviennent dans le cadre d'une procédure pénale n'ont toujours pas fait la preuve qu'ils obtiennent de bien meilleurs résultats que des sujets naïfs à qui l'on demande simplement d'écouter deux enregistrements et de répondre si oui ou non il s'agit de la voix d'un même locuteur.

En même temps, il faut bien le reconnaître, les connaissances diffusées par les médias sur la parole sont souvent affligeantes⁴⁸ : tous les spécialistes de parole en portent une part de responsabilité. C'est pourquoi le GFCP a fait un réel effort de clarification en soumettant à plusieurs dizaines de médias de la presse et de la radio un dossier contenant un exposé circonstancié sur l'identification vocale et les positions du GFCP⁴⁹. Les *empreintes vocales* et les « techniques révolutionnaires » d'identification vont poursuivre inexorablement leurs carrières médiatiques et juridiques si nous ne les démythifions pas.

Il est important que soient sensibilisés les 33 Cours d'Appel et les 181 Tribunaux de Grande Instance. Il faut maintenir les contacts déjà noués avec la Magistrature toujours en quête de l'amélioration de l'établissement des preuves ; avec la police et la gendarmerie, tout en restant fermes et vigilants sur les positions définies par le GFCP : mettre en place des bases de données (une première tentative de collaboration d'un Groupe de Travail du CNRS avec l'Institut de Recherche Criminologique de la Gendarmerie n'a pas abouti), des protocoles et procédures d'évaluation en identification des locuteurs en s'inspirant des travaux déjà effectués internationalement en reconnaissance de la parole et en reconnaissance du locuteur.

Enfin il faudrait que le Ministère de la Justice entende la position de la communauté scientifique : les conditions d'échange de point de vue sont d'autant plus favorables qu'aucune affaire mettant en jeu de manière déterminante les expertises vocales n'est en cours.

Dettes et reconnaissance

Un grand merci pour leurs informations, critiques, précisions et discussions à : Dominique Bertrand Daléchamps (Médecin, Grenoble), Serge Caillet (Directeur adjoint de l'Institut de Recherche Criminelle de la Gendarmerie, Rosny sous Bois), Gérard Chollet (ENST, Paris), Jean-Louis Croizier (École Nationale de la Magistrature, Bordeaux), Jacques Fombonne (Conseil d'État, Paris), Élisabeth Lhote (Laboratoire de Phonétique, Besançon), Yolanda Molina Ugarte (Avocate, Bayonne), Emmanuel Mazer (Leibniz, Grenoble), Claude Roques-Carmes (Laboratoire de Micro Analyse des Surfaces, Besançon), Michel Rumeau (Laboratoire Central de la Préfecture de Police, Paris) ; aux Services de documentation des journaux *Le Bien Public* (Dijon), *L'Est Républicain* (Nancy) ; aux membres du comité du GFCP, à Claude Legros, Président de la Société Française d'Acoustique ; aux membres de l'ICP, en particulier à : Christian Abry, Rafael Laboissière, Gordon Ramsay, Jean-Luc Schwartz ; et à Hélène Mathéry.

Mots clés : Reconnaissance du locuteur, identification du locuteur, expertise vocale juridique, déontologie

Key words : Speaker Recognition, Speaker Identification, Forensic Speaker Identification, Deontology

Summary

Among the sources of information used in legal identification, fingerprints and genetic data seem to provide a high degree of reliability. It is possible to evaluate the probability of confusing two individuals who might possess the same fingerprint characteristics or the same genetic markers, and to quantify the risk of a false alarm. By their very nature, these data do not vary significantly over the course of time, and they cannot be modified by a suspect.

⁴⁸ Cf. par exemple *L'affaire Martine Kempf*, pour la reconnaissance de la parole. En 1984-85, 220 articles de presse avaient vanté les mérites d'une jeune autodidacte supposée avoir inventé un appareil révolutionnaire pour la reconnaissance de la parole, le *Katalavox*. En fait il ne s'agissait que : la reconnaissance de la parole avait été inventée plus de 40 ans auparavant, commercialisée depuis plus de 15 ans Le *Katalavox*, n'avait absolument rien d'original : il existait sur le marché plusieurs dizaines d'appareils bien plus performants (Boë, Iranzo, 1994).

⁴⁹ *Le Canard Enchaîné*, *L'Humanité*, *Libération*, *Europe 1* et *France Inter* se sont fait les échos de ce problème.

The erroneous metaphoric term "voice fingerprint" or "voiceprint" leads many people (not only the general public) to believe that the voice is as reliable as the papillary ridges of the fingertips. This is not the case. A recording of speech is not a trace left on a surface by contact with part of an individual's body, nor is it a direct sample taken from the latter: it is nothing more than an indirect measurement of complex articulatory movements. The speech organs give rise to instantaneous acoustic pressure variations, and these are recovered by a transducer which converts them into variations. Like all human gestures, speech gestures are not reproducible over the course of time. The parameters used to describe speech clearly demonstrate dependence on the speed of articulation, on the pitch and loudness of the voice, on the psychological state of the speaker and on environmental stress. Automatic speech recognition is in fact directly confronted with this intra-speaker variability, which is intrinsically linked to the process of speech production.

To these factors must be added the effect of parameters describing transmission and recording, and the possibility that several voices or noise may have been superimposed. In the case where a measurement has been made over a telephone line, the characteristics of the microphone, the telephone line itself, and finally the recording device will also need to be accounted for. Experts do not always have access to all of these data. To these factors must also be added the possibility of imitation or disguise, and the possible use of a whole range of techniques for distorting the voice, ranging from simple spectral equalisation to the use of a vocoder or recent techniques such as voice morphing. The message may itself have been recorded by its author under conditions that are inaccessible to the investigators.

In certain inquiries the only evidence available may consist of a recording of a telephone conversation made by the suspect. The interest that reliable speaker identification techniques may arouse is understandable. In France there are no experts in the courts registered as specialists in voice identification. Instead, there are experts in "acoustics and vibration" who are recruited by policy and magistrates for legal identifications. Lawyers can call upon specialists in speech whose testimony will be assessed by judges.

« In the current state of the art, it does not seem possible to identify speakers by automatic methods based on procedures designed by an expert and using a speech sample recorded in optimal (i.e., laboratory) conditions - and even less so if the speech sample is of short duration (e.g., around twenty seconds), has a significant amount of background noise, was recorded in poor technical conditions and/or over a telephone, or comes from a speaker disguising or artificially modifying his or her voice. »

And yet certain magistrates in France attach importance to analyses of the voice which, along with other indices, should not be used except to help in directing an investigation. At present, a suspect can be held before judgement for several months as a result of a voice analysis.

In this communication, the authors will:

- detail the conditions under which, in France, voice analyses are carried out in the course of an investigation undertaken by the law;
- recall the statement taken by the French speech community about voice identification in the case of forensic analysis;
- attempt to define the limits and the difficulty (and impossibility) of producing a reliable statistical test;
- propose some elements to be discussed by speech scientists in collaboration with policy, magistrates to establish protocols in a scientific frame.

Résumé

Parmi les données utilisées dans le cadre de l'identification juridique les empreintes digitales et les données génétiques semblent présenter un degré de fiabilité élevé. Il est possible d'évaluer les probabilités de confondre deux individus qui auraient les mêmes points de repères digitaux ou les mêmes marqueurs génétiques et de chiffrer les risques de faux rejet. De par leur nature même, ces données ne varient pas (ou très peu) au cours du temps ; elles ne peuvent pas être modifiées par un suspect.

La terminologie métaphorique erronée « d'empreinte vocale » donne à croire (et pas uniquement au grand public) que la voix est tout aussi fiable que les crêtes papillaires des pulpes des doigts. Il n'en est rien. Un enregistrement de parole n'est ni une trace laissée sur une surface au contact d'une partie du corps d'un individu, ni un prélèvement direct opéré sur celui-ci. Comme tous les gestes de l'homme, les gestes de parole ne sont pas reproductibles au cours du temps. Les paramètres utilisés pour décrire la parole montrent bien leur dépendance avec la vitesse d'articulation, l'intensité de la hauteur de la voix, l'état psychologique du locuteur et les conditions de stress. De plus il faut évidemment tenir compte des paramètres de transmission et d'enregistrement, la possibilité d'une superposition de plusieurs voix ou de bruit. Dans le cas d'un prélèvement effectué à partir d'une ligne téléphonique vont entrer en ligne de compte les caractéristiques du microphone, celles de la ligne téléphonique et enfin celles de l'enregistreur. Les experts ne peuvent pas souvent disposer de toutes ces données. À cela vient évidemment s'ajouter les possibilités d'imitation, de déguisement et l'usage possible de toute une gamme de techniques de déformation de la voix allant du simple égaliseur de spectre, au vocodeur et aux techniques assez récentes de *morphing*. Le message peut lui-même avoir été enregistré par son auteur dans des conditions inaccessibles aux enquêteurs.

Dans certaines enquêtes le seul élément disponible est constitué par un enregistrement d'une communication téléphonique anonyme. On comprend tout l'intérêt que pourraient offrir des techniques fiables d'identification du locuteur. En France il n'existe pas d'experts auprès de tribunaux répertoriés comme spécialistes d'identification vocale. Ce sont des experts en « acoustique et vibrations » qui sont requis par les magistrats pour des identifications juridiques et les avocats peuvent faire appel à des spécialistes de parole dont le témoignage sera apprécié par les juges.

« Dans l'état actuel des connaissances il ne semble pas possible de pouvoir procéder à l'identification d'un locuteur, par des procédures mises en œuvre par un expert par une méthode automatique, à partir d'un échantillon enregistré dans des conditions. » (*Extrait de la Motion du Groupe Communication Parlée de la Société Française d'Acoustique du 7/09/1990, revotée le 10 octobre 1997*).

Et pourtant certains « experts » prétendent identifier de façon certaine la voix d'un suspect et certains magistrats accordent, en France, beaucoup d'importance à ces analyses. Actuellement une personne peut être placée en détention préventive pendant plusieurs mois à la suite d'une expertise vocale et elle peut être condamnée sur cette seule preuve.

Dans cette communication les auteurs :

- précisent les conditions dans lesquelles sont faites, en France, les expertises vocales dans le cadre d'une procédure pénale ;
- présentent un rappel historique des discussions et prises de position de la communauté parole française depuis 1990 ;
- tentent de cerner les limites, les difficultés d'une évaluation probabiliste ;
- rappellent les performances des meilleurs systèmes actuels de vérification du locuteur ;
- avancent enfin des éléments de réflexion et des propositions qui pourraient être discutées par les spécialistes de parole en collaboration avec la police, la gendarmerie et la magistrature pour faire progresser la recherche de preuve dans un cadre scientifique et aboutir à des protocoles bien balisés.

Références

- Al Bader H.O. 1992. The problem of voice disguise. *10^e Colloque International de Police Scientifique*. Lyon, 24-27 nov.
- Bimbot F., Chollet G. 1998. In *Handbook of standards and ressources for spoken language systems*. Ed. by Gibbon H.D., Moore R., Winski R. Mouton de Gruyter.
- Boë L.-J. 1998. L'identification juridique de la voix : le cas français. Historique, problématique et propositions. *RLA2C Speaker Recognition and its Commercial and Forensic Applications*, 222-239, Avignon, France.
- Boë L.J. 1994. *Invention médiatique et découverte scientifique. Un sérieux cas d'amnésie vocale*. Les Cahiers de l'ICP, Monographie, Grenoble 119 p.
- Boë L.J., Pascal D., Raymond D. 1984. Les caractéristiques individuelles de la voix. *Bibliographie indexée*. [1261 titres]. Note Technique NT/LAA/TSS/226. CNET, Lannion.
- Bolt R., Cooper F.S., David E.E. Jr., Denes P.B., Pickett J.M., Stevens K.N. 1970. Speaker identification by speech spectrograms : A scientists' view of its reliability for legal purposes. *Journal of the Acoustical Society of America* 47, 2(2), 597-612.
- Bricker P.D., Pruzansky S. 1976. Speaker recognition. In *Contemporary Issues in Experimental Phonetics* (N.J. Lass ed.), 295-326. Academic Press, New York.
- César D. 1996. Recherche d'invariants et de paramètres caractéristiques dans les signaux de parole. Thèse de Doctorat, Sciences pour l'Ingénieur. Université de Franche Comté, Besançon.
- Champod C., Meuwly D. 1998. The inference of identity in forensic speaker recognition. *RLA2C Speaker Recognition and its Commercial and Forensic Applications*, 125-133, Avignon, France.
- Doddington G. 1985. Speaker recognition-identifying people from their voices. *IEEE* 73(11), 1651-1664.
- Doddington G. 1998. Speaker recognition evaluation methodology – An overview and perspective –. *RLA2C Speaker Recognition and its Commercial and Forensic Applications*, 60-66. Avignon, France.
- Duda R.O, Hart P.E. 1973. Pattern classification and scene analysis. John Wiley, Chichester.
- Fombonne J. 1996. *La Criminalistique*. Que Sais-je, PUF, Paris.
- Galton F. 1892. *Finger prints*. Macmillan and Company, London. Reed. 1965, Da Capo Press, New York.
- Genoud D. 1998. *Reconnaissance et transformation de locuteurs. (Quelle identité pour la parole ?)*. Thèse, Systèmes de Communication. École Polytechnique Fédérale de Lausanne, Suisse.
- Hecker M. 1971. *Speaker recognition : An interpretive survey of litterature*. ASHA monographs, 16.
- Jeffreys A.J., Wilson V., Thein S.L. 1985. Individual-specific fingerprints of human DNA. *Nature* 316, 76-79.
- Kersta L.G. 1962. Voiceprint identification. *Nature* 196, 1253-1257.
- Nadler M., Smith E.P. 1993. Pattern recognition engineering. John Wiley, Chichester.
- Nolan F. 1983. *The phonetic bases of speaker recognition*. Cambridge University Press, Cambridge Grande Bretagne.
- Przybocki M.A., Martin A.M. 1998. NIST Speaker recognition evaluations. 331-335. *1rst International Conference on Language Resource & Evaluation*, Grenade, Espagne.
- Stefani G., Lévassieur G., Bouloc B. 1996. *Procédure pénale*. 16^e édition. Dalloz, Paris.
- Tosi O. 1979. *Voice identification. Theory and legal applications*. University Park Press, Baltimore.
- Van Lancker D.R., Cummings J.L., Kreiman J., Dobkin B.H. 1988. Phonagnosia : A dissociation between familiar and unfamiliar voices. *Cortex* 88, (24), 195-209.
- Van Lancker D., Kreiman J., Emmory K. 1985a. Familiar voice recognition : pattern and parameters. Part I : Recognition of backward voices. *Journal of Phonetics* 13, 13-38.
- Van Lancker D., Kreiman J., Emmory K. 1985b. Familiar voice recognition : pattern and parameters. Part II : Recognition of rate-altered voices. *Journal of Phonetics* 13, 39-52.
- Yarmey A.D. 1991. Description of distinctive and non-distinctive voices. *Journal of the Forensic Science Society* 31, 421-428.

Pour se tenir au courant des travaux dans le domaine de la reconnaissance du locuteur on peut consulter les actes des deux derniers congrès européens dans ce domaine :

Workshop ESCA on Automatic Speaker Recognition Identification Verification (1994)
Martigny, Suisse.

Speech Recognition and its Commercial and Forensic Applications RLA2C (1998), Avignon,
France.

Pour les évaluations coordonnées en 1998 par le *National Institute of Standards and Technology* (NIST) concernant les performances des systèmes de :
<http://www.nist.gov/itl/div894/894.01/spkrec98.htm>

La motion de 1991-1997

L'identification du locuteur pose de véritables problèmes tant au niveau déontologique que scientifique. Les aspects techniques ne nous semblent pouvoir être étudiés que par des laboratoires de recherche ou des industriels dotés de réelles compétences et de moyens spécifiques dans le domaine de la parole.

Dans l'état actuel des connaissances il ne semble pas possible de pouvoir procéder à l'identification d'un locuteur, par des procédures mises en œuvre par un expert par une méthode automatique, à partir d'un échantillon enregistré dans des conditions optimales, et a fortiori si celui-ci :

- est de durée très limitée (une vingtaine de secondes),
- avec fond sonore important,
- est enregistré dans de mauvaises conditions techniques et / ou à partir du téléphone,
- provient d'un locuteur qui a déguisé ou modifié artificiellement sa voix.

Le bureau du Groupe Communication Parlée (GCP) de la Société Française d'Acoustique affirme que l'identification d'un individu par sa voix est à l'heure actuelle un problème à sa connaissance non résolu. Les travaux publiés jusqu'ici indiquent que l'expertise auditive directe n'est pas fiable, pas plus que l'examen visuel ou automatique de résultats d'analyse de la voix (spectrogrammes, etc.). Les mauvais scores obtenus en laboratoire sont évidemment aggravés en situation de terrain, quand il s'agit d'identifier un locuteur sur un bref enregistrement (souvent téléphonique) accompagné de bruit de fond, à supposer même qu'il ne cherche pas à déformer sa voix.

Les membres du GCP ont travaillé et travaillent encore sur les problèmes des caractéristiques individuelles des locuteurs, y compris sur la question de leur reconnaissance automatique. En tant que scientifiques, ils sont évidemment prêts à apporter leurs compétences dans tout projet de recherche qui puisse faire avancer les connaissances dans ce domaine, ne serait-ce qu'en proposant des procédures rigoureuses d'évaluation de systèmes d'identification du locuteur.

Cependant, leur expérience leur permet justement de mesurer le peu de fiabilité des solutions apportées jusqu'ici à ce problème. Étant donné les évidentes conséquences sociales et morales de prétentions imprudentes sur ce sujet, le bureau du GCP tient à affirmer qu'à son avis les soi-disant experts en identification du locuteur ont encore à apporter la preuve vérifiable de leurs capacités. En particulier, par souci déontologique, il conviendrait que tout spécialiste démontre sa compétence en identification du locuteur avant d'accepter de procéder à une quelconque expertise (policière, judiciaire...).

Loin de vouloir évacuer ce problème, il semble fondamental au Groupe Communication Parlée de la Société Française d'Acoustique de bien définir et de préciser – dans les conditions optimales – quelles sont les limites actuelles des méthodes scientifiques. Si un organisme souhaite faire progresser les recherches dans ce domaine, les membres du GCP, s'ils sont sollicités, se proposent dans le cadre de leurs laboratoires respectifs de :

- contribuer à toute recherche, constitution de bases de données, expérimentation... qui puisse faire avancer la connaissance des caractéristiques individuelles de la voix,
- participer à l'élaboration de protocoles, à des évaluations d'experts, de logiciels ou de matériels qui se présenteraient pour résoudre les problèmes d'identification juridique.

Annexe 1. Motion adoptée, le 7 septembre 1990, par le Bureau du Groupe Communication Parlée (GCP) et reconduite intégralement, le 10 octobre 1997, par le comité du Groupe Francophone de la Communication Parlée (GFCP) de la Société Française d'Acoustique (SFA).

La pétition de 1999

Depuis de nombreuses années, les scientifiques spécialistes de l'étude de la voix, notamment le GFCP (Groupe Francophone de la Communication Parlée) de la SFA (Société Française d'Acoustique) ont pris position contre l'utilisation des expertises vocales à des fins d'identification judiciaire. Une motion en ce sens a été votée en 1990 par le GFCP et reconduite intégralement en 1997. Cette motion rappelle les limites scientifiques et technologiques de l'identification du locuteur et s'inquiète de l'absence d'évaluation des techniques utilisées dans le domaine judiciaire.

Le terme *d'empreintes vocales* fréquemment utilisé dans les médias donne à croire que la voix présente des caractéristiques tout aussi fiables que celles des empreintes digitales ou génétiques. Il n'en est rien. Un enregistrement de parole n'est pas comparable à une image d'une partie des doigts, ni à un prélèvement opéré sur le corps d'une personne. L'enregistrement ne reflète que les mouvements de l'air provoqués par le déplacement des organes de la phonation qui, comme tous les gestes humains, sont non déterministes, peu reproductibles et très facilement modifiables. Pour ces raisons, la fiabilité des techniques actuelles d'identification de la voix n'est pas comparable à celle des empreintes digitales ou génétiques et ne permet pas d'aboutir à des conclusions formelles.

La SFA, par l'intermédiaire du GFCP, ne peut reconnaître de validité aux expertises vocales, car elles ne sont soumises à aucun protocole d'évaluation fondé sur des bases scientifiques. En raison d'un cas judiciaire récent, dont le procès en appel aura lieu en Juin 1999, les spécialistes réaffirment clairement leur position et invitent leurs collègues scientifiques et tous les citoyens à exprimer, en signant cette pétition, leur très vive préoccupation quant à l'utilisation par la Justice des expertises vocales, telles qu'elles sont actuellement pratiquées.

Les soussigné(e)s :

- 1) demandent que les expertises vocales ne soient plus utilisées par la Justice tant qu'elles n'auront pas été validées scientifiquement,
- 2) tiennent à souligner la gravité des conséquences d'une décision de Justice fondée sur ce type d'expertise, qui pourrait constituer un précédent pour d'autres décisions de Justice,
- 3) souhaitent la définition et la mise en place de procédures permettant d'évaluer au préalable la fiabilité de toute technique d'identification vocale qui pourrait être proposée pour réaliser des expertises.

Annexe 2. La pétition lancée par la SFA et le GCP pour l'arrêt des expertises vocales tant qu'elles n'auront pas été validées scientifiquement.